Empirical Studies on the Diffusion and Valuation of the Internet

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Mainak Sarkar

Dissertation Director: Patrick Bayer

May, 2005

Abstract

Empirical Studies on the Diffusion and Valuation of the Internet

Mainak Sarkar

2005

In my dissertation, a collection of three essays, I study the diffusion of the personal computer and the Internet across households in the United States. Existing studies in this context are of a descriptive nature and suffer from serious methodological problems. They are therefore unable to investigate relevant policy questions, such as the predicted future dimension of the so-called Digital Divide and the dynamic impact of government programs initiated to bridge this divide. Using household data I find that the digital divide exists, i.e. substantial differences in access rates exist across different groups in the population. I am also able to predict that (ceteris paribus) this divide will not close in the near future, as some economists have argued. I forecast the exact magnitude of the divide across various dimensions such as income, education, race, etc. Second, I find strong evidence in favor of social learning or network effects that are commonly cited as justification for government interventions. This suggests a different set of policy prescriptions such as selective tutoring, compared to existing price-based policies such as subsidies to encourage the adoption of the Internet. Third, I estimate household level price elasticity for Internet access and find it to be non-negligible. I show that earlier studies that argued against such subsidies based on the low elasticities estimated were seriously biased, in part since they ignored the differentiated nature of the Internet services available to the consumer. Using elasticity estimates it is also possible to calculate approximately the entire consumer surplus derived from these new information technologies.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Economists have long recognized the central role played by new products, processes, and practices, i.e. changes in technology, in the evolution of the industrial society. Historical evidence (e.g. Mokyr (1990), Solow (1957), Griliches (1996)) confirm that advances in technological knowledge are the single most important contributing factor to long-term gains in productivity and economic growth. Following Schumpeter the three phases of technological change are traditionally identified as invention, innovation and diffusion. Invention may be defined as the first illustration of a scientific principle or proposed solution to a problem, whereas innovation is the first commercial application of the same. Diffusion is the third and final phase of the process whereby the said commercial application reaches mass acceptance.

Whereas no widely accepted theories exist explaining why, when and where invention or innovation occurs, a very large literature does exist investigating the diffusion phase of technological change, starting with the seminal study by Griliches (1957). The central questions in this context being why do certain technologies 'make it' and not others, and also what factors affect their speed of diffusion. Over the years numerous factors have been suggested and empirically tested in this context, for example, uncertainty surrounding costs or demand, high switching costs etc., are typically considered hindrances and conversely high profitability and an accompanying first-mover advantage etc., are considered conducive to faster diffusion.[1] History is rife with examples of revolutionary inventions or innovations that failed to capture enough converts to be economically viable (e.g. Betamax video recorders or Dvorak keyboards), therefore diffusion is critical since without it the economic significance of a new technology is likely to be trivial.

Also diffusion is commonly seen as an economic inefficiency or deadweight loss that is ideal for some form of policy intervention. Technological inefficiency is defined as the usage of older products, processes or practices when more modern superior alternatives exist. Therefore some economists have proposed a more proactive role for the government in encouraging the adoption of newer technologies to accelerate the diffusion process or equivalently to lower the deadweight loss to the economy. However, detractors point out that the government had proven repeatedly inept in regulating established markets,thus it can scarcely be expected to successfully navigate the shifting sands of new technologies in a fast changing world. In a context where even technological experts are often not sure which of several competing technologies shall eventually prevail. Also due to regulatory delay government interventions in new technology can be counterproductive, by preventing competition or delaying the launch of new products and services (for example, Hausman (1998) and (1999) considers the case of cell phones and automated answering services). In this context note that a critical distinction is often made in the literature between 'process' and 'product' innovations, whereas the former refers to improvements in the production process, the latter leads to direct gains to the consumer from new products and/or the creation of new markets. The diffusion studies reported later

[1] For an excellent survey of the literature see Geroski (2000) and Hall and Khan (2003).

consider new product innovations only.

The studies reported here consider the diffusion of two new information technologies, the personal computer or PC and the Internet. Although the PC has been widely available for more than two decades, it rapidly gained in popularity only with the launch of the Internet in the early nineties. Initially launched by the government to connect the computers at universities and defense laboratories in the sixties, the Internet was born in its current user-friendly avatar as the World Wide Web a little more than a decade ago. They are unique among all new technologies given both their level and speed of diffusion relatively early in their life cycle. Also few other technologies have had such widespread impact across so many sectors of the economy (and are projected to incrementally do so in the near future, see below), here in the US and across much of the developed world.

These technologies also exhibit a strong income and education bias, i.e. most of the early adopters came from the highly skilled and also high-income professional class in most countries. Therefore a widespread concern was expressed in policy circles regarding the so-called 'Digital Divide', which is the concern that certain groups in the population may not have access to these new technologies and therefore be somehow handicapped, for example in the labor market, in an increasingly wired world those without a working knowledge of computers are likely to find it harder to find a job and/or may be paid significantly less.[2] Note that most new products such as cars or televisions start out as being too expensive to be within the budget of the average consumer but over time economies of scale and innovation leads to a widening consumer base. These new information technologies however require a relatively higher investment of time and money, it can be time-consuming and costly

[2] Empirical studies have found a positive wage differential associated with using the computer as well as the Internet at work.

to learn how to use computers or the Internet, when compared to learning how to operate a television for instance. Such switching costs can operate as a significant barrier towards adoption and lead to slow diffusion rates and a very large deadweight loss given its relatively disproportionate benefit to the economy (due to the network good nature of the new technologies).

The rest of this chapter is organized as follows; we consider some of the salient features of diffusion and these new technologies. We mention the critical importance of these technologies as noted by scholars across many fields. Next data sources for these studies are discussed and descriptive statistics reported. Finally the studies in this volume are briefly described and motivated.

1.1 Diffusion

Diffusion or a dynamic process of adoption of new technologies primarily occurs due to the following reasons; first, all new technologies start out as the dark horse and require a leap of faith from the early adopters regarding its utility/profitability and its longevity in the market. Second, almost all new technologies are characterized by a rapidly improving quality and/or falling prices, which may be due network effects, learning from others (lowering switching costs) or economies of scale. Therefore the adopter plays a waiting game and adopts when the gain from waiting is counterbalanced by the loss in benefit from consumption in the current period. Third, the characteristics of the population of potential users itself might be changing over time increasing the utility from adoption. For example skills in information technology might be increasingly valuable over time as employers realize the potential of the new technology.

In a world with perfect information (i.e. in the absence of uncertainty) the market for new technologies would reach equilibrium in the same period it is introduced. Potential users will know perfectly how useful the new technology is for them and will behave rationally to adopt the technology if their utility from adoption is higher compared to the best alternative (accounting for all switching costs). The supplier(s) knowing perfectly the size of the market will optimally price their product to maximize profits. In which case there should be no dynamics involved neither any inefficiencies, since social surplus is maximized as standard neoclassical theory would predict provided this new market is perfectly competitive, even otherwise standard policies exist to maximize welfare, such as profit taxes etc. Therefore diffusion studies are studies in non-equilibrium dynamics. Alternatively diffusion can also be interpreted as a multiple equilibria switching regime where each period the technology improves and a new equilibrium point is set. Therefore even in the absence of any uncertainty a diffusion process can arise purely through the evolution of quality or prices of the new technology.

A salient feature of diffusion processes is that usage or adoption rates follow an S-shaped curve over time. Such a curve typically exhibits inertia initially when the technology is introduced, few people adopt early on, however the diffusion process accelerates over time. Subsequently the growth of users declines over time until saturation is achieved when all potential users are already using the product. Using marketing science terminology we can divide the set of users into four groups with the inflexion point on the curve defining the division between the two groups called adopters and imitators. Adopters are those who are first to adopt the new technology and imitators learn from them about the existence and/or usefulness of this new technology and adopt later. Adopters can be further subdivided into early and late adopters and similarly for the imitators. Note that social welfare may be maximized in the presence of uncertainty through policies that either augment the information set of consumers early on or provides a subsidy for early adoption, i.e. there is a role for technology policy.

1.2 Technology

The Internet is a vast world-wide network of computers that can communicate with each other using the Transmission Control Protocol/Internet Protocol (TCP/IP). It originated as the ARPANET in the late sixties, developed by the Advanced Research Projects Administration (ARPA), a division of the U.S. Defense Department. It was developed to link together universities and high-tech defense contractors. The ARPANET was subsequently succeeded by the NSFNET in the mid-eighties created by the National Science Foundation or NSF to connect it's supercomputer centers. NSFNET provided the high-speed backbone for the Internet to develop, although currently there are several backbones which are privately operated and the NSFNET has ceased to exist since 1995. The structure of the Internet as it exists now consists of three layers, at the bottom layer are local users connecting to Local Area Networks (LAN). These are connected to local geographical networks which are in turn connected to the high-speed backbone. A small portion of the Internet is the World Wide Web (WWW) which consists of files written in the Hypertext Markup Language (HTML) that can be displayed on any computer via a browser program. This has spurred the adoption of the Internet by consumers since it provides a user-friendly graphical interface.

Internet service at home is typically provided by Internet service providers (henceforth, ISP). In most geographical locations there are a multitude of ISPs. These companies allow the consumer to use her personal computer (PC) / Web TV to connect to the LAN operated by these companies which are in turn connected to the Internet. Until recently the dominant form of connection at home was using the modem built into the PC or Web TV to connect to the ISP using the telephone line. Recently however alternatives have become available that allows the user to connect to the Internet using the cable used for Cable TV programming, and also telephone lines for high-speed Digital Subscriber Line (DSL) connections. Alternatively wireless modems are also available that do not use the phone line. Typically the latter ones provide faster connections to the Internet and cost more, both the modem and the contracts are more expensive.

The consumer generally signs a contract with the ISP for either a fixed number of hours of connection or unlimited usage for a fixed fee per month. Recently a number of ISPs like NetZero have started providing Internet service at home to consumers for free, however these services generally have several unattractive attributes like an advertising banner on the screen that the user cannot turn off as well as higher level of congestion such that web pages take longer to display. Alternative pricing schemes have been suggested. For example Mackie-Mason and Varian (1995) suggests a usage based pricing instead of a flat fee in order to reduce congestion.

1.3 Information Revolution

The Internet is at the epicenter of the so called Information Revolution which was predicted to change the way most markets operate and lead to a boon in productivity and economic growth. Now that the evidence is in, most empirical studies do find positive and significant contributions for the Internet and its associated technologies, in the amazing growth spurt enjoyed by the US economy throughout the last decade. However, the dimensions of such an impact have been estimated to be more modest compared to earlier predictions, for example see Jorgenson (2001) and Litan and Rivlin (2001). For earlier predictions see the *Digital Economy 2000* (NTIA 2000) report published by the Department of Commerce. Apart from past gains few disagree regarding its future potential, it is unlikely that all gains from a networked world has been exhausted. Perhaps most promisingly for most developing countries the Internet holds out great promise through a boost in globalization and free dissemination of information and trade in goods and services. In these senses it is truly a revolutionary technology.

The core contribution of the Internet is to provide a fast and cost-effective method of transmission of information from multiple sources to multiple recipients. Perhaps the dominant use of the Internet currently is for e-mail and transfer of files. Although a large number of firms have also started using the Internet as a retail outlet for goods and services. All transactions in this sector are generally known as E-commerce and it can denote anything from paying bills to shopping. Economists have been interested in this sector since the welfare implications of E-commerce can be quite substantial for the consumer. For a recent survey see Borenstein and Saloner (2001). They point out that:

> 'The Internet creates value by vastly lowering the cost of transferring many types of information, on a one-to-one, one-to-many, or many-to-many basis.'

A priori one expects the Internet to reduce search costs and transaction costs and thereby lead to more efficient functioning of the market. This is particularly true for information goods and goods with low transportation costs relative to their value. If the transportation costs are high then the Internet provides a channel for the consumer to augment her information set about the product. For example one might check the consumer reports online before buying an automobile. On the cost side

E-commerce can reduce costs by lowering distribution costs like warehousing etc., since frequently there are economies of scale involved in inventory management and it can also lower sales costs like maintenance of a showroom and sales force etc. Replacing paper transactions by electronic ones (for example in banking) can lead to substantial savings. On the demand side it can lead to a better matching of buyers and sellers by lowering search costs. However, the consumer needs to know exactly what she wants, since the information that can be transmitted via the Internet is limited, for example one cannot sample the product online and for that one has to go to a brick and mortar store.

The Internet promises to have a profound impact on the labor market through lowering search costs for job searches thereby leading to better matches between employers and employees. Also the location of the worker becomes irrelevant in a wired world. For other issues concerning the labor market see Autor (2001). The Internet has also had a major impact on the financial market for example see Barber and Odean (2001). Perhaps the biggest impact of the Internet is likely to be in terms of the so called business-to-business (B2B) sector with which we are not concerned here. In the retail market lower search costs can lead to increased competition in the market. For example one can search for millions of websites for a particular product using the new shop bot technology in a matter of seconds. One implication of this is that price dispersion should fall in all markets, particularly for goods with low transportation costs. Recent evidence indicates that online prices for goods like books and CDs are indeed lower and price changes are more often which may indicate a more efficient functioning of the market for example see Brynjolfsson and Smith (2000) and Bailey (1998). A number of studies including (Brynjolfsson and Smith 2000) find that online price dispersions without adjustment for market shares is comparable to conventional markets. The sellers have sought to beat this price competition through differentiating the purchase experience by customizing their website for users and other methods.

1.4 Universal Service

It is worthwhile to view the debate surrounding the 'digital divide' as continuation or extension of an earlier debate regarding telephones. Starting with the 1934 Telecom Act, *'universal service'* defined as the easy and affordable access to new communications technologies for all individuals within the United States, has been enshrined as a policy goal for the FCC. This has been achieved primarily through legislation, the latest instance being the Telecommunications Act of 1996. Under this controversial policy a tax on toll calls had been used to subsidize the monthly access price of telephones.[3] Also there exists other targeted federal/state subsidy programs for low income families such as the *LinkUp* program.

These policies have long been controversial since economists apart from debating the legitimacy of such a goal[4] have also long concluded that the price elasticity of access (defined as the percentage change in the penetration rate of telephones in a locality for a unit percentage change in access prices) are extremely low. Therefore any price based policies are not likely to be the best instruments for achieving such a goal. Alternatives include the provision of information, the provision of public telephones etc. Whatever benefit is derived by consumers from such subsidies is far outweighed by the deadweight loss of taxes used to finance such subsidies.

[3] The FCC operates under the restriction that all such subsidies need to be endogenously funded and resources from the general budget are not available to it.

[4] Typical justifications given in the past such as telephones were necessary for access to emergency services, news and information etc. were not viewed as essential by most economists since close substitutes or alternatives usually exists.

1.5 Data

The data for this study was obtained from the Census and is part of the *Current Population Survey (CPS)* conducted by the Bureau of the Census for the Bureau of Labor Statistics (BLS). This data is publicly available online at the BLS website.[5] The CPS has been conducted by the Census for over fifty years and it is a monthly survey of approximately 50,000 US households. The CPS was primarily designed to obtain a snapshot of the U.S. labor market. The data includes information on a variety of demographic characteristics including age, sex, race, marital status, and educational attainment of household members. The labor market data includes detailed information on each household member's occupation, industry, and class of worker etc.

Periodically supplemental questions on a variety of topics are also added to the regular CPS questionnaire. We use data from one of these supplements that the Census calls *Internet and Computer Usage Supplement.*[6] The CPS survey conducted in the following months included this supplement: November 1994, October 1997, December 1998, August 2000 and September 2001. Respondents were asked in addition to the regular questions on demographics and labor market variables whether they use computers at home/work and what are the primary purposes it is used for. Similarly they were asked whether they have an Internet connection at home and if so how do they connect and to what purpose do they use the Internet, for example searching for jobs, reading the news etc. Additionally the 2000 version of the survey also asked people whether they had a high-speed Internet connection (Cable / DSL).[7] The price paid for monthly service is only available for two waves, 1998 and

[5] http://www.bls.census.gov/cps/computer/computer.htm

[6] This data was collected by the Census on behest of the NTIA for their Falling Through the Net series of publications (see above), studying the Digital Divide.

[7] Note that we only use data for cable and DSL connections and these connections are referred

2000 respectively. Therefore these are the only years used for estimation below.

For an overview of the methodology followed in designing the survey refer to *Technical Paper 63RV* (2002).[8] The sampling method used by the Census to construct the sample is *multistage stratified random sampling without replacement.* This is done with the objective of achieving complete coverage of the eligible population. In the first stage certain counties are selected in each state that are representative of other counties in the same state i.e. with similar population characteristics, and primary sampling units (PSU) are defined. Typically PSUs are either a single county or contiguous counties. Then the PSUs are combined to form strata using a clustering algorithm, with one PSU selected from each strata. Metropolitan Areas (MSA) form their own strata within each state provided they are one of the 150 largest MSAs. The second stage involves selection of housing units from each PSU, this is done using the 1990 Decennial Census of Population and Housing and the Building Permits Survey. Since this data is old it is supplemented by the Building Permits Survey which is an ongoing one conducted also by the Census. Finally the survey uses a rotating panel, with each household following a 4-8-4 pattern, with each household interviewed for four months, then rested for eight months and then again interviewed for four months before they are retired permanently. Each month one-eighth of the sample is being interviewed for the first time, another one-eight interviewed for the second time and so on. Because of the rest period only six-eighth of the sample is common between two consecutive periods. This data is publicly available as *use files* from the census website (www.bls.census.gov/cps/datamain.htm).

We also obtained data about the demographic characteristics of the states and metropolitan areas (MSAs) like population, ethnicity of the residents etc. from var-

to as broadband in the text, the CPS definition is substantially broader including wireless, satellite etc.

[8] For additional documentation on methodology refer to the CPS website www.bls.census.gov/cps/

ious publications of the Census. Some of the data is from the recently concluded 2000 Census whereas other variables were obtained from earlier publications like the Economic Census of 1997 etc. The Census used this data to construct very detailed cross-tabulation tables of computer and Internet usage across various demographic characteristics, this information has been published by the National Telecommunications and Information Agency and is available online (see NTIA (2000)).

Additional data on median income and population of MSAs and states were also obtained from the census web site. The median income estimates were obtained from the *Small Area Income and Poverty Estimates (SAIPE)* survey, whereas the population estimates were obtained from Census 2000.

1.6 Descriptive Analysis

The descriptive statistics for the data used for this study is reported in table (1.1) and (1.2). Since adopting the Internet is a household level decision all variables refer to the head of the household as designated by the CPS. Only the characteristics of the 1998 and 2000 sample are reported here, the samples analyzed for the other years were found to be very similar in nature. Apart from age all other variables reported are dummy variables and the sample mean therefore represents the percentage of the overall sample which belongs to this category. These figures roughly correspond to the distribution of these variables obtained separately from the 2000 Census. Sampling weights are almost always used for the estimation, where the weights are defined as the inverse of the probability of selection or alternatively it approximates the actual number of families with similar characteristics in the overall population that this household is meant to represent. The variables used for the three studies reported later are broadly similar although not identical and they are defined later.

Table 1.1:
Descriptive Statistics

Variables	1998	2000
Internet	0.278	0.436
Price	4.836	7.749
Age	47.69	47.67
Male	0.584	0.555
upto $20,000	0.280	0.248
$20,000–35,000	0.229	0.223
$35,000–50,000	0.166	0.157
$50,000–75,000	0.169	0.176
$75,000+	0.156	0.196
No HS/GED	0.171	0.160
HS/GED/Some College	0.533	0.529
College Degree	0.205	0.216
Advanced Degree	0.261	0.272
Black	0.122	0.125
Hispanic	0.089	0.095
Married	0.545	0.544
Single	0.181	0.186
Employed	0.680	0.680
Household size	2.588	2.604
No. of children	0.604	0.597
Rural (Non-MSA)	0.229	0.225
Central City	0.254	0.248
MSA (large)	0.475	0.478
South	0.359	0.363
Computers (≥ 2)	0.107	0.145
Leased	0.006	0.006
Comp. bought 00		0.120
Comp. bought 99		0.157
Comp. bought 98	0.140	0.121
Comp. bought 97	0.108	0.056
Earlier	0.192	0.080
No computer	0.559	0.466
Use outside home	0.200	0.238
Median Income (1000s)	34.673	34.683

Table 1.2:
Descriptive Statistics (cont..)

Variables	1998	2000
MSA sample		
Price	5.278	8.256
Median Income (1000s)	33.038	33.108
% w/ computers	0.467	0.561
% w/ computers ($\geq$ 2)	0.122	0.161
% w/ latest yr. comp.	0.150	0.126
Utilization Variables*		
E-mail	0.712	0.763
Online Courses	0.200	0.190
Search information	0.613	0.601
Phone	0.065	0.058
News	0.517	0.520
Search jobs	0.177	0.200
Job related	0.378	0.356
Shopping	0.296	0.392
Games	0.044	0.047

**Conditional on access.*

Table 1.3 reports the breakup by technology for Internet access. Unfortunately we do not have data on prices for all years but only for the years 1998 and 2000. Starting with the 2000 sample the CPS also had questions on broadband technologies used by the households. The top half of the table gives the breakup between the major technologies that can be used by households to access the net, whereas the bottom half reports the market share of various broadband technologies which might be interesting in its own right.[9] The baseline technology used by most households to

[9] There has been much debate in recent times, particularly in regulatory circles, over the asymmetric regulation of two related broadband technologies, cable and DSL. Local telephone companies are obliged by law to allow (for a charge) the use of their facilities to competing ISPs offering DSL services to the consumer. Whereas there is currently no such requirements for cable service providers despite the fact that most cable franchises enjoy monopoly status in most of their home markets across the country.

access the net remains the dialup modem with prices for access remaining roughly stationary over the period for which we have data (1998–2000). The other alternative touted for people who are reluctant to learn to use the computer simply to access the net is the Web TV, which allows the consumer to check e-mail and generally browse the net by connecting a set top box (similar to cable TV) to their television. We find that although web TV usage increased from 1998 through 2000 it starts to fizzle out by 2001 when population usage fell dramatically from around 1.7% to less than 1%. Broadband technologies have enjoyed significant growth in recent times with their share of the Internet access market growing from around 10% in 2000 to about 18% by the end of 2001. Among the broadband technologies we find that cable which was available earlier had more than fifty percent share of this market and actually expanded its share to over 65% by the end of 2001, with DSL actually losing market share.[10] This is inspite of the fact that cable was more expensive compared to DSL on an average.[11] For completeness we also report the market share of other technologies like cellular and the older ISDN. The newest sample has a somewhat different classification methodology and therefore these figures are omitted in the table.[12]

1.7 Aggregate Diffusion Trends

We start of by reporting simple trends observed in the data for the two main variables of interest for this study, the ownership of computers and access to the Internet. As noted earlier most new innovations have been observed to follow a S-shaped curve of

[10] DSL technology was marred by frequent problems with installation which have been subsequently resolved.

[11] Note that the prices reported are generally lower than what an informal search over the Internet reveals since a lot of consumers had promotional temporary deals which unfortunately we cannot distinguish from the long term regular price paid for service.

[12] In the 2001 sample all other technologies are pooled together in the category *others*.

Table 1.3:
Type of Access

Type of Access	1998		2000		2001
	(%)	Avg. Price	(%)	Avg. Price	(%)
Dialup	24.9	17.4	37.44	16.842	41.62
		(8.46)		(9.33)	
Web TV	1.28	18.04	1.72	20.137	0.62
		(8.58)		(10.47)	
Broadband			4.35	26	9.35
				(15.02)	
a) DSL			32.48	23.83	30.05
				(14.99)	
b) Cable			51.66	29.45	65.35
				(14.87)	
c) Cellular / Satellite			5.06	19.44	
				(11.55)	
d) Other (ISDN)			10.8	19.12	
				(12.51)	

Standard errors in parentheses

diffusion. A simple model generating such a pattern of diffusion is the *logistic growth model* used by Griliches (1957). Let P_{it} be the percentage of the population using the Internet in market i at time t, and let K_i be the ceiling or equilibrium value for this market i.e. the number of final users that we expect will ever use the Internet in this market. This model assumes that ceiling values are stationary and do not change as the technology improves over time. The model can be defined as follows:

$$P_{it} = \frac{K_i}{1 + e^{-(a_{it} + b_{it} t)}} \tag{1.1}$$

where a_i is interpreted as the *origin* of the diffusion process for market i, and b_i is the *slope* of the linearized trend and measures the speed of diffusion in market i. Applying the logistic transformation and adding a normal error term leads to the following linear relationship:

$$log\left(\frac{P_{it}}{K_i - P_{it}}\right) = a_i + b_i t + \epsilon_{it} \qquad (1.2)$$

which can be estimated using ordinary least squares method. Usually the parameters a_{it} and b_{it} are defined as functions of the characteristics of the market and/or technology.

Table 1.4:
Diffusion Process

Year	Computer (%)	Internet (%)
November 1994	24.1	6.1
October 1997	36.6	18.3
December 1998	42.1	26.2
August 2000	51.8	41.9
September 2001	56.6	50.6

Source: Own calculations using CPS data.

For our purposes we are only interested in aggregate diffusion for the whole country. In table 1.4 below we report the aggregate diffusion for the these two technologies, we find that computers which have been available for much longer had ownership of around 24% by the beginning of this study and increased to about 57% by the end (2001). Whereas the Internet which was available to the general public only in the early nineties had a usage level of about 6% by the beginning of the study and increased to about 51% by the end of this period. This data is then used to fit the logistic growth model described above. A significant advantage of aggregate diffusion models is that it has been observed to fit the data extremely well for various new goods, despite its parsimonious representation and limited behavioral basis. The observed trends are reported in figures 1.2 (a)-(b) below. Figure 1.2 (b) fits the logistic model for the trend in Internet usage and similarly figure 1.2 (a) does so for computer ownership. These models are estimated in two steps, first the ceiling value or maximum usage for the country estimated by maximizing the fit (R^2) via a grid search.[13] The maximum in both cases is uniquely defined and in the second step we use this value of K to construct the dependent variable and estimate a and b respectively.

Another interesting feature obtained as a byproduct of this analysis is that we can find the maximum usage levels for both of these technologies. We find that the maximum level of usage for the Internet to be around 84% at its peak whereas the PC reaches universal adoption i.e. ceiling value $K = 100$. Needless to say these estimates need to be taken with caution since it has been observed in numerous cases that accuracy of such forecasts increases with more data and also with a higher level of current adoption i.e. later stages of the diffusion process.

1.8 Studies

Diffusion is a temporal, social and spatial phenomenon, information regarding new technologies spreads through various communication channels through the economy or social system over time. The first study reported in the next chapter deals with the temporal aspect of it, i.e. it traces diffusion patterns for the PC and the Internet for the United States, across multiple dimensions such as income, education, race etc. This addresses the concern regarding the 'digital divide', i.e. using the techniques outlined there we can trace the gap in penetration rates across these dimension, for example between different races over time. This allows us to answer the fundamental

[13] We search for K_i over the following interval: [usage in 2001 + 5% , 100%]. We divide it into a fine grid and for each potential value of K_i we estimate the OLS regression of equation 1.2 above.

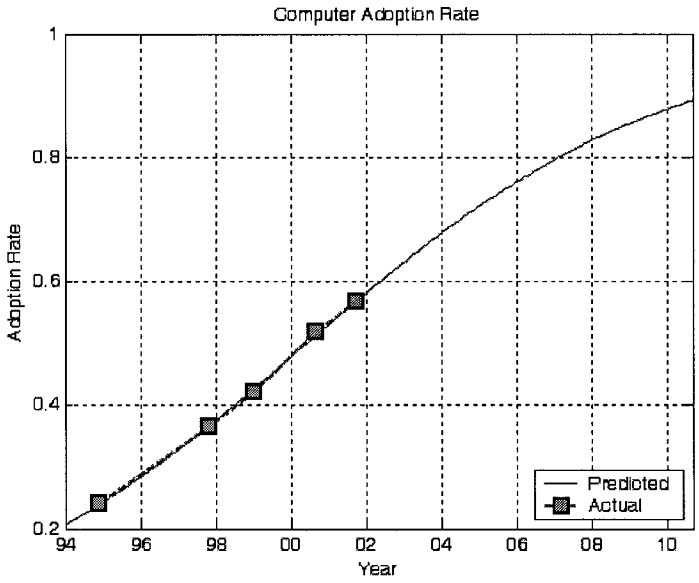

Figure 1.1: Ownership Trend for Computers

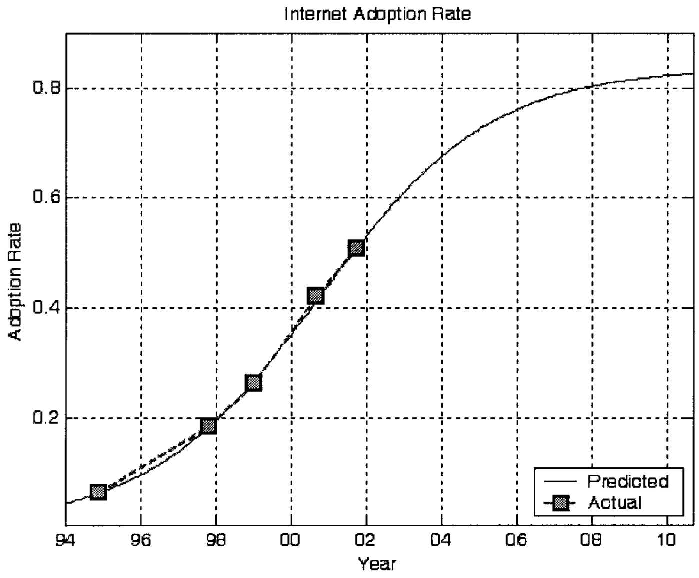

Figure 1.2: Ownership Trend for Internet

question in this debate of whether there is a role for the government in promoting Internet access. The next chapter investigates the role played by social networks or so called 'neighborhood effects' in the diffusion of new information technologies. This has potentially far-reaching implications regarding how technology policies need to be designed in order to be most effective. For instance with strong social networks facilitating diffusion a policy based on selective tutoring might be more effective compared to price based policies such as subsidies etc.

At the height of the 'digital divide' debate the government initiated a number of policies to promote Internet access across the population. These policies were very much in line with earlier policies undertaken to promote universal service for telephones such as subsidies for basic access, provision of phones in public places etc. Such policies have been controversial in the context of telephones earlier on since it was believed that the price elasticity of access was so low that subsidies were likely to be costly and ineffective. Some authors have raised a similar concern regarding the policies surrounding these new information technologies as well. Unfortunately there exists few studies in this fields and the ones that do exist suffer from certain serious flaws as shown in the final chapter of this volume. We estimate a search model for Internet adoption at home and using it are able to calculate the price elasticity of access. It also allows us to draw certain inferences regarding the size of the total consumer surplus resulting from these new technologies.

Chapter 2

Digital Divide: Myth or Reality

The digital divide is broadly defined as the concern that certain groups in the population might not have access to information technology and therefore be somehow handicapped in their lives (for example they will have fewer employment opportunities in the future in an increasingly wired job market etc.). In many ways these concerns are a continuation of long-standing policy goals in the United States, Canada as well as in many other OECD countries of *universal service*. This is defined by the 1934 Telecommunications Act as follows:

> "...to make available, so far as possible, to all people of the United States a rapid efficient nation-wide and worldwide wire and radio communications service with adequate facilities at reasonable charges."

In practice this definition as well as the policies advocated for its achievement has evolved over time.[1] In its current form, the Telecom Act of 1996 extends this concept to the provision of new, high-speed telecom services to public institutions such as libraries, schools and medical institutions. Also known as the *E-rate* program it

[1] In order to achieve these goals, in the past explicit subsidies such as the Linkup and Lifeline programs has been undertaken by the FCC, as well as implicit subsidies provided for local telephone rates.

uses revenue generated by taxes on long distance calls to subsidize Internet access for these institutions (about 2 billion dollars). Additionally Hausman (1998) finds that these programs cost about 2.25 billion dollars to administer every year. The general public also enjoys implicit subsidies from the *Internet Tax Freedom Act* of 1998 which initially placed a three year moratorium on all taxes on Internet access and has since then been extended for an additional three years.[2]

Many authors have pointed out that in the context of 'universal service' such policies fly in the face of economic logic, since most developed countries have telephone penetration rates of over 90%. Also most econometric studies report a very low price elasticity of access, for residential demand for telephones (see Crandall and Alleman (2002) and the studies cited there). Similarly for the Internet it is hard to justify the policies undertaken and the large efficiency costs generated therein, since Kridel, Rappoport, and Taylor (1999a) reports similar findings for Internet access. Earlier studies such as Beckert (2000) also report a low price elasticity of demand for bandwidth.[3] The evidence provided in favor of the digital divide is flimsy at best (see Compaine (2001)). The series of studies done at the behest of the Clinton administration[4] provides limited information in this regard since the methodology used there is primarily descriptive (cross-tabulations etc.) and static in nature. Similarly other scientific studies such as Hoffman, Novak, and Schlosser (2001) which uses a simple static discrete choice framework to test for a digital divide across various races is not entirely satisfactory.

Melnikov (2000) and Gandal, Kende, and Rob (2000) show that in the context of new technologies static models of discrete choice are inadequate and a dynamic model

[2] Set to expire this November the House of Representatives voted overwhelmingly in favor of extending this bill, on Sept. 17, 2003.

[3] Elasticities were estimated using data generated from a controlled experiment conducted on the Berkeley campus, also known as the INDEX project.

[4] Reported in the Falling Through the Net series of publications, available online at the FCC website as well as summarized in Compaine (2001).

with foresight is required. They show that for new technologies rapid improvements in quality introduces a potential bias into the estimates obtained from static discrete choice models, due to intertemporal substitution. Low levels of adoption might simply be due to the option value of waiting as new and better technologies come along (and/or prices fall) and not because of the digital divide. Melnikov (2000) also shows that static estimates of the value attached to quality implies an exploding sales pattern over time as quality improves exponentially, however in reality for most new goods such a pattern is not observed in real data. Therefore they stress the need for estimating forward-looking models of consumer behavior where future benefits from improvements in technology are endogenized.

In this paper I present such a model of technology adoption and show that it can generate the patterns observed in the real world. I go on to show that the salient features of the model can be adequately summarized under certain circumstances by parametric duration models. I use publicly available data from the *Current Population Survey (CPS* to estimate these models. The data is in the form of repeated cross-sections, and naturally the question arises whether dynamic duration models can be estimated using such data. I provide evidence in the affirmative, using Monte Carlo simulations I show that the maximum likelihood estimator in this context is consistent and even efficient for certain models. Therefore this paper should have two contributions, first it estimates a dynamic model of technology adoption with an application to the Internet, therefore we can test for the existence of the digital divide as well as forecast its dimensions in the short to medium run if it exists, this has serious policy implications as discussed before. Additionally this study shows that a broad class of diffusion models can be estimated using repeated cross-sectional data, which opens up new potential sources of data for any field where such models are currently used. Also Besley and Case (1993) claims that in the context of adoption of new technology, since self-reported adoption times tend to be notoriously unreliable i.e. have very high measurement errors, discrete choice data of current usage might provide better estimates.

The rest of the paper is laid out as follows, in section 2 we discuss related studies, following which in section 3 we present a simple model of technology adoption. Section 4 introduces the most commonly used duration models and section 5 discusses estimation strategies using RCS data. Section 6 presents monte carlo evidence regarding the MLE and in section 7 we present our main results and finally section 8 concludes.

2.1 Related Work

This study draws its inspiration from several sources: the marketing literature on new product diffusion, in economics the literature on diffusion of new technology and also the literature in sociology on diffusion of innovations and learning in social networks and lastly the statistical literature on survival analysis.

Schumpeter called diffusion the third pillar of technical progress along with invention and innovation. There exists a very large literature in economics on the diffusion of new technology, studying for the most part adoption decisions by firms, for various process innovations.[5] This literature is too large and diverse to be adequately summarized here, the reader is instead referred to the excellent surveys by Geroski (2000) and more recently Hall and Khan (2003). The empirical literature is usually dated to have originated with the seminal contribution by Griliches (1957), studying adoption decisions of farmers, for new varieties of corn seeds. Gruber and Verboven (2001) applies a similar methodology to estimate the diffusion of mobile telephones in

[5] Process innovations are defined as improvements in the production process as opposed to product innovations which are improvements in a final good or service.

the European Union. Numerous issues have been considered in this context both on the demand side (adopter side) such as firm size, market concentration etc.,[6] as well as the supply side (technology and supply features) such as improvements in quality, uncertainty in utility, seller concentration etc. Many of the insights developed in this literature do not translate to this case since this study focuses on consumer adoption decisions.

This study is instead closer in spirit to the marketing literature on the diffusion of new goods, since it models adoption decisions by households. The workhorse model in this context is the Bass (1969) model, which has been remarkably successful over time in predicting diffusion patterns for numerous goods. For an excellent survey of this literature refer to Roberts and Lattin (2000) and Mahajan, Muller, and Bass (1991). A useful way to classify the various models is by their levels of aggregation, for instance the models considered by Griliches and Bass study diffusion at the market level. These models have found wide applicability across numerous studies particularly due to their parsimonious representation of the diffusion process, usually summarized by a few variables which are then related to the characteristics of the new technology or the adopter. Only market level data is required for estimation. The literature on diffusion of innovations in sociology[7] (see for example Rogers (1995))is conceptually close to the marketing literature, the diffusion process is explained by an *epidemic model* of learning by consumers.[8] Note that for most new products / technologies an S-shaped market level adoption curve is observed i.e. initially adoption proceeds slowly but accelerates over time, all models mentioned in this

[6] Schumpeter also hypothesized that market power should accelerate diffusion; others have pointed out theoretical reasons against it. Numerous empirical studies in this field therefore seek to estimate the relationship between firm size and adoption decisions.

[7] Diffusion is defined more broadly in this context as any new social behavior.

[8] Information about the new product spreads like an epidemic through contact between an infected person (current user) and an uninfected (uninformed) person. Therefore a larger infected population leads to faster adoption.

study generate such a curve.

Alternatively, a separate class of models stresses consumer heterogeneity as the driving force behind the diffusion process. These disaggregate consumer level models are usually more intuitive since they have a basis in consumer utility theory. However such models require consumer level micro data for their estimation and for forecasting purposes as well, thereby limiting their use. It is argued that consumers are heterogeneous in terms of their utility for the new product and therefore have diverse reservation values, which in turn leads to staggered adoption dates, i.e. a diffusion curve at the market level (Davies (1979) first considered such models). It is common to use duration models to estimate these models, see for example Hannan and McDowell (1984), Rose and Joskow (1990) and Berndt, Pindyck, and Azoulay (2000). Note that in this context although heterogeneity can be explicitly tested for, consumer learning or network effects are not identified separately. A third category of models explicitly specifies and estimates consumer learning, see for example Chatterjee and Eliashberg (1990) and Erdem and Keane (1996).[9] Policies to accelerate adoption rates for new technologies are considered by Stoneman and David (1986).

Lastly for a current review of survival analysis from an econometrics perspective refer to Berg (2000). Duration models have been used widely in the economics literature to study diverse phenomenon such as government program impact on unemployment spells (Meyer (1990)), criminal recidivism (Schmidt and Witte (1989)), runs on banks (Henebry (1996)) and currency crises (Glick and Rose (1999)).

[9] A dynamic programming model with Bayesian learning process is usually assumed.

2.2 A Simple Structural Model

In this section we present a stylized model of technology adoption. This model is a modified version of the model introduced by Cameron and Heckman (1998) [*henceforth CH*], which studies the impact of family background variables on schooling decisions for five cohorts of American men. Others such as Davies (1979) had studied similar models in the technology diffusion literature, Geroski (2000) calls this broad class of models *probit models.* Note that conceptually the decision to terminate further education is very similar to adoption of new goods and/or technology. Therefore many of the insights derived by the authors in the context of schooling are relevant to individual adoption decisions as well. We first report several critical features derived by the authors which serves as a cautionary tale for the diffusion literature.

Numerous authors have estimated logit and probit models with cross-sectional data on adoptions. In particular a number of authors studying the digital divide had used such tools (see discussion above). With multiple cross-sections or with a single cross-section and recall data[10] earlier authors such as Goolsbee and Klenow (2002) had estimated period specific adoption probabilities over time. Let D_t be a dummy variable denoting usage/adoption at time t, then the probability of adoption in period t conditional on not having adopted by period $t-1$ and given a set of time-varying covariates $\mathbf{X}_t$, i.e.

$$\Pr(D_t = 1|\mathbf{X_t} = x_t, \quad D_{t-1} = 1) = P_{t,t-1}(x_t) \tag{2.1}$$

[10] Self-reported past date of adoption.

is usually parameterized as a standard logit or probit model as follows:

$$\begin{aligned} Logit \quad P_{t,t-1}(x_t) &= \frac{exp(x_t'\beta_t)}{1+exp(x_t'\beta_t)} \quad or, \\ Probit \quad P_{t,t-1}(x_t) &= \Phi(x_t'\beta_t) \end{aligned}$$

These models which formulate the consumer's decision as a static problem are fundamentally flawed since adoption decisions typically are intertemporal (since improvements in quality are enodgenized) and therefore valuation/beliefs depend on the whole history of past shocks in more complicated ways than captured by a simple logit/probit formulation. CH show that behavioral models that can generate such behavior implicitly makes strong assumptions such as myopic consumers and/or a martingale process for the period specific shock to valuation. Additionally they show that,

- In the presence of omitted variables / unobserved heterogeneity, dynamic selection over time makes the coefficients biased in ambiguous ways,[11] making the coefficients harder to interpret.
- Theorem 4-5 in the CH study show that in the presence of unobserved heterogeneity and if both **X** and β is the same for all transitions then the model is non-parametrically unidentified and depends upon strong distributional assumptions for identification.

Next we consider a modified version of the simple behavioral model presented in CH. It consists of forward-looking, profit maximizing, heterogeneous individuals maximizing the discounted present value of consumption. The adoption decision can be framed in terms of an optimal stopping problem. There is a return as well as

[11] The sign of the bias depends critically on distributional assumptions about the unobs. heterogeneity term.

a cost associated with postponing adoption, the return in this case comes from a downward trend in hedonic prices (price adjusted for quality improvements[12]) that is almost always observed for all new technologies. The cost is in the form of forgone benefits of consumption in the current period.

Formally given individual characteristics $\mathbf{X} = x$, let the cost from waiting be $C(t|x)$. We assume that this is weakly convex and increasing in waiting time t. As long as per period benefits are strictly positive, total cost will increase over time. The convexity assumption says that benefits (usually) rise more than proportionately over time; this is not unreasonable for most new technologies and is particularly appropriate for the Internet given the explosive nature of its growth in the recent past. The Internet is a strong network good, with quality directly proportional to the number of users, therefore as the number of users increases so does the number of potential correspondents for e-mail, chat etc., as well as websites / sources of information. Therefore per period foregone benefit from consumption can be assumed to rise at least initially. Also assume that $c(0|x) = 0$ for all x, which is not unreasonable for pure network goods as it is worthless with no other users.[13] Assume that the returns function $R(t)$ is strictly concave (at least upto a point $\bar{t}$) and weakly increasing in t. This is justified if quality increases or price decreases with certainty early in the life of all new technologies, but this peters out over time. Also assume that $R(0) > 0$, which says that everyone knows with certainty that quality will improve. Without loss of generality we assume the R function is the same for everyone since all individual specific differences can be absorbed in the cost function. Notice that subjective discount factors are embedded in both the returns and costs functions.

[12] Improvements in characteristics such as reliability, lowering of uncertainty in benefits through consumer learning etc.

[13] This is a simplifying assumption that ensures an interior solution for the consumer's problem, no loss of generality results since adoption can happen at $t + \epsilon$ with $\epsilon \rightarrow 0$ in the limit.

Optimal adoption time is then the solution to the following maximization problem:

$$\max_{t} R(t) - C(t|x), \quad t \in [0, \infty) \tag{2.2}$$

Given our assumptions about the shape of the returns and cost functions this function is well behaved and concave with a unique maximum which is positive since $R(0) > 0$ and $C(0|x) = 0$. This model retains the essential feature of earlier diffusion models with heterogeneity since any factor that increases benefits or raises the cost of waiting necessarily lowers reservation values leading to an earlier adoption times. For simplicity we also assume the following:

Assumption 1 *The cost function is multiplicatively separable, i.e.* $C(t|x) = c(t)\kappa(x)$.

Assumption 2 *The individual effect can be decomposed into observed and unobserved components, i.e.* $\kappa(x) = \lambda(x)\epsilon$ *where* ϵ *is unobserved factors.*

Assumption 3 *The unobserved factors are independent of* $\boldsymbol{X}$*, and distributed as follows:* $E(\epsilon) = 1$*. Also we assume that cost is non-negative i.e.* $\epsilon, \lambda(x) \geq 0$.

Here unobserved factors represent all omitted variables that influence the adoption decision observed by the individual but not by the analyst.[14] Later on we will assume a random effects model where the unobserved factor could be interpreted as unobserved ability or technological sophistication.

Example 1: Let the return curve be a quadratic of time $R(t) = at - bt^2$ for $t \leq a/2b$ and $R(t) = a^2/4b$ for $t > a/2b$, with suitable $a, b > 0$, this curve is concave and increasing in t upto $a/2b$. Also let the cost curve be $C(t|x) = ct\lambda(x)\epsilon$ and $c > 0$,

[14] If we write $\kappa(x) = exp(x'\beta)$ then let x_o be observed variables and x_u be unobserved and correspondingly β_o and β_u their coefficients, then $\kappa(x) = exp(x_o'\beta_o + x_u'\beta_u) = \lambda(x)\epsilon$ where $\lambda(x) = exp(x_o'\beta_o)$ and $\epsilon = exp(x_u'\beta_u)$.

this is weakly convex and also increasing over time. Then the first order conditions from problem (2.2) implies the following:

$$\begin{aligned} R'(t^*) &= C'(t^*|x) \quad or, \\ a - bt^* &= c\lambda(x)\epsilon \quad or, \\ t^* &= \frac{a - c\lambda(x)\epsilon}{2b} \end{aligned} \tag{2.3}$$

If we assume that unobserved factor ϵ is distributed as normal with unit mean and variance σ^2, then the probability of adoption by time T can be written as:

$$\begin{aligned} \Pr\left(t^* \leq T|x\right) &= \Pr\left(\frac{a - c\lambda(x)\epsilon}{2b} \leq T\right) \\ &= \Pr\left(\frac{a - 2bT}{c\lambda(x)} \leq \epsilon\right) \\ &= 1 - \mathbf{\Phi}\left((1/\sigma^2)\left\{\frac{a - 2bT}{c\lambda(x)} - 1\right\}\right) \end{aligned} \tag{2.4}$$

where $\mathbf{\Phi}$ is the cumulative distribution of the standard normal variate. For simplicity of notation using the fact that both the return and cost curves are at least weakly increasing and therefore $R', C' \geq 0$, define the function:

$$exp[\rho(t)] = R'(t)/c'(t) \quad t \in [0, \bar{t})$$

Then by the definition of $c(t) = C(t|x)/(\lambda(x)\epsilon)$ and using the assumptions made earlier regarding the shape of the curves, i.e.

$$R'(t) > 0, R''(t) < 0 \quad and \quad C'(t) > 0, C''(t) \geq 0 \Rightarrow c'(t) > 0, c''(t) \geq 0$$

we can show that $\rho(t)$ is a monotonous and therefore invertible function of t.[15] Specifically we can show that

$$\frac{d\rho(t)}{dt} = \frac{d}{dt}[\log(R'/c')] = \frac{c'^2}{R'}\left(\frac{R''}{c''} - c'\right) \tag{2.5}$$

by the concavity of R and the weak convexity of c the first term is negative and since both are increasing functions of time the second term is as well, which implies $\rho'(t) < 0$. Therefore given ϵ, the optimal stopping time using this notation is:

$$t^* = \rho^{-1}\{\log(\lambda(x)) + \log(\epsilon)\} \quad = \rho^{-1}\{x'\beta + \log(\epsilon)\}$$

Then we can write the probability of failure by time T as:

$$\Pr(t \leq T|\mathbf{X} = x) = \Pr\left[\frac{exp(\rho(T))}{\lambda(x)} \leq \epsilon\right]$$

Letting $\lambda(x) = exp(-x'\beta)$ as before we see that:

$$\Pr(t \leq T|\mathbf{X} = x) = \Pr\left(\rho(T) + x'\beta) \leq \log(\epsilon)\right) \tag{2.6}$$

If we assume that $\log \epsilon$ is distributed with pdf $g(\log \epsilon)$ then we can derive the distribution of adoption times as follows; in particular if we assume that it is distributed normal with mean zero[16] and variance $\sigma^2_{\log \epsilon}$, then we can show that this gives us the standard probit model (see below). When the data is interval censored i.e. adoption is known to have occurred within an interval of time (discrete case), this assumption leads to an ordered probit model (this is the model used by CH). Note that alterna-

[15] Under the assumption of weak concavity of $R(t)$ one can show that equilibrium adoption time is always less than $\bar{t}$, given a weakly convex $c(t)$ function, therefore $\rho(t)$ is monotonic in the relevant range.

[16] This follows from assumption 3 above that $E(\epsilon) = 0$.

tive parametric models of discrete choice can be derived using different assumptions about the distribution of $\log \epsilon$.

$$\begin{aligned} \Pr(t \leq T | \mathbf{X} = x) &= \int_{\rho(T)+x'\beta}^{\infty} g(\log \epsilon) d \log \epsilon \\ &= 1 - \mathbf{\Phi}\left(\frac{\rho(T) + x'\beta}{\sigma_{\log \epsilon}}\right) \end{aligned} \tag{2.7}$$

Manski (1988) shows that such models are identified upto affine transformations, which implies the need for the assumptions, $E(\log \epsilon) = 0$ which fixes the location and $\sigma^2_{\log \epsilon} = 1$ to normalize the scale.

2.2.1 Uncertainty

In this model we assume either that consumers know about benefits and costs with certainty or the uncertainty about benefits and/or costs remain constant over time, i.e. it is endogenized into the original decision process. If individuals receive new information every period then they face a new optimization problem every period. Estimating such a structural model of consumer learning with uncertainty as considered by Chatterjee and Eliashberg (1990) and Erdem and Keane (1996), requires fairly detailed information about the consumer and/or strong assumptions need to be made about the information updating process. Given the nature of the data we use it is well beyond the scope of this paper.

In the next section we present a simple stochastic model of consumer learning, where consumer valuations follow a random walk with drift. Compared to the Bayesian learning models used by the other studies this specification has the added advantage of having a simple closed form solution.

2.3 Duration Models

For simplicity we consider only continuous time models here, since it makes derivations of various functions considerably easier and can be extended to a discrete setup with minor modifications. Duration models are defined either in terms of a hazard rate or equivalently using an underlying distribution of time to adoption. Let us define time to adoption T as a random variable with cumulative distribution function $F(t)$ and distribution function denoted by $f(t)$. Then the *hazard rate* is defined as the probability of failure in the interval Δt conditional on survival until time t i.e.

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t} \tag{2.8}$$

and by definition this can be shown to be,

$$h(t) = \frac{f(t)}{S(t)} \quad \textit{where} \quad S(t) = 1 - F(t) \tag{2.9}$$

The function $S(t)$ is sufficiently important in our analysis that it is worth defining separately. Typically referred to as the *survivor function* it refers to the proportion of the total population that has not failed yet at time t, or

$$S(t) = \Pr(T \geq t) \tag{2.10}$$

2.3.1 Behavioral Model

First we consider what kind of a duration model is implied by the model of technology adoption presented in the last section. Note that equation (2.7) above defines the distribution of adoption times t given the distribution of the unobserved heterogeneity term $\log \epsilon$ i.e. $g(\log \epsilon)$. Using the definition of the survivor function, the relation in (2.9) and using the Leibniz rule if we differentiate equation (2.7) we find

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{F'(t)}{1 - F(t)} = -\rho'(t)\left\{\frac{g(\rho(t) + x'\beta)}{1 - F(t)}\right\} \tag{2.11}$$

We consider two examples of standard symmetric distributions for the heterogeneity terms and show they lead to standard duration models considered below. First assume that $\log \epsilon$ is distributed as a logistic distribution, i.e.

$$G(\log \epsilon) = \frac{exp\{\log \epsilon\}}{1 + exp\{\log \epsilon\}} \qquad g(\log \epsilon) = \frac{exp(\log \epsilon)}{\{1 + exp(\log \epsilon)\}^2}$$

Using this in (2.11) gives us the following hazard rate:

$$h(t) = G(\log \epsilon) = -\rho'(t)[1 + exp\{\rho(t) + x'\beta\}]^{-1} \tag{2.12}$$

We know that $\rho'(t) < 0$ (from (2.5) above), therefore if we assume $exp[\rho(t)] = t^\alpha$ where $\alpha < 0$. We can write the hazard rate as follows:

$$h(t) = \frac{(-\alpha)t^{\alpha-1}exp(x'\beta)}{1 + exp(x'\beta)t^\alpha)} \tag{2.13}$$

which is simply the hazard rate for the log-logistic model (see table 2.1 below).

Also if we assume that $\alpha = -1$ then $\rho(t) = -\log t$, then by using the definition of the survivor function from (2.9) and (2.11) we can write

$$S(t) = 1 - F(t) = \Phi\left(\frac{-\log t + x'\beta}{\sigma_{\log \epsilon}}\right) \tag{2.14}$$

note that this survivor function is identical to the lognormal model where $\log t$ is distributed as

$$\log t \sim N(x'\beta, \sigma^2_{\log t}) \quad where \quad \sigma_{\log t} = \sigma_{\log \epsilon}$$

as defined below in table 2.1.

Example 2: A simulation exercise was performed to study the dynamic individual and market level behavior predicted by this model. We drew one hundred samples each with $N = 5000$ data points, we randomly generated a single covariate $x \sim N(0, 0.25)$ and a constant with parameters $\beta_0 = 1, \beta_1 = 1$. The return function was taken to be $R(t) = exp[(1-t)/10]$ which is increasing and concave in t, the cost function was taken as $c(t) = 1 - exp[(t-50)/10]$, with $t \in [0, 50]$.[17] The resulting failure time distribution is symmetric about the mean and approximately normal as expected. We plot the simulated hazard rates (averaged over all samples) in figure (2.1). We find that the hazard is non-monotonic, increasing initially and then declining. Most data obtained from real studies also follow a similar pattern.[18]

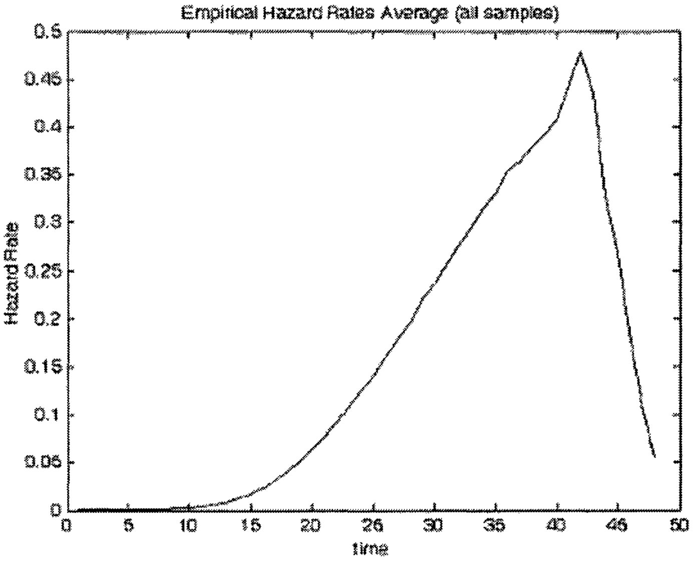

Figure 2.1: Simulated Empirical Hazard

[17] Which implies that it is increasing and convex as assumed earlier.

[18] For examples see Lancaster (1990)

2.3.2 Parametric Specifications

The various models discussed here vary in terms of two important characteristics. First the behavior of the underlying hazard rate over time with no covariates or constant covariates, i.e. given the characteristics of the person and given that he has not adopted by time t, is he more or less likely to adopt as t increases. This is known as *duration dependence* and it may be positive, negative or constant depending on whether the underlying hazard increases, declines or stays constant over time. The second crucial difference lies in the modeling of consumer heterogeneity, depending on specification while some models allow the covariates to affect only the location of the distribution other models allow the location, scale and shape of the distribution to change with the covariates. For an extensive survey of the various models refer to Berg (2000) and Lancaster (1990).

The most commonly used models in this literature are summarized in table 2.1. Perhaps the most widely used model is the Weibull partly due to its simplicity, how-

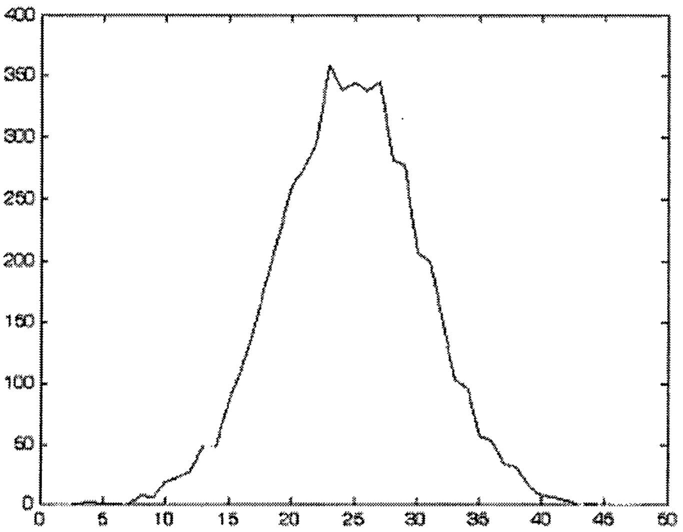

Figure 2.2: Simulated Distribution

Model	Hazard Rate	Survivor Function	Shape parameter	Other pars.
Weibull	$\alpha\lambda^{\alpha}t^{\alpha-1}$	$\exp\{-(\lambda t)^{\alpha}\}$	α	$\lambda = \exp\{-X'\beta\}$
			Monotonic	
Lognormal	$\frac{\phi(y)}{\sigma t(1-\Phi(y))}$	$1 - \Phi(y)$	μ, σ	$\log(T) \sim N(\mu, \sigma^2)$
			Non-monotonic	$\mu = \exp(X'\beta)$
Log-logistic	$\frac{\lambda\alpha t^{\alpha-1}}{1+\lambda t^{\alpha}}$	$\frac{1}{1+\lambda t^{\alpha}}$	λ, α	$\lambda = \exp(X'\beta)$
			Non-monotonic	
Prop.	$g(X)h_0(t)$	No closed	$h_0(t)$	$g(X) = exp(X'\beta)$
Hazard		form	Flexible	
Cont.	$\nu \exp(X'\beta)h_0(t)$	No closed	$h_0(t)$	$g(X) = exp(X'\beta)$
Mixture		form	Flexible	$\nu_i \sim f(\nu; \eta)$

Table 2.1: Various Duration Models

ever it has a monotonic underlying hazard rate which may or may not be appropriate in this context. If consumers learn about the new technology then one expects the baseline hazard to increase ($\alpha > 1$). On the other hand there is also dynamic selection bias, i.e. the population left behind each period might have a lower average ability (or any other unobserved variables not included in λ), leading to a declining hazard over time ($\alpha < 1$). Thirdly the hazard rate can be constant over time (if the adoption process is entirely random). More realistic non-monotonic hazards are provided by the other two parametric models reported in table 2.1. The lognormal model also widely used has an inverted U-shaped hazard with initially increasing and then declining hazard rates, which is more commonly observed in real world data. It has a single maxima depending on $X'\beta$, also it can be homoscedastic or heteroscedastic $\sigma = \sigma(X)$. Also note that in the Weibull the covariates act as a scale factor increasing or decreasing the hazard proportionately for all t which is not very flexible, whereas in the lognormal and the log-logistic model the location and the shape of the hazard rate depends on the covariates. An additional advantage of the log-logistic model is that there exists closed form expressions for both the hazard and the survivor function. The log-logistic model in non-monotonic only if $\alpha \geq 1$.

Apart from these parametric models numerous studies have estimated a proportional hazards model which assumes that the baseline hazard rate and the covariates that affect the hazard rate are multiplicatively separable. This model is flexible enough to include all types of duration dependence since the underlying hazard $h_0(t)$ is estimated non-parametrically. However it has the shortcoming that non-parametric estimation depends heavily on the multiplicative separability of the baseline hazard which might be a strong assumption in certain contexts. An extension of the proportional hazards model is the so called *mixture* models, which assumes an unobserved heterogeneity term ν also enters the hazard rate multiplicatively. The distribution of ν can be assumed to be either discrete (finite mixtures) or continuous.

Among discrete mixing distributions the most popular choice is the binomial distribution, and similarly the gamma distribution for the continuous case given that it has the unique advantage of being sufficiently flexible and also the likelihood function has a closed form solution. An alternative approach suggested by Heckman and Singer (1984) assumes a discrete distribution and maximizes over the number of points of support. Also known as the *NPMLE* (non-parametric MLE) although conceptually attractive in reality we found just as numerous authors before that it has frequent convergence problems. Elbers and Ridder (1982) shows that such mixture models are identified under fairly mild conditions.

2.3.3 A Stochastic Learning Model

In this section we show that a stochastic learning model (for example match models considered by Jovanovic (1979)[19]) can lead to hazard rates that are very similar to parametric models considered above. We assume that individuals each period observe a noisy signal regarding the value of the new technology (specific to them). Using this signal they update their beliefs every period using Bayes Rule. Then one can show that this stochastic process of consumer valuations follows a simple *random walk with drift.*[20] This process in continuous time is also known as the *Wiener* process. Let $z(t)$ denote consumer valuation which is updated as new information comes in every period (instantaneously in a continuous time setup). The Wiener process can be written as follows:

$$dz(t) = \mu dt + \eta(t)\sigma\sqrt{dt} \tag{2.15}$$

[19] He used a similar setup to estimate optimal tenure in a job search model with match specific uncertainty and learning over time.

[20] For proof see Jovanovic (1979). Melnikov (2000) considers similar processes for their simplicity.

where $\eta(t)$ are independently distributed standard normal shocks i.e. $\eta(t) \sim N(0,1)$. Without loss of generality assume $z(0) = 0$ or that consumers have no information on the new technology (at the instant) when it is launched all advertising and promotional activities take place after product launch. In that case increments in $z(t)$ are independent normal variates with mean μt and variance $\sigma^2 t$ respectively.

If reservation values for adoption are either fixed $\alpha(x)$ or declining[21] $\alpha(x) - \gamma(x)t$,[22] using a standard result on the Wiener process we can show that time to adoption is distributed as a duration model also known as the Inverse Gaussian distribution which is:[23]

$$f(t) = \frac{\alpha}{\sigma t^{3/2}} \phi \left(\frac{\alpha - \mu t}{\sigma \sqrt{t}} \right) \quad \forall\, t \geq 0 \tag{2.16}$$

where $\phi(y)$ is the pdf of the standard normal. This model is very similar to other duration models presented before in terms of hazard rates, therefore we do not estimate this separately. We also take it as added justification for the duration models fitted to data.

Note that in the last section we derived a duration model starting with a simple behavioral model with no idiosyncratic shocks to consumer beliefs. Whereas in this section we showed that a duration model may also be the result of consumer learning with heterogeneity. Is there any way to differentiate the two? This question has been considered by numerous other authors as well. Unfortunately in our instance we could not find a consistent method to econometrically identify and test this hypothesis. In a separate paper Sarkar (2003) tests this hypothesis using a different approach, by identifying each individual's potential network of contacts and finds strong evidence in support of it.

[21] As prices fall or quality improves.

[22] Formally the two cases are very similar since the time trend γt can be absorbed into the drift term μ.

[23] For example see Lancaster (1990) for proof.

2.4 Estimation

2.4.1 Extreme Censoring

Frequently in the real world the data available to the investigator is right censored (observed data is $\min(T, C)$ with censoring at time C) and the method for controlling for this in the estimation process is well documented, for example see Lancaster (1990). This is usually achieved by rewriting the log-likelihood function to incorporate censoring. In the real world of course other more complicated forms of censoring is sometimes observed in the data. One of them is *interval-censoring* that arises routinely in biostatistics, for example the onset of disease can only be known to have occurred between two test dates which might be sufficiently far apart. A number of authors have estimated statistical failure models that take this kind of censoring into consideration. This type of censoring is usually referred to as "case 2" interval censored data in the literature. Huang and Rossini (1997) considers the asymptotic properties of MLE estimates of semi-parametric models with this kind of censored data.

However the data we have is in the form of repeated cross-sections (henceforth RCS), since the survey was conducted over several waves. RCS data can be considered to be an extreme form of interval-censoring, the only information available from the sample is that failure or transition occurred before t_j when the jth wave of the sample was collected. It is closer to "case 1" interval censoring, in this case what is observed is

$$(t_j, \delta, X) \in \mathbf{R}^+ \times \{0, 1\} \times \mathbf{R}^d$$

where $\delta = 1_{\{T \leq t_j\}}$ indicating whether T has occurred or not by time t_j. Other situations where such data arise naturally are animal tumorigenicity experiments the existence of tumors can be verified only at natural death or sacrificing the animal which is done at different times, see Finkelstein (1986).

In this context, Huang (1996) shows that the MLE estimates of a semi-parametric model is asymptotically efficient. RCS is very similar to "case 1" censoring since each observation is interval-censored over the interval $(-\infty, t_j]$ for the j wave of the survey. The only difference being that in most survey data this interval is the same for all observations collected in each wave. Intuitively this interval is less informative than knowing that failure occurred in a relatively short interval (t_1, t_2) in the disease studies, or when intervals are randomly selected for different individuals.

However we find that all is not lost, certain parametric family of failure models can be estimated by rewriting the likelihood in terms of the survivor functions (defined above). The MLE estimator retains the advantages of maximum likelihood estimation, i.e. it is consistent and distributed as $\sqrt{N}$ asymptotic normal.[24]

The data we have consists of $m = 1, 2, \ldots, M$ cross-sections of N_m individuals at times $t_1, t_2, \ldots, t_M$. We will follow convention and denote individual i observed at time t as $i(t)$). Therefore, let $X_{i(t)}$ denote the characteristics of individual i in the survey collected at time t. The variable we are interested in is coded as a binary variable $y_{i(t)} = 1$ if individual $i(t)$ is a user of the new technology and $y_{i(t)} = 0$ otherwise.

We outline here two additional assumptions that we need to make for duration models to be identified in this context.

Assumption 4 *Adoption is an absorbing state, i.e. if* $y_{i(t)} = 1 \rightarrow y_{i(t^+)} = 1$ *for all* $t^+ \geq t$.

Assumption 5 *Alternatively assume that the analyst has at her disposal a variable*

[24] It is easy to verify the sufficient Kiefer-Wolfowitz conditions in this case.

that summarizes whether the individual ever used the new technology given that she is not using it now, i.e. $z_{i(t)}|(y_{i(t)} = 0) \in \{0, 1\}$.

Note that the first assumption is very common in the literature on adoption and this is usually justified by noting that adoption usually implies that the new technology is superior/more productive compared to older preexisting technologies, for rational consumers. More generally in this setup if this is unlikely to be true, the model is also identified under the weaker condition that individuals may terminate usage but the data indicating past usage is available to the analyst.

If both assumptions are violated then our methodology outlined below fails, no duration model can be estimated using such repeated cross-sectional discrete choice data. The intuition for this is simple, $y_{i(t)}$ helps us partition the sample into two sets as follows, denoting the unobserved adoption time as t_a if $y_{i(t)} = 1$ we know that for individual $i(t)$ adoption took place before the survey was conducted (at time t) i.e.

$$y_{i(t)} = \begin{cases} 1 \;\; iff \;\; t_a \leq t \\ 0 \qquad\quad t_a \in [t + \epsilon, \infty) \end{cases}$$

If adoption is not an absorbing state then this one-to-one relationship does not hold anymore since $y_{i(t)} = 0$ includes two groups of people those who have not yet adopted the technology i.e. $t_a \in [t + \epsilon, \infty)$ as well as those who had adopted but since then have stopped using it $t_a \leq t$. However all is not lost as long as we have another variable that serves the same purpose as $y_{i(t)}$ did before, i.e.

$$z_{i(t)}|(y_{i(t)} = 0) = \begin{cases} 1 \;\; iff \;\; t_a \leq t \\ 0 \qquad\quad t_a \in [t + \epsilon, \infty) \end{cases}$$

In the absence of such information one has to make further assumptions regarding the proportion of users who adopt and subsequently stop using the technology to be able to estimate any kind of a duration model.

2.4.2 Likelihood

The likelihood in this case is the standard discrete choice likelihood with probabilities of success given by the survivor function from before. Since following our discussion from before we can partition the data into two sets of past adopters and future adopters, therefore we can write the likelihood function as follows

$$\mathcal{L} = \prod_{y_{i(t)}=1} Prob(\tau_i \leq t) \prod_{y_{i(t)}=0} Prob(\tau_i > t) \tag{2.17}$$

Under the assumption that individuals are independently and identically distributed $(i.i.d)$ we can simplify the log likelihood function as follows (after taking logarithm and using the definition of the survivor function $S(t)$ from (2.9) above),

$$\mathcal{LL} = \sum_{t\in\{t_1,t_2,\ldots,t_M\}} \sum_{i=1}^{N_m} \Big[y_{i(t)} \log(1 - S_i(t)) + (1 - y_{i(t)}) \log S_i(t) \Big] \tag{2.18}$$

Since most standard parametric distributions $(F(t))$ have closed form solutions for the survivor function $S(t)$ this log-likelihood can be maximized to obtain the maximum likelihood estimates (MLE) of the parameters. For example using the definition of the Weibull hazard rate from table (2.1) we can write its log-likelihood as follows:

$$\mathcal{LL} = \sum_{t\in\{t_1,t_2,\ldots,t_M\}} \sum_{i=1}^{N_m} \Big[y_{i(t)} \log(1 - exp\{(exp(-X'\beta)t^{\alpha})\}) - (1 - y_{i(t)})(exp(-X'\beta)t^{\alpha}) \Big] \tag{2.19}$$

similarly for other parametric forms discussed before. Some shortcomings of this approach include the heavy computational burden involved in maximizing non-linear functions.

2.4.3 Semi-parametric estimation

It is evident that non-parametric estimation methods such as the Kaplan-Meier cannot be applied in this context since the actual durations are not known.[25] In theory this can be done for instance with 'case 1' interval censoring when the censoring is at random times for each observation. However with RCS data most observations share a common censoring time which is simply the date when that particular wave of the survey was conducted. However we show that semi-parametric estimation might be possible in this context. Consider the mixed proportional hazard model introduced earlier (see table (2.1) above), which in some sense is the most general semi-parametric model out there. The usual approach taken in the literature is to allow one component of this mixture to vary freely and to specify the other, for example Meyer (1990) assumes a unit mean gamma distribution for the unobserved heterogeneity term and allows a non-parametric specification of the baseline hazard. Conversely, others assume a parametric form for the baseline hazard and allows the heterogeneity distribution to be flexible (for example see Heckman and Singer (1984)).

In order to estimate the proportional hazard model in this context, we need the survivor function for the model, in order to derive the likelihood. We use the fundamental relation in this context:

$$S(t) = \exp\left(-\int_0^t h(s)ds\right) \tag{2.20}$$

and using the definition of mixture models from table (2.1) (for the moment assume

[25] Horowitz (1999) and (1996) discusses various methods for semi-parametric estimation of both the baseline hazard as well as the mixture distribution.

that ν is known and the survivor function conditional on ν is S_ν), we get

$$S_\nu(t) = \exp\left(-\int_0^t \nu \exp(x'\beta) h_0(s) ds\right) \tag{2.21}$$

In order to evaluate this integral we could use either a flexible parametric form for the baseline hazard rate $h_0(t)$, say a second order polynomial which can capture the U-shaped empirical hazards often obtained in the real world. Therefore assuming $h_0(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2$, gives us

$$S_\nu(t) = \exp\left(-\nu \exp(x'\beta)[\alpha_0 t + \alpha_1 t^2/2 + \alpha_2 t^3/3]\right) \tag{2.22}$$

Alternatively from (2.21)

$$S_\nu(t) = \exp\left(-\nu \exp(x'\beta) \int_0^t h_0(s) ds\right) \tag{2.23}$$

Then let us define the variable $\exp(\gamma(t)) = \int_0^t h_0(s)ds$ which when substituted in (2.23) gives us

$$S_\nu(t) = \exp\left(-\nu \exp[x'\beta + \gamma(t)]\right) \tag{2.24}$$

as before let the heterogeneity term be distributed as $\nu_i \sim f(\nu; \eta)$ then the unconditional survivor function can be found by taking expectations using the distribution of the unknown heterogeneity term.

$$S(t) = \int \exp\left(-\nu \exp[x'\beta + \gamma(t)]\right) f(\nu; \eta) d\nu \tag{2.25}$$

Similarly for the polynomial case. The log likelihood is obtained by substituting expression (2.25) in (2.18) above. The $\gamma(t)$ terms are called *splines*, in this case with five waves of data, there are five splines to be estimated alongside the usual parameters β and, with unobserved heterogeneity, η also needs to be estimated. In this context it is customary to assume a gamma distribution for this heterogeneity term with unit mean and variance σ^2 in order to avoid numerical integration, since the gamma distribution provides a closed form solution for equation (2.25), as follows:

$$S(t) = \left[1 + \sigma^2 \exp\{x'\beta + \gamma(t)\}\right]^{1/\sigma^2} \tag{2.26}$$

2.5 Monte Carlo Results

In this section we present evidence to support our claim that duration models can be estimated using discrete choice data in the form of RCS. This claim is not immediately obvious given the heavily censored nature of the data. We show empirically that all the models we consider are indeed identified and the MLE estimator is efficient for small samples in certain situations. Usually surveys with RCS data such as expenditure surveys have a broad coverage with very many observations. This fact somewhat counteracts the loss of information from the censoring and we find that one can get arbitrarily close estimates of the true values, particularly with parametric models when they are correctly specified. We largely follow the methodology laid out by Hendry (1984) in the context of Monte Carlo studies.

For uniformity and to maintain comparability across models, in all the studies reported below we usually consider a single covariate x and a constant term with coefficients β_1 and β_0 respectively. We start off with the Weibull model given its widespread usage in the literature, the results are reported in table 2.2. We consider two variants of the model, the top half of the table reports evidence from a monotonically declining hazard ($\alpha = 0.5$) and the bottom half considers an increasing hazard rate ($\alpha = 1.5$). To highlight the loss of information from RCS data we contrast it with data containing actual durations measured upto the second decimal place which (almost surely) rules out any ties in the actual data.[26] The parameters are chosen for the *data generating process* (henceforth DGP) such that about $10 - 15\%$ of the observations are censored at $t = 100$ (most real life data would contain similar censoring percentages). Similarly the distribution for the single covariate x was selected with the same goal. Note that in most models discussed here a higher value of x implies longer durations. We found that for all the studies reported here the scale factor (of time) chosen did not affect the estimates, particularly for the models with actual durations, however they had a very large impact on estimates obtained from the RCS data. Intuitively this makes sense since RCS provides snapshots of the failure process and if most failures occur early on in the process, the cross-sections collected later on contain very little additional information.

The DGP proceeds as follows, we first generate the covariate x from a normal distribution with mean 2 and variance 1. We generate 100 samples each of size 1000, containing actual durations. Since it is important for comparison purposes to use the same data for the different methods (see Hendry (1984),Baker and Melino (2000)), if we need a data set of $N = 100$ we use the first 100 obs. for each sample and so on. In the second step we use this data to generate the RCS data, without loss of generality we take four cross-sections equi-spaced over time and of equal size (five for the proportional hazards model reported below). The discrete data in the form of RCS are generated at times $t_{(j)} = \{4, 8, 12, 16\}$.[27] We found that the results do not vary significantly for the RCS data provided the cross-sections are of comparable size, time intervals do not vary significantly and collectively they contain sufficient variation in terms of failure percentages.

[26] In reality usually the data is far more discretized.

[27] Since almost 50% of adoptions occur by period 20, these times were calibrated to match proportions of adoptions in the real data.

Table 2.2:

Weibull Hazards with Repeated Cross-sections

MLE	α	β_0	β_1	Obs.
True Values	0.5	1.0	1.0	
Durations	0.51	0.98	1.00	100
	(0.05)	(0.47)	(0.22)	
RCS	0.50	-25.11	22.50	100
	(0.35)	(96.18)	(75.67)	
RCS	0.50	0.92	1.08	500
	(0.12)	(0.50)	(0.47)	
RCS	0.51	0.99	1.01	1000
	(0.09)	(0.33)	(0.21)	
True Values	1.5	1.0	1.0	
Durations	1.51	1.00	1.00	500
	(0.06)	(0.07)	(0.03)	
RCS	1.51	0.96	1.03	500
	(0.21)	(0.22)	(0.15)	

SE in parentheses. No ties, 100 samples.

For the Weibull model we found that in both cases with actual durations known the parameter estimates are very close to the true values even with very small samples $N = 100$. Not surprisingly for RCS data we need a much larger sample to obtain reasonable estimates as expected. However the estimates obtained are consistent, arbitrarily close estimates of the true value can be obtained with RCS data for samples of size 500 or larger in this setup. For very large samples of 1000 or $10,000$ (not reported) we found little difference between the two models. Also we found that the shape of the baseline hazard (increasing / decreasing / constant) does not have any impact on these conclusions.

We next consider by turns two other parametric models widely used for their simplicity. The top half of table 2.3 reports the results from the *log-logistic* model and bottom half does so for the *lognormal* model. For the log-logistic we assume the parameter value of $\alpha = 1$, which generates a U-shaped hazard. Here we find that surprisingly the log-logistic is highly efficient and converges to the true values for very small samples of around 100 only, with a larger sample the coefficients are highly significant as well.

For the lognormal model we consider a homoscedastic model with $\sigma = 1$. We find that the RCS data actually performs better for very small samples of 100 in terms of consistency, compared to actual durations, and with larger samples of 500 or more it is also efficient. Therefore in both cases we found that sample sizes of at least 1000 were more than sufficient with RCS data to consistently estimate the true parameters of the model. Typically census collected survey data sets (such as expenditure surveys) tend to have hundreds of thousands of observations (see below), therefore we can expect the estimates to be highly significant.[28]

[28] In this context we also tried the MCEM algorithm, which is an implementation of the standard *EM* algorithm widely used in this literature, using monte carlo integration. We found that with RCS data this algorithm performs adequately, the estimates are actually better with sample size of 500 than with actual durations. Using this approach potentially any model that can be estimated

Table 2.3:

Lognormal/Log-logistic Hazards with Repeated Cross-sections

Log-logistic	α	β_0	β_1	Obs.
True Values	1.0	1.0	1.0	
Duration	1.01	0.96	1.03	100
	(0.08)	(0.25)	(0.19)	
RCS	1.05	0.99	1.08	100
	(0.49)	(1.05)	(0.27)	
RCS	0.97	0.90	1.02	500
	(0.20)	(0.45)	(0.14)	
RCS	1.00	1.00	0.99	1000
	(0.16)	(0.35)	(0.08)	

Lognormal	σ	β_0	β_1	
True Values	1.0	1.0	1.0	
Duration	0.88	1.31	0.72	500
	(0.03)	(0.09)	(0.04)	
RCS	1.1	0.96	1.036	500
	(0.352)	(0.227)	(0.164)	
RCS	0.99	0.98	1.00	1000
	(0.20)	(0.15)	(0.10)	

SE in parentheses. No ties, 100 samples.

Table 2.4:

Proportional Hazards with Repeated Cross-sections

	Model I	Model II		
True ($\beta = 1$)	$\hat{\beta}$	$\hat{\beta}$	$\hat{\sigma}^2$	$\sigma^2 = 1/\eta$
(a) Actual Durations	0.998	0.566		1
	(0.05)	(0.04)		
(b) Discrete data (with ties)	0.953	0.537		1
	(0.05)	(0.04)		
(c) Discrete data (coarse grid)	0.795	0.448		1
	(0.04)	(0.03)		
(d) Repeated Cross-sections	0.479	0.354		1
(5 waves)	(0.03)	(0.03)		
(e) With EM correction for het.		1.221	0.295	1
		(0.03)	(0.10)	
(f) Actual durations		0.712		0.5
		(0.05)		
(g) With EM correction for Het.		1.137	0.321	0.5
		(0.02)	(0.07)	
(h) Repeated Cross-sections		0.349		0.5
(5 waves)		(0.03)		

All max. partial likelihood est. except RCS data.
SE in parentheses. 100 samples N=1000

The proportional hazards results are reported in tables (2.4) and (2.5). In the first table we consider two versions the standard one and the mixture one i.e., with and without unobserved heterogeneity. Model II includes a gamma unobserved heterogeneity term (several variants are considered). It has been well documented that with interval censored data and particularly with unobserved heterogeneity the partial likelihood estimates[29] are seriously biased. We also show in the top half how the bias increases with increasing discretization of time[30] See Lancaster (1990) for an EM correction in this context which controls for such heterogeneity. As expected the bias worsens with the variation in the heterogeneity term (as measured by the variance of the gamma distribution). In table 2.5 we consider two alternative ways of estimating a mixture model with RCS data. We find that the estimates are consistent with no heterogeneity, or when it is explicitly controlled for in the estimation process. Polynomial specifications of the baseline hazard usually performs better with or without heterogeneity. However when the heterogeneity is controlled for a spline based piecewise constant hazard is highly efficient and unbiased.

2.6 Data

The data used for this project was discussed earlier in details in chapter 1 and will not be repeated here.

using actual durations can also be estimated with RCS data, since actual durations can be treated as unobserved data and conditioned on. However we abandoned this approach due the extreme computational burden involved with even modest sample sizes. Note that it has been suggested that one needs 10,000 random draws for an accurate estimate of the expectation step.

[29] Standard methodology used for semi-parametric estimation of these models.

[30] As observations are only observed at more discrete (longer) intervals of time.

Table 2.5:

Prop. Haz. with RCS II

(Quasi-likelihood approach)

RCS data (5 waves)	$\hat{\beta}$	$\hat{\sigma}^2$
A. Polynomial baseline hazard		
DGP: $\beta = 1$, $h_0 = 0.05$, *no Het.*		
(a) *no heterogeneity*	1.017	
	(0.08)	
DGP: $\beta = 1$, $h_0 = 0.1$ *w/ Het.*		
(b) *gamma het.* $\sigma^2 = 0.5$	0.76	
	(0.06)	
(c) *gamma het.* $\sigma^2 = 2$	0.50	
	(0.05)	
B. Spline baseline hazard		
(d) *gamma het.* $\sigma^2 = 0.5$	0.59	
	(0.11)	
(e) *gamma het.* $\sigma^2 = 2$	0.40	
	(0.07)	
splines and het. correction		
(f) *gamma het.* $\sigma^2 = 0.5$	1.065	2.24
	(0.32)	(1.52)
(g) *gamma het.* $\sigma^2 = 2$	0.70	1.70
	(0.18)	(1.13)

SE in parentheses. 100 samples N=1000

2.6.1 Variables of Interest

We use the CPS data to construct the variables of interest as follows. We use data from five cross-sections and in four of them (2001, 2000, 1998 and 1997) households were explicitly asked whether they had access to the Internet, if they had either a personal computer or Web TV at home. If not then they were asked whether they had ever used the Internet from home. We use the responses to these questions to construct the primary dependent variable *Internet*, which is a dummy variable taking on the value one if the household is a current or past user of the Internet and zero otherwise. However, for the 1994 sample[31] respondents were asked whether they had a personal computer at home, the specifications of the computer and various other usage questions, for example does anyone in this household use the computer for reading the news etc. In this case we infer that the household had access to the Internet if the household had a computer with a modem *and* if the respondent answered yes to any of the questions regarding usage that require an Internet connection.

For education the baseline case is taken to be no high school diploma or equivalent (GED). The following categories are subsequently included as dummy variables, a) high school diploma or GED, b) some college but no degree or an associate degree in a vocational or academic program, c) bachelors degree and, d) any advanced degree including a master's degree, or professional or doctorate degree. The CPS following the Census 2000 convention classifies Hispanics as an ethnicity and not as a separate race, i.e. being of parental origin from certain South/Central American countries, therefore racially they are classified as either white or black. However in our study we found substantial differences with other whites and blacks and do control for them as

[31] In 1994 the Internet was still a highly specialized technology only used by a few people in academics and in the military. We include this sample since theoretically the current expansion of the Internet can be traced back to the invention of the World Wide Web (WWW) by Tim Berners-Lee in 1991, which predates the sample.

a separate ethnic group. We take the baseline case as whites of non-Hispanic origin and use dummy variables to control for black non-Hispanic households and Hispanic (both whites and blacks) households, and Asians.[32]

2.7 Results

2.7.1 Discrete Models

We start by presenting the familiar evidence usually cited in support of the digital divide, using standard discrete choice models such as the logit and probit. These models are useful in providing a snapshot of the diffusion process particularly when a single cross-section is available to the analyst. Based on the simple behavioral model presented earlier we estimate a logit and probit model with year specific dummies included for the pooled sample. These estimates are presented in table 2.6. We find that age lowers the probability of adoption, higher income and education raise this probability. The racial divide is also documented with blacks and Hispanics much less likely to adopt the Internet. Also we note the rural urban divide in adoption patterns.

2.7.2 Duration Models

In order to use duration models we need to specify the exact duration of the process, for this two relevant dates are required, first the date of origin i.e. the date from when the good / technology is available to the household for adoption and second, the actual date of adoption. Unfortunately we could not locate any data on the initial availability of the Internet by geographic location. Therefore as origin we take January 1993, since this was the year when by most indicators the Internet began its

[32] Which also includes Pacific Islanders and natives of Hawaii etc.

explosive growth. Each survey date is coded as months from this date. In table 2.7 we report the estimates from the three standard duration models most often used in empirical work. The coefficients all have the expected signs, note that a negative sign in this context implies a positive influence i.e. it moves the mean adoption time of the distribution to the left on the time axis. We find that increasing age of the householder delays adoption and similarly the higher the family income and higher the education level of the householder the more likely it is to adopt the Internet early. Not surprisingly we find a digital divide in terms of a difference in adoption timing among various racial and ethnic groups even after controlling for other demographic variables. The only surprisingly result is that we find a negative sign for Asian origin, however the estimate is not significant. The distributional parameters implies a monotonically increasing hazard rate for the Weibull since $\alpha > 1$, and a U-shaped one for the other two models (by definition for the lognormal and, since $\alpha > 1$ for the log-logistic).

In table 2.8 we take the models from before and add a set of geographic dummy variables for the northeast, Midwest and the west (south being the excluded dummy). We also add rural and central city dummies to measure the urban versus rural divide in technology usage as well as any *inner city* phenomenon. Note however that central cities as defined in the CPS are fairly large areas and measure the whole downtown of any metropolitan area and only excludes the suburbs. Unfortunately we could not obtain data at a more disaggregate level. In this table we also report the results for the proportional hazards model and the mixture model with gamma heterogeneity. The standard effect of income, age, gender, education and race stays the same as before. As expected both living in rural areas and in central cities lowers the probability of adoption. Geographically living in the northeast increases adoption probabilities all else equal, the Midwest is actually behind the south in diffusion rates and the west is significantly ahead (California strongly influences this result). Therefore we find that there is some truth to the rural urban digital divide from these estimates.

2.7.3 Model Selection

We considered two alternative model selection criteria commonly used in the literature the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) defined as follows:

$$AIC: \quad -2\log(\hat{L}) + 2K \qquad BIC: \quad -2\log(\hat{L}) + log(N)K$$

where K is the number of parameters, N is sample size and $\hat{L}$ is the maximized value of the log-likelihood. In this context they yield identical results as follows, first the three common parametric models presented in table 2.7 have the same number of parameters and the same sample size therefore comparing their likelihoods we find that the log-logistic has highest log-likelihood value and therefore is the clear choice. Unfortunately due to the complex weighing scheme used for the QMLE this likelihood is not directly comparable to the other models. However from table 2.8 we find that the proportional hazards has a lower value of both statistics ($AIC_{LL} - AIC_{PH} = 1504$ and $BIC_{LL} - BIC_{PH} = 1561$). Similarly we can show that both values are even lower for the mixture model, therefore we conclude that among all the models considered the mixture model with gamma heterogeneity and splines as baseline hazards do the best in describing the data.

Table 2.6:
Simple Adoption Model

Discrete	Logit		Probit	
Models*	(1)	(2)	(3)	(4)
Age of Householder	-0.026	-0.026	-0.015	-0.015
	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.191	0.198	0.112	0.116
	(0.014)	(0.014)	(0.008)	(0.008)
Income	0.842	0.841	0.477	0.476
$25,000-50,000	(0.018)	(0.019)	(0.010)	(0.010)
$50,000-75,000	1.497	1.496	0.864	0.863
	(0.021)	(0.021)	(0.012)	(0.012)
$75,000	2.020	2.016	1.176	1.174
	(0.022)	(0.023)	(0.013)	(0.013)
Education1	0.713	0.702	0.381	0.375
(No HS degree)	(0.028)	(0.029)	(0.015)	(0.015)
Education2	1.323	1.308	0.739	0.730
(HS / some college)	(0.028)	(0.029)	(0.016)	(0.016)
Education3	1.652	1.622	0.933	0.917
(College degree)	(0.030)	(0.031)	(0.017)	(0.017)
Education4	1.849	1.821	1.046	1.030
(Graduate Degree)	(0.033)	(0.034)	(0.019)	(0.019)
Hispanic	-0.651	-0.670	-0.373	-0.384
	(0.029)	(0.031)	(0.017)	(0.018)
Black	-0.793	-0.786	-0.456	-0.451
	(0.026)	(0.027)	(0.015)	(0.015)
Asian	-0.001	-0.021	0.002	-0.009
	(0.379)	(0.040)	(0.022)	(0.023)
Rural	-0.256	-0.246	-0.150	-0.143
	(0.018)	(0.036)	(0.010)	(0.021)
Central City	-0.071	-0.094	-0.039	-0.053
	(0.017)	(0.019)	(0.010)	(0.011)
MSA dummies	No	Yes	No	Yes
Log-likelihood	-93,024	-92,040	-92,973	-91,991

Standard errors (robust) in parentheses. N=220,758

**All specifications used household weights.*

All include year dummies for four years.

Table 2.7:
Parametric Duration Models I

	Weibull	Lognormal	Log-logistic
Age of Householder	0.014	0.017	0.025
	(0.000)	(0.000)	(0.000)
Male	-0.082	-0.101	-0.0145
	(0.003)	(1.869)	(0.011)
Income	-0.555	-0.564	-0.838
$25,000-50,000	(0.002)	(0.016)	(0.011)
$50,000-75,000	-0.918	-1.020	-1.489
	(0.011)	(0.020)	(0.014)
$75,000 & above	-1.186	-1.409	-2.050
	(0.017)	(0.026)	(0.015)
Education1	-0.526	-0.444	-0.711
(No HS degree)	(0.002)	(1.549)	(0.011)
Education2	-0.867	-0.853	-1.296
(HS / some college)	(0.006)	(0.008)	(0.015)
Education3	-1.016	-1.072	-1.609
(College degree)	(0.019)	(0.009)	(0.016)
Education4	-1.112	-1.206	-1.813
(Graduate Degree)	(0.014)	(0.011)	(0.019)
Hispanic	0.315	0.379	0.558
	(0.018)	(0.018)	(0.023)
Black	0.437	0.513	0.76
	(0.007)	(0.017)	(0.017)
Asian	-0.013	-0.026	-0.041
	(0.018)	(0.022)	(0.032)
Constant	4.158	5.515	8.018
	(0.006)	(0.004)	(0.008)
Distribution	$\alpha = 1.295$	$\sigma = 1.170$	$\alpha = 1.567$
parameters	$\log \alpha = 0.259$	$\log \sigma = 0.157$	$\log \alpha = 0.449$
(shape of pdf)	(0.003)	(0.002)	(0.002)
Log-Likelihood	-93,925	-93,994	-93,837

Standard errors in parentheses. N=220,758

Table 2.8:
Duration Models II
Proportional Hazard cols.(3-4)

	Weibull	Log-logistic	Log-logistic	No Het.	Gamma
			QMLE		Het.
Age of Householder	0.014	0.025	0.025	0.018	0.028
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Male	-0.086	-0.155	-0.154	-0.123	-0.199
	(0.007)	(0.011)	(0.008)	(0.007)	(0.012)
Income	-0.546	-0.827	-0.844	-0.708	-0.893
$25,000-50,000	(0.015)	(0.017)	(0.009)	(0.012)	(0.016)
$50,000-75,000	-0.901	-1.465	-1.484	-1.168	-1.60
	(0.015)	(0.017)	(0.011)	(0.012)	(0.02)
$75,000 & above	-1.157	-2.008	-2.047	-1.501	-2.187
	(0.021)	(0.018)	(0.013)	(0.013)	(0.023)
Education1	-0.521	-0.707	-0.687	-0.68	-0.757
(No HS degree)	(0.014)	(0.015)	(0.013)	(0.016)	(0.037)
Education2	-0.852	-1.278	-1.272	-1.112	-1.40
(HS / some college)	(0.014)	(0.016)	(0.012)	(0.015)	(0.034)
Education3	-0.999	-1.588	-1.579	-1.314	-1.761
(College degree)	(0.017)	(0.018)	(0.012)	(0.016)	(0.04)
Education4	-1.099	-1.798	-1.762	-1.446	-1.986
(Graduate Degree)	(0.029)	(0.022)	(0.009)	(0.018)	(0.043)
Hispanic	0.315	0.627	0.639	0.471	0.724
	(0.017)	(0.023)	(0.016)	(0.017)	(0.032)
Black	0.429	0.747	0.752	0.571	0.854
	(0.014)	(0.018)	0.012	(0.016)	(0.029)
Asian	0.029	0.041	-0.009	0.042	0.057
	(0.019)	(0.031)	(0.017)	(0.021)	(0.045)

Standard errors in parentheses. N=220,758
Non-parametric splines used for baseline hazard.

Table 2.8:

Duration Models II (cont.)

Proportional Hazard col. (3–4)

	Weibull	Log-logistic	Log-logistic QMLE	No Het.	Gamma Het.
Northeast	-0.037	-0.083	0.013	-0.091	-0.144
	(0.009)	(0.015)	(0.010)	(0.011)	(0.03)
Midwest	0.034	0.05	0.096	-0.049	-0.062
	(0.012)	(0.015)	(0.010)	(0.01)	(0.019)
West	-0.101	-0.19	-0.133	-0.179	-0.271
	(0.011)	(0.015)	(0.010)	(0.01)	(0.028)
Rural	0.12	0.217	0.319	0.165	0.256
	(0.011)	(0.013)	(0.020)	(0.01)	(0.02)
Central City	0.049	0.087	0.084	0.063	0.088
	(0.011)	(0.013)	(0.008)	(0.009)	(0.021)
Constant*	4.116	7.997	8.014		
	(0.008)	(0.011)	(0.011)		
Distribution	$\alpha = 1.3$	$\alpha = 1.575$	$\alpha = 1.585$		$\sigma^2 = 0.297$**
parameters	$\log \alpha = 0.262$	$\log \alpha = 0.454$	$\log \alpha = 0.461$		$\log \sigma^2 = -1.214$
(shape of pdf)	(0.02)	(0.002)	(0.002)		(0.02)
Log-Likelihood	-93,703	-93,607	$-213,584^{\#}$	-92,851	-92,212

Standard errors in parentheses. N=220,758

Non-parametric splines used for baseline hazard.

**Constant for the proportional hazard model is not identified.*

***σ^2 is the variance of the gamma heterogeneity term*

# This likelihood is weighted and therefore not comparable to the others.

Table 2.9:

Duration Models III

Log-Logistic Model

Fixed Effects, Large Urban Sample#

	MSAs*	Counties*	10 Largest	Restricted
			MSAs	Model**
Age of Householder	0.026	0.027	0.025	0.025
	(0.001)	(0.002)	(0.001)	(0.001)
Male	-0.166	-0.163	-0.194	-0.195
	(0.02)	(0.058)	(0.031)	(0.03)
Income	-0.804	-0.844	-0.801	-0.796
$25,000-50,000	(0.024)	(0.178)	(0.093)	(0.046)
$50,000-75,000	-1.453	-1.451	-1.484	-1.478
	(0.028)	(0.038)	(0.047)	(0.049)
$75,000 & above	-1.99	-2.059	-1.963	-1.955
	(0.03)	(0.954)	(0.049)	(0.051)
Education1	-0.631	-0.69	-0.611	-0.611
(No HS degree)	(0.036)	(0.255)	(0.059)	(0.06)
Education2	-1.162	-1.246	-1.123	-1.126
(HS / some college)	(0.046)	(0.195)	(0.058)	(0.059)
Education3	-1.496	-1.529	-1.409	-1.409
(College degree)	(0.037)	(0.2)	(0.061)	(0.061)
Education4	-1.707	-1.702	-1.685	-1.68
(Graduate Degree)	(0.045)	(0.216)	(0.066)	(0.065)
Hispanic	0.631	0.67	0.659	0.673
	(0.034)	(0.079)	(0.056)	(0.054)
Black	0.707	0.805	0.693	0.708
	(0.03)	(0.064)	(0.058)	(0.046)
Asian	0.04	0.086	-0.033	-0.025
	(0.04)	(0.41)	(0.626)	(0.063)
Central City	0.107	-0.087	0.174	0.187
	(0.017)	(0.039)	(0.185)	(0.031)
Log-Likelihood	-39,199	-17,981	-13,180	-13,190
MSAs/ Counties	75	31	10	10
N	92,567	37,816	32,695	32,695

*Standard errors in parentheses. *Only MSAs/counties with* $N \geq 500$.
#All specifications include other geographic variables.
***Restricted model refers to baseline model with no fixed effects.*

2.7.4 Quasi-maximum likelihood

Most survey data is collected through *stratified* random sampling, i.e. the population is divided into stratas and then randomly some strata are selected and households from that strata are sampled. Sample weights are usually provided which gives the inverse of the probability of selection for the household or the number of similar households in the population. Within strata households are usually selected based on demographics which implies endogenous sampling i.e. selection of sample depends on X. Earlier Hausman and Wise (1981) had shown that in such cases estimates of the linear model using weights (usually sample weights) can provide consistency and asymptotic normality. Wooldridge (2001) derives similar results for a broad class of *M-estimators* which includes the maximum likelihood as a special case, he shows that with endogenous sampling an unweighted estimator might be inconsistent but still retains the feature of asymptotic normality. He also shows that a weighted version of MLE, using sample weights which is generally referred to as *quasi maximum likelihood (QMLE* in the literature, is both consistent and asymptotically normal. Therefore we reestimate the QMLE for the log-logistic model reported in table 2.8. We find that none of the main coefficients change significantly from their earlier unweighted estimates. Only changes are in the estimates for Asian which changes sign but is insignificant and, in the geographic dummy variables for the northeast and the Midwest, the former changes signs however both turn out to be insignificant.

2.7.5 Testing for heterogeneity

A potentially serious issue, mentioned in the literature, is the presence of unobserved heterogeneity due to either unobserved variables such as ability or measurement error. Heckman and Singer (1984) show through monte carlo simulations that the existence of such factors seriously biases the results obtained. There are several ways to test for unobserved heterogeneity (see discussion above). The simplest way is to assume a random effects model with the unobserved factor distributed across the population as a unit gamma distribution. The standard nested test for unobserved heterogeneity in this context verifies whether the variance of the estimated gamma distribution is zero. The variance is reported in table 2.8, since it was constrained to be positive in the estimation procedure, $\sigma^2 = 0$ implies here $\log \sigma^2 = -\infty$, which can be safely rejected at all levels of significance. However we do find that the variance estimated is small and almost negligible at 0.3. Similarly a likelihood ratio test rejects the hypothesis of no unobserved heterogeneity. Specifically in this context the restricted model is the standard proportional hazards model (column 4) and the unrestricted model is the mixture model (column 5), denoting the respective log-likelihood values as L_R and L_U, we can write the test statistic as follows:

$$-2[L_R - L_U] = -2[-92,851 + 92,212] = 1278 \quad \Rightarrow \quad \Pr(\chi^2_1 \geq 1278) \approx 1.0$$

In table 2.9 we take a different approach by assuming clustering, that is we define aggregate fixed effects for each location, i.e. we assume that people living in different locations fundamentally differ in terms of their unobserved ability, however for simplicity all observations from that location share the same fixed effect.[33] An example might be San Francisco (with Silicon Valley) compared to any other location in the country, the group effect essentially captures the fact that a priori one expects a higher likelihood of adoption for people living there. Even at the aggregate MSA level the data contains more than 300 unique locations and we found given the highly non-linear nature of the log-likelihood it was not feasible to include all such

[33] We consider this a compromise since we do not have enough data to identify true individual fixed effects.

variables. Note that for rural consumers we do not have sufficient data to estimate locational fixed effects, therefore we restrict our sample to large MSAs or counties with large populations (most MSAs contain a number of counties). Based on the monte carlo simulations reported earlier we decided that a sample of 500 was reasonable to estimate the log-logistic model and so we only chose MSAs or counties with more than 500 observations in the pooled sample. This left us with data on 75 MSAs and 31 counties.

In our first specification we define the locational fixed effects as a linear function of the characteristics of that location i.e. $\delta_k = Z_k\eta$ where Z_k are MSA/county characteristics such as income, age or educational distribution. This simplifies estimation since we can write:

$$\lambda_i = \exp\{X_i'\beta + Z_{ik}\eta\}$$

The estimates for the MSA and county level are reported in the first two columns of table 2.9. We do not find any significant differences from our baseline estimates in table 2.8, although they are estimating somewhat different model, the former is estimated for the whole country and the latter primarily for large urban centers and densely populated suburbs. Estimates remain similar in substance although standard errors rise due to fewer observations and also due to multicollinearity between X and Z. We find our estimates for the racial divide is actually larger and still highly significant. For the counties we find a reversal of sign for the central city dummy which is due to insufficient observations.[34] As before a likelihood ratio test of the restriction of $H_0 : \eta = 0$ is overwhelmingly rejected.

Instead of projecting the locational fixed effects on characteristics of the location we now allow a more flexible specification of the unobserved heterogeneity term by including a constant fixed effect for each location. However this flexibility comes at

[34] Note that the rural dummy from before is dropped due to the nature of the sample.

a cost, we found convergence to be a serious problem the more dummy variables we added. Therefore we settled on a compromise, we picked out only ten MSAs with the most number of observations and estimated the model with nine dummy variables for locations. The estimates are reported in column 3 in table 2.9, and in the next column we report the estimates from the restricted model (our baseline log-logistic model) for comparison. As before a likelihood ratio test rejects the null hypothesis of no heterogeneity at 5% level of testing.

$$-2[L_R - L_U] = -2[-13,190 + 13,180] = 20 = \Pr(\chi_9^2 \leq 20) = 0.982$$

2.7.6 Other Models

We considered by turns the non-parametric MLE suggested by Heckman and Singer (1984), however as noted by other authors (for example see Baker and Melino (2000)), we found the maximization routine failed to converge. In this situation others have arbitrarily assumed a binomial distribution and estimated the model, however for lack of space we do not report these results here. Also we found the split population model mentioned above to be extremely unstable and almost always failed to converge particularly for larger samples, we also do not report those results here.

2.8 Predictions for Individuals

2.8.1 Predictive Power

A potential use for such models is to predict the adoption of new technology by individuals. In this section we consider how well does the models presented above achieve that goal. Since adoption is a discrete event and the duration models presented here provide adoption probabilities at each date, one can calculate the goodness of fit measure R^2 which is the correlation coefficient between the dependent binary variable and predicted probabilities. However it is well known that in the context of limited dependent variables this measure does not have the explained variation interpretation as in linear regression models (see for example Maddala (1983)). To measure graphically the goodness of fit of the models considered here we take two of the parametric models and plot their distribution function (cdf $F(t|x;\beta)$) against the actual distribution function (adoption rates) obtained from the data, we consider the lognormal and the log-logistic model here in figures 2.3 (a)–(b).

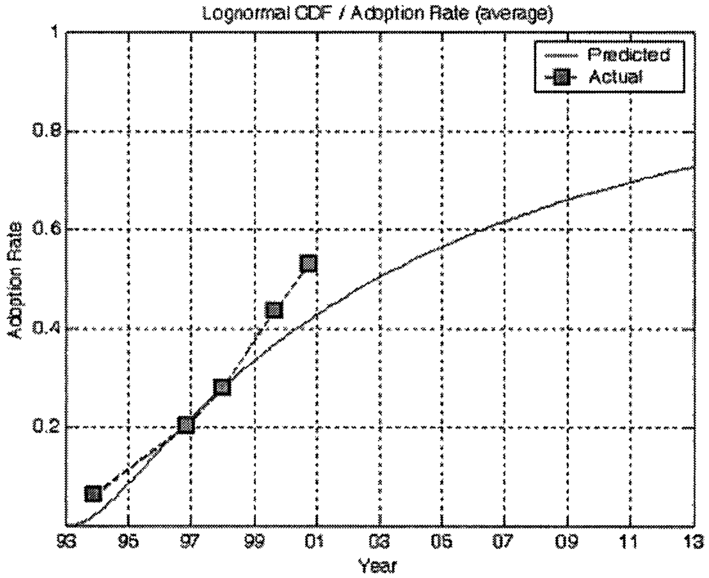

Figure 2.3: Lognormal Fitted vs. Actual

A more intuitive approach used by Schmidt and Witte (1989) is to predict individual adoption probabilities and choose a cutoff such that people with predicted probability higher than this are predicted to adopt and vice versa. From our perspective a highly stylized dynamic model can be considered a huge success if it can reasonably predict adoption in the real world. Then such models can be used to

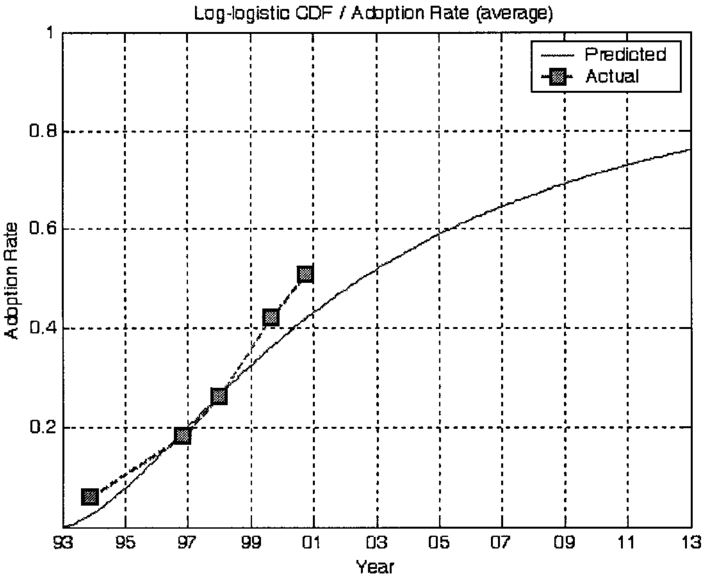

Figure 2.4: Log-logistic Fitted vs. Actual

solve one of the key issues of marketing a new product that is to identify the early adopters and encourage them through incentives or information. Alternatively from a policy perspective in the context of the digital divide, it is imperative to identify the groups in the population who are the least likely to adopt in the near future such that incentives can be better targeted towards them.

It is well known that in general any econometric model fits well to data used to estimate it, since we are interested in the forecasting powers of the models presented, intuitively we want to check for the out-of-sample properties of the estimates. We therefore divide the sample into two halves picked randomly[35] with one half used to estimate the model and other half used for validation of the model. The pooled sample after division leads to an estimation sample of size 110,673 and a validation sample of 110,085.

[35] Random sampling without replacement such that each observation is selected for estimation with probability half.

Table 2.10:
Accuracy of Individual Predictions

Upper Percentile	All Years	Lower Percentile	All Years
0.5	89.7	99.5	29.7
1.0	89.5	99.0	29.4
5.0	86.9	95.0	27
10.0	81.5	90.0	24.3
20.0	73.0	80.0	19.2
30.0	65.7	70.0	14.7
40.0	59.1	60.0	10.6
50.0	52.5	50.0	7.5
60.0	46.6	40.0	5.0
70.0	41.4	30.0	3.4
80.0	37.0	20.0	2.1
90.0	33.2	10.0	0.9
95.0	31.5	5.0	0.3
99.0	30.3	1.0	0.5
99.5	30.1	0.5	0.2

Log-logistic model. Estimation $N = 110,673$.
Validation sample $N = 110,085$

The evidence is presented in table 2.10.[36] The predictive success of the model therefore can be summarized by two statistics, the false positive rate, i.e. how many are predicted to adopt by the model and who do not and similarly the false negative rate defined as the converse. The table is to be read as follows, first the data is arranged in ascending order of adoption probability and for certain percentile values the actual adoption rates are calculated. For example from columns 1–2 of table 2.10 the actual adoption rate for the top percentile of the population (arranged based on predictions by the model) is actually 89.5%. The *false positive rate* can be calculated from this table, given that the adoption rate for the whole sample 30.1%, we arrange the data in ascending order of probability and take the top 30% of the population and

[36] The log-logistic model is used for prediction purposes with all variables except the fixed effects included.

calculate this statistic as 34.3%.[37] Similarly columns 3–4 of the same table can be used to calculate the *false negative rate*, if the data is arranged in descending order of predicted adoption probabilities using the same cutoff as before of the bottom 70% who are predicted not to adopt only 14.7% do. Also we note that the prediction improves over time (not reported) which is expected since the model only explains part of the variation over time.[38]

2.8.2 Forecasting Diffusion Patterns

An attractive feature of duration models is that it allows us to forecast adoption rates at various levels of aggregation once the parameters of the underlying model has been estimated. We consider four dimensions of the digital divide over the next several years and plot the results implied by the full model in figures 2.5 and 2.7. The divide in terms of race has perhaps received the most attention, we find in figure 2.5(a) that this divide remains for the next several years with a $10-15\%$ difference in adoption rates for the Internet among various races, note that there is hardly any difference between Hispanics and blacks although both lag from the population majority. In case of income (figure 2.5(b)) we find that divide actually widens over the next few years before all economic groups in the population approach similar rates of adoption well into the future.

The divide when expressed in terms of education in figure 2.7(a), shows that the difference within the various groups with some college education or higher to be very small and closes fast, although those without a high school degree tend to lag behind them for a while into the future. Lastly the FCC has expressed much concern over the divide between urban and rural areas, we do not find evidence of any such divide

[37] Since from the table among the top 30% of the predicted adoptions 65.7% do and the rest don't.

[38] Given a time trend any model with some predictive power, the fit will improve over time.

(once all other demographic variables are controlled for) either now or developing in the near future.

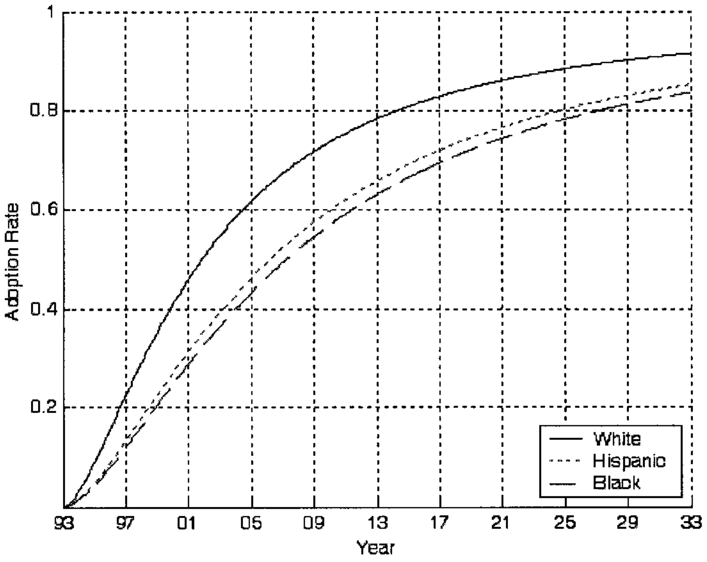

Figure 2.5: Predicted Racial Divide

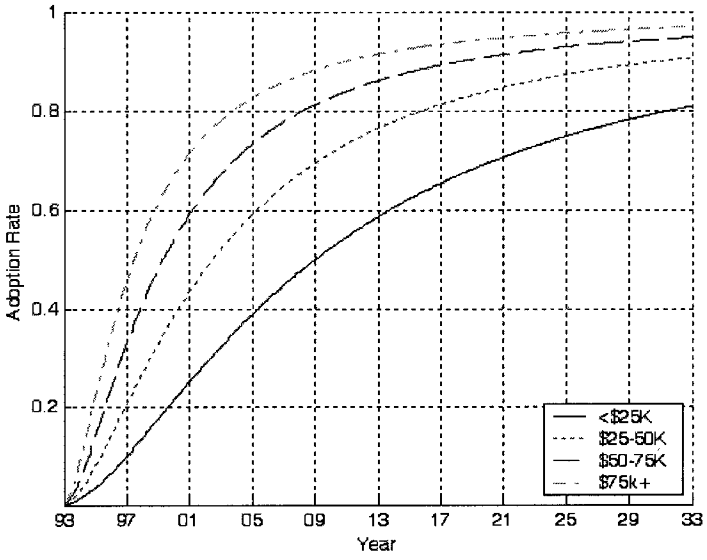

Figure 2.6: Predicted Income Divide

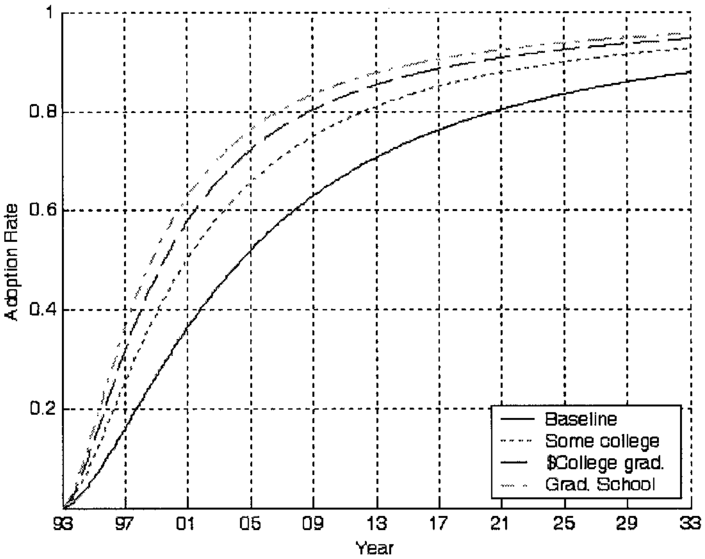

Figure 2.7: Predicted Education Divide

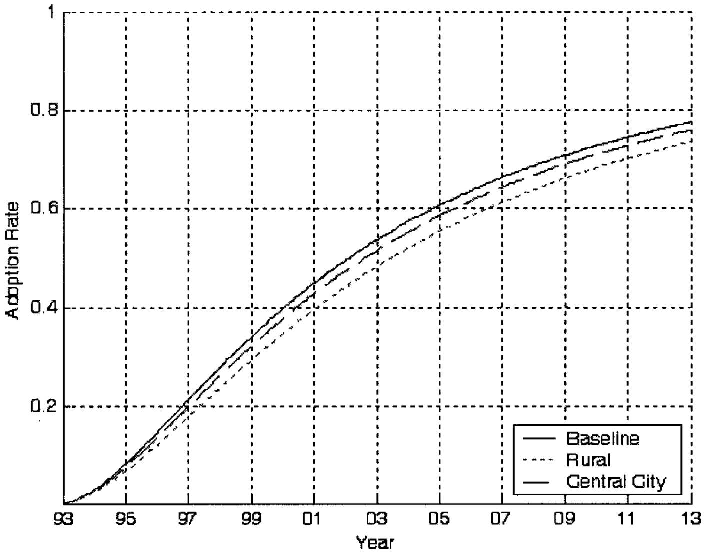

Figure 2.8: Predicted Urban vs. Rural Divide

2.9 Conclusion

The main contribution of this paper is twofold, first I show that a range of duration models can be estimated using repeated cross-sections data. I also apply this methodology to the question of the digital divide, a topic which has generated much controversy in recent times, since significant subsidies have been allocated by the government to bridge this divide. I show that such models can provide a useful heuristic treatment which might be of independent interest (who adopts first etc.), as well as provide forecasts for future adoption levels. Additionally they also allow us to test for any heterogeneity in adoption patterns in the population, both observed and unobserved. To summarize our findings, we find that the digital divide is largely a temporary phenomenon which is forecast to close in the short to medium run by itself, with existing policies. However in the short run differences in access will persist at least for the next two decades, which by itself might be considered significant.

My current work focuses on extending this model to the standard application of duration models, which is *program evaluation.* One of the main features of these models is that they allows us to test for differences in diffusion processes. Therefore one can estimate the impact of programs such as the E-rate program which subsidizes access to the Internet for schools and libraries, in terms of its overall impact on the diffusion process for various socio-economic groups, i.e. in bridging the so-called digital divide. Static models used by other authors are inadequate in this context for reasons discussed before and a dynamic model is called for.

Chapter 3

Testing for Neighborhood Effects

An Ethnicity Based Approach

A number of authors have pointed out the importance of *peer effects* in determining individual behavior. They have drawn attention to the fact that in the real world one often observes pockets of underdevelopment in an otherwise developed region, typically characterized by high crime rates and poverty. Examples of this phenomena are the so called *inner cities* in America. Although there is widespread consensus as well as empirical evidence as to the existence of such phenomena, there is no widely accepted economic theory explaining it. The social network literature seeks to explain such phenomena in terms of peer effects. If individuals living in such an underdeveloped area, interact primarily with others from a similar background / locality, they are more likely to receive the wrong signals (through imitation) and are less likely to receive any beneficial information such as job opportunities, etc. This can lead to an extremely intractable problem of a human development trap. Peer effects in this context offer both a positive and negative lesson for policymakers. On the negative side it implies that any naive policy prescriptions that improve the external situation (such as creating jobs) without addressing this problem directly is likely to be at best inefficient and at worst costly and ineffective. On the bright side this also implies that policies that explicitly account for such peer effects can lead to a proportionately higher impact, through the so-called *social multiplier* effect. For example in order to improve the grades of all students, selective tutoring of certain students might be a cheaper and more effective solution in the presence of peer effects. When compared to more expensive and time consuming solutions such as reducing class sizes, for example see Boozer and Cacciola (2001).

Sociologists have long emphasized such linkages, empirical evidence on how people form opinions suggests that individual behavior is affected by attitudes and behaviors of those around them i.e. their peers or others in their social group. In particular they distinguish two important linkages, first individuals typically wish to conform to group behavior also known as *peer effect*, more broadly this is the role played by culture or conventions in society. Second, individuals learn and are typically influenced by opinions and attitudes of people they interact with. Economic models in numerous fields stress on similar interactions, however the empirical evidence is limited. Notable attempts include those by Evans, Oates, and Schwab (1992) studying the impact of peer effects on teenage behavior and Glaeser, Sacerdote, and Scheinkman (1996) studying the relationship between crime rates on social interactions.

In a seminal contribution Manski (1993) noted that most empirical studies in this field utilize models which are not econometrically *identified*, therefore any conclusions drawn from such models is likely to be spurious. This implies care needs to be taken in interpreting a correlation between individual behavior and group behavior, as evidence of peer effects. In order to prove causality either strong behavioral assumptions need to be made or suitable instruments need to be found. This study follows the second approach of suggesting a set of suitable instruments.

Here we study the technology choices that people make in particular the timing of adoption of a new technology. The implicit assumption is that of incomplete information, i.e. people are not aware of the existence, quality or the price distribution of a new good. Also for such a good the technology and therefore the price (hedonic price – adjusted for quality improvements) might be falling fast over time making it harder for individuals to make an optimal adoption decision. In such a situation we expect people to learn about the new good / technology from their peers or others in their social group. We assume that for ethnic minorities particularly for recent immigrants to this country there exists considerable *clustering*, i.e. they tend to overwhelmingly live in areas where the majority community is of similar ethnic background, the so-called *ethnic enclaves.* Such enclaves are almost always predominantly populated by recent immigrants, due to reasons of language or cultural barriers, hence they stay insulated from the general community at large. Theoretically the relevant social networks for such individuals are well-defined and simply the members of the enclave. This observation is the key identifying condition in this study. Therefore we use ethnicity to construct measures (indices) of the size and quality[1] of each individual's social network.

Given the recent concerns expressed in policy circles regarding the so-called Digital Divide, which is defined as certain groups in the population not having access to new technologies such as the Internet, the presence of peer effects in the diffusion of new technology has certain strong implications. Similar to the grades example mentioned earlier, any policy that improves external situations without taking this into consideration is likely to be an expensive failure. For example the subsidies provided through the *Internet Tax Freedom Act (1998)*[2] may be the least effective

[1] By quality we mean how good and reliable is it as a source of information in general.

[2] See Hausman (1999) for the expenses involved in administering such a program.

way of achieving universal access to new technologies. A much better option might be selective tutoring of certain individuals in the community. Programs such as Microsoft's donations of computers and software to schools in poorer neighborhoods have a much better chance of success compared to centralized federal government programs.

3.1 Related Studies

This paper largely follows the methodology laid out in Bertrand, Luttmer, and Mullainathan (2000) (henceforth BLM). BLM used census data to study welfare usage among various ethnic groups. Social networks were identified by the language spoken at home. Borjas (1995) studied the role of social networks via ethnicity on intergenerational mobility of skills. He utilized CPS data as here, where the language spoken at home is not available, therefore he used national origin and restricted the sample first and second generation immigrants. This is the approach taken here.

However the main difference from the BLM paper is that the selection bias works in our case in the opposite direction, in their case the selection bias was positive i.e. they were more likely to find evidence of neighborhood effects when none existed. Whereas in our case the selection bias is negative in the sense that we are likely to find no evidence of such effects even if they existed.

There is a large theoretical literature taking such neighborhood level interaction into consideration. For excellent surveys of this literature refer to Brock and Durlauf (2001b) and Moffitt (2001). Additional game theoretical models with a similar theme has also been proposed by a number of authors such as Banerjee (1992) (herd behavior) and Bikhchandani, Hirschleifer, and Welch (1992). Other notable papers in the game theory literature on learning include Bala and Goyal (1995) and (1998).

The empirical literature by comparison is small and growing, notable contributions include Glaeser, Sacerdote, and Scheinkman (1996), studying the impact of peer groups on crime rates.

3.2 Model

3.2.1 Standard Specification

The econometric model that we are interested in is as follows, we follow the notation set out in Manski (1993). Let y be the outcome variable and it is usually binary (takes one of two values) which might be adoption of a new technology, teenage pregnancy or dropping out of high school. Let x be a set of individual attributes that determine which reference group an individual belongs to it might be demographic (for example race, background) or economic (example family income). Other factors that affect outcome y are summarized by z, for example socio-economic status and ability (proxy measures are used when the relevant variable is unobserved) and u is an independent error term (that is not correlated with (x, z). The main equation we are interested in estimating is as follows

$$y = \alpha + \beta E(y|x) + x'\delta + z'\eta + u \tag{3.1}$$

where the coefficients $\theta = (\alpha, \beta, \delta, \eta)$ are to be estimated. The existence of neighborhood effects can be tested in this setup by the hypothesis $H_0 : \hat{\beta} = 0$.

Manski (1993) distinguishes between three distinct types of neighborhood effects in this setup. This distinction is important since they have very different implications for the policymaker, and is directly relevant to our study therefore we mention them here,

Endogenous effects: this is the propensity of individuals to conform to the majority behavior in the group.

Exogenous (contextual) effects: when individual behavior depends on group behavior through the exogenous characteristics of the group.

Correlated effects: individuals behave in similarly fashion due to the fact that they have similar characteristics and / or face similar institutional environments.

Note that only *endogenous effects* have implications for the policymaker and the other two effects do not, since changing the behavior of the group on average does not affect individual behavior. Manski goes on to show that endogenous effects can only be identified if the econometrician is willing to impose considerable structure on the model i.e impose severe restrictions on the data generating process. He calls this problem of identification of the true model in this context as the *reflection problem.* However, Brock and Durlauf (2001b) show that the situation is not as hopeless as it appears to be since Manski's criticism only holds for linear models and for a large group of non-linear models (for example discrete choice models like Logit, Probit etc.) neighborhood effects are indeed identified under fairly general conditions.

3.2.2 An Alternative Approach

Goolsbee and Klenow (2002) investigates the same problem using data on a single cross-section of about $100,000$ households with self-reported date of purchase of a computer. Using this data they generate a multiyear retroactive sample of households. They estimate a linear probability model which can be written as follows,

$$y_{ijt} = \alpha + \beta \bar{y}_{j,t-1} + X'\gamma \tag{3.2}$$

where y_{ijt} is an indicator variable for individual i in location j at time t (they restrict their sample to the pool of potential adopters i.e. those who have not bought a computer yet), and $\bar{y}_{j,t-1}$ is average level of usage (ownership) in area j at time $t-1$. The intuition being that the higher the ownership of computers in a particular locality, the more information is available to the consumer and therefore the higher the likelihood of adoption of the new technology. They use regressors X which are demographic controls like income, age and education of the household. They deal with unobserved variables like technological sophistication using instruments like ownership of other technology goods like cd-players, survey question on attitudes toward technology etc. They find positive network / learning effects in the diffusion of computers.

We argue that their model is not identified, heuristic proof is provided here. Their study suffers from the problem of a simultaneity bias. Since a diffusion phenomenon has been observed for most new technologies, leading to an S-shaped adoption curve, a positive coefficient always arises in their setup. Note that there are two common explanations for diffusion, first learning from others as emphasized by the authors, and second due to improvements in quality and / or falling prices.[3] Both would show up as a positive coefficient (β) in equation 3.2 above. They use various instruments to control for unobserved variables, but we claim that even if there is no omitted variable bias, given the simultaneity implied by the diffusion curve, a positive β cannot be treated as conclusive evidence of the presence of neighborhood effects. In other words even if there were no differences between the MSAs (i.e. all observed and unobserved factors were perfectly controlled for), if the diffusion phenomenon was entirely driven by improvements in quality and / or falling prices (i.e. no learning or

[3] If individual valuations are distributed as a normal curve and quality improves linearly an S-curve results, similarly for a rectangular distribution of consumer preferences and a non-linear improvements in quality / prices.

network effects whatsoever) a positive coefficient would still be observed provided all MSAs were on the initial concave part of the S-curve (which would be the case since typically the inflexion point occurs at 50% adoption rate for a symmetrical curve and adoption rates for PCs were definitely less than that for the sampling period).

We believe a more robust evidence of the existence of neighborhood effects is obtained through a more direct approach, i.e. when social networks can be clearly identified. Since this is not possible for the general population at large we settle for certain ethnic groups. However the tradeoff here is that since we deal with relatively small minorities in the population their behavior may not be indicative of the larger population, i.e. it is possible albeit unlikely that neighborhood effects are strong for ethnic minorities but relative unimportant for the general population.

3.3 Data Description

3.3.1 Sources

The data used for this project was describer earlier in chapter 1 above. We also obtained data about the demographic characteristics of the states and metropolitan areas (MSAs) like population, ethnicity of the residents etc. from various publications of the Census. Some of the data is from the recently concluded 2000 Census whereas other variables were obtained from earlier publications like the Economic Census of 1997 etc. For a detailed analysis of computer and Internet usage across various demographic groups, see NTIA (2000).

Table 3.1:

Summary Statistics for chosen Ethnic Groups

Variable	Mean	U.S. Mean (MSAs Only)
Age	41.89	42.47
Male	0.4834	0.4762
Education		
No School	0.058	0.0118
Upto High School, No Degree	0.4245	0.2571
High School Degree	0.1873	0.2736
Some College	0.1781	0.2611
College Degree or more	0.1521	0.1964
Married, spouse present	0.5403	0.5335
Married, spouse absent	0.0489	0.0184
Widowed	0.0304	0.0234
Divorced	0.0571	0.0841
Separated	0.0709	0.0714
Never married	0.2524	0.2692
Children present	0.3790	0.3554
Foreign Born	1.00	0.1361
Years since entry (foreign born)		
0-5	0.1983	0.0276
6-10	0.1754	0.024
11-15	0.1423	0.0193
16-20	0.1221	0.0162
21+	0.3618	0.049
Speaks no/poor English	0.2967	0.0396
Born abroad of American parents	0.0502	0.0073
Naturalized citizen	0.3575	0.0511
Not a citizen	0.5193	0.0702
Observations	576,476	6,316,668

3.3.2 Summary Statistics

In table 3.1 we compare our sample with the general sample in the CPS. We find a similar age distribution and a similar gender composition as the general U.S. sample. When it comes to education we find that on an average the ethnic sample we deal with is less educated with around sixty one percent of the householders having either a high school degree or at least some schooling, the corresponding figure for the general population is fifty three percent.

Other statistics on the structure of the household indicate that more of the immigrant population are married with spouse absent compared to the general population and also the marriage rates are higher though the difference is small. Also they have a somewhat lower divorce rates and more households have children in them (again the differences are small). Only thirteen percent of the U.S. population are foreign born whereas our sample consists entirely of people born elsewhere. The rates of immigration (as can be seen from the data on years in the country) have been roughly constant with a slight rise in the last five years. A large section of the population speaks either no English or very poor English. The citizenship data indicates that in our sample we have primarily non-citizens or naturalized citizens and very few people born abroad of American parents (5% only).

In table 3.2 we report the contact availability as defined above by ethnicity at the MSA level. The first two rows give the summary statistics for the whole sample. As noted above given our definition of contact availability if individuals were evenly distributed across the country this figure would be close to one. However we find evidence that individuals from the chosen ethnicities tend to congregate in particular neighborhoods. [4] One expects that people with first language English or who are

[4] An alternative explanation might be that those living in highly concentrated neighborhoods are likely to be sampled more which implies a certain *sample selection bias* therefore the contact availability of the sample would be higher than any random sample chosen from the population of

reasonably fluent in English will mingle better with the general population and that is what we find in the data. The lowest contact availability figures are for the English, followed closely by the Canadians and Indians in the sample. The surprising result is about the Germans living in very diversified localities. The highest level of concentration we observe is for the Cubans (everyone ends up in Miami?) and similarly for people from the Dominican Republic and Haiti.

Tables 3.3 and 3.4 report the ownership of computers by ethnicity. In table 3.3 we consider only people born outside the country and non-citizens whereas in 3.4 we consider second generation Americans i.e. households where either parents of the householder were born outside the country. We find that for first generation Americans and immigrants (non–citizens) computer ownership varies a lot by ethnicity and it is substantially different from the U.S. average value. For the purpose of this table we pool together all the different waves of the CPS and we find the U.S. mean ownership of computers is around 52% whereas for people of other ethnicities it goes from a low of 21% for Mexicans to a high (significantly higher than the U.S. mean) of 78% for Indians. This shows that the mean quality of the network i.e. the average information that a person gets from her network is not the same across all ethnicities, this is crucial towards the identification of the model. However this also raises the serious question of selection bias which we will discuss in more details later. Whereas from table 3.4 we find that for second generation Americans the difference with the mainstream population is smaller but not insignificant for example an average household with a second generation polish householder is likely to have a computer only 30% of the times compared to the U.S. mean of 53% and similarly for Italian households (36%) and Mexicans (36%) respectively[5]. Note that our findings

people of each ethnicity.

[5] On the positive side some ethnicities far outstrip the mainstream population in terms of computer ownership like the Philippines (67%), however the number of observations make this estimate

follows those of Borjas (1995) who finds that for labor earnings the distribution for second generation immigrants is much closer to the general population compared to the first however differences do persist at least for certain ethnicities. This raises the puzzle in their paper as well as here why do certain ethnicities assimilate faster than others even after controlling for other demographic variables like income and education (as reported later).

3.4 Econometric Specification

The dependent variable we use in our analysis is $comp_{i(t)jk}$. Where each wave of the survey is indexed by t so if there are T waves $t \in \{1, 2, \ldots, T\}$. Following convention used in the unbalanced panel / repeated cross-section literature we denote household i observed in wave t as $i(t)$ so if there are $n(t)$ observations (which might vary across t) in wave t then $i(t) \in \{1, 2, \ldots, n(t)\}$. Also let the population be divided into K mutually exclusive communities or particular ethnicities (sets) indexed by $k \in \{1, 2, \ldots, K\}$, similarly let there be J geographic areas in the sample indexed by $j \in \{1, 2, \ldots, J\}$. Then our dependent variable is a binary variable, which is an indicator variable taking the value one if the household owns a computer and zero otherwise, for the $i(t)$ household observed in wave t belonging to ethnicity k and living in area j.

Our primary regressor variable of interest is constructed in two steps. First as mentioned above we define the *quality* of the network i.e. the information that one can expect from the average person belonging to that particular ethnicity. This is measured here as in BLM by $\overline{comp}_{kt}$ which is average computer ownership of each ethnicity k with the mean taken across all geographic regions for each survey date t.[6]

not very robust.

[6] Local effects are controlled for by regional dummy variables which enter as other regressors in

Table 3.2:

Contact Availability at MSA level by Ethnicity

MSA CA	Mean	Std. Dev.	Min.	Max.	Obs.
Sample	5.51	9.709	0.004	52.085	576,476
Log MSA CA	0.949	1.24	-5.641	3.953	576,476
Puerto Rico	4.281	3.117	0.011	11.545	42,125
Germany	1.225	0.626	0.072	4.189	27,696
Italy	2.713	1.832	0.02	8.186	26,443
Poland	3.316	3.385	0.031	24.301	16,352
England	1.222	0.472	0.068	2.572	17,668
Russia	2.666	1.938	0.023	5.733	13,005
China	4.085	4.137	0.018	13.74	23,109
India	1.881	1.432	0.025	9.472	17,982
Japan	2.28	2.799	0.032	12.363	14,435
Korea	2.386	1.565	0.022	5.778	19,510
Philippines	3.74	3.346	0.011	13.241	40,421
Vietnam	3.471	3.312	0.021	10.634	22,279
Canada	1.713	1.29	0.067	10.665	29,631
El Salvador	5.495	3.432	0.005	8.861	18,488
Mexico	4.226	2.629	0.004	11.154	168,855
Cuba	32.954	24.215	0.009	52.085	34,448
Dominican Republic	11.438	6.622	0.006	34.26	12,487
Haiti	9.476	6.929	0.011	20.713	8,010
Jamaica	5.348	3.533	0.037	9.858	11,859
Colombia	5.18	4.589	0.015	13.515	11,673

Table 3.3:

Computer Ownership by Ethnicity

(First generation/foreigners)

	Computer in Hh.			
Country	Yes (%)	No (%)	Obs.	% of Sample
USA	52.51	47.49	42,164	88.44
Puerto Rico	25.33	74.67	225	0.47
Germany	58.72	41.28	218	0.46
Italy	40.91	59.09	110	0.23
Poland	42.86	57.14	98	0.21
England	61.26	38.74	111	0.23
Russia	44.44	55.56	72	0.15
China	63.75	36.25	160	0.34
India	78.61	21.39	187	0.39
Japan	62.5	37.5	96	0.2
Korea/South Korea	61.29	38.71	124	0.26
Philippines	58.74	41.26	223	0.47
Vietnam	40	60	100	0.21
Canada	52.41	47.59	187	0.39
El Salvador	30.22	69.78	139	0.29
Mexico	21.01	78.99	1104	2.32
Cuba	35.56	64.44	180	0.38
Dominican Republic	25.95	74.05	131	0.27
Haiti	44.58	55.42	83	0.17
Jamaica	53.72	46.28	121	0.25
Colombia	45.33	54.67	75	0.16

Table 3.4:
Computer Ownership by Ethnicity
(Second Generation)

Country	Computer in Hh.		Obs.	% of
	Yes (%)	No (%)		Sample
America	53.36	46.64	38072	90.3
Puerto Rico	50.66	49.34	152	0.36
Germany	48	52	350	0.83
Ireland	52.05	47.95	146	0.35
Italy	36.01	63.99	572	1.36
Poland	29.46	70.24	224	0.53
Sweden	38.46	61.54	78	0.18
England	57.06	42.94	170	0.4
Scotland	50	50	70	0.17
Russia	43.94	56.06	198	0.47
Japan	50.65	49.35	77	0.18
Philippines	67.14	32.86	70	0.17
Canada	48.54	51.46	410	0.97
Mexico	36.52	63.48	356	0.84

Ethnicity coded as (a) country father was born in, or
(b) country mother was born in if father US-born

To motivate this measure consider this simple setup with a completely homogeneous population (of K ethnicities), with a proportion p_k using the technology. Only users of the technology have the following information, they know whether it is better than other alternatives available in the market or not. The probability that the average person sampled from this population will have this information is p_k. With a homogeneous population if we form a random sample of identical individuals the probability that we will get this information is proportional to the size of the sample.[7]

Therefore in the second step we define the *quantity* of the network as CA_{jk} which is a measure of the *contact availability.* By contact availability we mean the average number of contacts that a person might have, for example an Italian American living in a neighborhood dominated by Italian Americans is likely to have a far larger social network than one living in say an Irish American neighborhood. The indexing denotes the fact that this measure only varies across ethnic-MSA cells since we use census data to construct this measure for each MSA by ethnicity and the census is conducted only every ten years, a more appropriate measure is hard to find that is closer to the survey dates. At the time of this study the 2000 census had been completed but the data files were not publicly available. As mentioned before the identification of our model (just as in BLM) of estimating true learning effects strongly depends on how isolated each ethnic/language group is from the mainstream, by that we mean how often and how closely do they interact with others in society who do not belong to their community as defined by ethnicity or language. For example this is more likely to be true for new immigrants or people who do not speak English very well. We interact the quality and quantity measure to estimate

the estimated equations.

[7] In this setup we know that random variable x is distributed as a binomial distribution, therefore we know that the random variable $y = \sum_{i=1}^{N} x_i$ formed from a sample of $\{x_1, x_2, \ldots, x_N\}$ observations, is distributed as a Bernoulli distribution with the probability of exactly n successes defined as $P(y = n) = p^n(1-p)^{N-n}$. Therefore a larger N increases this probability for all n i.e. $dP/dN > 0 \quad \forall n, N$.

the average information variable. The baseline model considers a multiplicative model however we do experiment with other specifications as well.

If we assume a linear probability model[8] then the equation we estimate is as follows,

$$comp_{i(t)jk} = \xi + (CA_{jk} * \overline{comp}_{kt})\alpha + X_{i(t)}\beta + \phi_{i(t)jk} + \gamma_j + \delta_k + \tau_t + CA_{jk}\theta + \epsilon_{i(t)jk} \quad (3.3)$$

Since we pool together the data to form a reasonably large sample for hypotheses testing and therefore from now on we shall suppress the time subscript given that we do not have a true panel dataset but instead repeated cross-sections of observations at different time periods. The time trend or component will be captured by the time dummies τ_t. However this raises the problem of *path dependence*, see for example Arthur (1989) since with new goods and technologies a diffusion process is often observed (see discussion above). Thus making it hard to assume that the sample was obtained from the same population for each of the survey years.

In equation (3.3) above $X_i(t)$ is a set of demographic and economic controls like income, education, type of household etc. for individual $i(t)$. We assume that the error term is $\epsilon_{i(t)jk} \sim N(0, \sigma^2)$. Note that the above specification assumes a particular structure for the unobservable household level fixed effects. Similar to a variance components model, we assume that this effect can be decomposed into a year effect common to all households surveyed in a given year for example this might be a time trend (linear or non-linear) or any idiosyncratic shock to the diffusion process as long as it is common to all individuals $i(t)$, we measure this by τ_t. Second, we assume that there are unobservable factors that are common to particular MSAs measured by the MSA level fixed effects γ_j and these are time or invariant and the

[8] In the linear probability model the dependent variable is $E(y) = Prob(y = 1) = x\beta + \epsilon$ where x is a set of regressors and ϵ is an iid error term.

same for all individuals across waves, for example the preponderance of high-tech firms in and around San Francisco implies residents would have a higher probability of having a computer compared to almost any other locations in the country. The third component of the unobservable factor is an ethnic level fixed effects already discussed before that some communities might be more receptive to new technology because of their culture or other reasons, these are measured by the dummy variables δ_k. Note that these variables are also fixed over time and the same for all individuals surveyed, across the years. Although extensive this does not cover other more complicated situations involving the interaction terms of these fixed effects, say for example ethnic Chinese living in New York might be very different from those living in Houston in ways that are unique across ethnicities, i.e. the MSA and ethnic fixed effects do not entirely capture this. However given the large number of ethnic–MSA cells and the limited number of observations for each it is almost impossible to estimate these interaction terms and we accept this as a shortcoming of our study. The only way around this problem is to assume random effects i.e. these effects are jointly distributed across the population which is known and whose parameters can be estimated. We discuss this issue in greater details below.

Last we include the contact availability term separately to control for *selection bias.* Let $\phi_{i(t)jk}$ be the household level unobserved (by the econometrician) factors that affect the household's technology decision but which are not captured by the group (cell) level dummy variables included above like year, ethnicity etc., i.e. each household is unique in it own way apart from belonging to a particular ethnicity and so on. In our analysis we consider residential choice and thereby the size of the social network as exogenous or at least not correlated with the unobserved variables conditional on $X_i(t)$. If individuals choose where to live based on economic reasons or demographic reasons that we control for like income, say a poor person is more likely to live in a poorer neighborhood, then sorting is based on income and given that we include income in our regressions below there is no bias in the estimates for neighborhood effects. However if people choose where to live based on some unobserved variable say something like technological sophistication. Then the error term depends on $X_i(t)$ and the estimates obtained are biased. To test and control for this we include contact availability separately since if the factors that affect residential choice (and therefore contact availability) are the same as those that affect residential choice then the OLS estimates of the equation above are biased. Therefore a test for selection bias would be to test whether $\theta = 0$. Note that this argument is far easier to make for new technology rather than welfare choice. The case we are concerned about is the following, say Indians who migrate to this country predominantly are either software professionals a disproportionate amount of whom end up in Silicon Valley and the rest or cab drivers who settle all over the country. Even after controlling for income and education assuming the cab drivers make similar amounts and also have a college degree. Then the correlation between living in a high contact area like Silicon Valley and owning a computer is spurious and the underlying factor that affects both is technical training or profession. In our analysis below we do control for profession and this seems to make a difference. Therefore in some sense this last term CA_{jk} measures the size of the selection bias. In this setup the presence of learning or neighborhood effects is tested by the following $\alpha \neq 0$, otherwise social effects do not exist.

Measures of contact availability: A simple measure of contact availability would be C_{jk} which is the number of people of ethnic origin k living in area j at the time of the study. However a better measure of contact availability would normalize this measure for the size of the region considered i.e. the total population A_j of region j at that time. The argument for this derives from geographic proximity, since people living in areas with a low population density, for example someone living in a sprawling MSA spread over a large geographic area is less likely to come into contact anyone in particular anyone within their social network even if that person has a very large network a priori. Since measures of population density at the MSA level or PUMA level cannot be found we use the total population instead the reasoning being that the census constructs MSAs such that they have the same geographic area. Therefore total population proxies for population density in our analysis. Therefore our second measure of contact availability is $CA_{jk} = C_{jk}/A_j$. In our analysis we are particularly interested in differences from the mean values of contact availability, i.e. what happens to someone living in an area with higher/lower than average density of people of the similar ethnic origin. Thus we normalize our contact availability measure using the average size of that ethnicity in the U.S. population L_k/P_{US}, where L_k is the total number of people of ethnic group k living in this country and P_{US} is the total population of the United States. Consider the counterfactual if individuals of all ethnicities were randomly distributed across the country then this measure would be one for all individuals. For any region j the proportion of people of ethnicity k would be exactly L_k/P_{US}. So according to this measure if CA_{jk} is greater than one then one is living in an area with higher than average social contacts and vice versa. Since the deviation from the average tends to be very small in much of our analysis we consider instead the logarithm of this measure, i.e. we define contact availability for person i as follows,

$$CA_{jk} = ln\left(\frac{C_{jk}/A_j}{L_k/P_{US}}\right) \tag{3.4}$$

As mentioned above since the CPS our primary source of data does not have information on language spoken at home we use the country of origin approach to ethnicity, i.e. we define ethnic and social groups by which country the particular person came from. Therefore first we restrict the sample to recent immigrants and naturalized citizens to this country, all people born outside the country of parents who were not American citizens. We drop all observations for households not living in MSAs due to two reasons, first the number of observations (since most recent immigrants tend to live overwhelmingly in urban areas) do not allow for estimation of region specific fixed effects. Also the availability of both information and the product (PCs) outside the metropolitan areas is a serious question on which we have no data on. We use data from the CPS we calculate the average ownership of computers for each group j at time t ($\overline{comp}_{kt}$). When deciding which ethnic groups to include in this study the binding constraint was calculating this measure of the quality of the network for each of the survey years. Since we are using census data to construct the quantity measure this could be done at a fairly disaggregated level (MSA or PUMA) by ethnicity for most regions across the country.[9] However the CPS sample for immigrants is small enough that when it is decomposed by ethnicity and survey year, only for a certain number of ethnic groups can this measure be calculated with reasonable certainty. This process yields about twenty ethnic groups for which we have sufficient number of observations to calculate the quality of the network for each of the survey years. These are interacted to form the primary variable of interest. We report the results of our estimation in the next section.

[9] Unfortunately the CPS does not provide data disaggregated at the PUMA level which are smaller than MSAs, my enquiries at the BLS met with no success

3.5 Estimation

3.5.1 Differences–in–Differences

We start of with the basic differences-in-differences estimator commonly used in other studies. We can divide the sample into two groups those belonging to ethnicities with higher than median (across ethnicities) usage of computers and those with below median usage. Similarly we can divide the sample into two groups based on those living in areas with higher than median contact availability and those living in below median areas. In this strategy taking the difference between the high and low computer usage is equivalent to using ethnic fixed effects and similarly the difference between the two groups based on contact availability is similar to using fixed effects controlling for above average contact availability. The difference of the two differences gives the coefficient on the interaction term i.e. if a person belongs to an ethnicity with higher than median computer usage and lives in a an area with higher than median contact availability is she more likely to have a computer. Table 3.5 below uses the sample from one wave of the CPS i.e. 2001 to do this calculation the logic being that individuals across waves may not be easily comparable if there is an underlying diffusion process taking place. The table is organized as follows: the rows report the mean of the dependent variables for people belonging to ethnicities with lower than median and then higher than median usage of computers and the columns represent mean of the dependent variable across people living in high contact availability areas and low ones compared to the median respectively. The fourth column gives the differences between the columns and standard errors are reported in parentheses.

We find that for people belonging to ethnicities with below median computer usage being in a high contact availability area helps whereas the reverse is true for eth-

Table 3.5:
Difference–in–difference
(2001 sample only)

	Low CA	High CA	$\triangle CA$
Low Computer Use	0.3493	0.3629	0.0136
	(0.0803)	(0.0756)	(0.1103)
	$N = 812$	$N = 1072$	
High Computer Use	0.6927	0.6853	-0.0136
	(0.0638)	(0.0704)	(0.095)
	$N = 736$	$N = 483$	
Diff-in-diff			-0.021
Estimates			(0.1456)

Table 3.6:
Difference–in–difference
(Sample 1997,1998, 2000 and 2001)

	Low CA	High CA	$\triangle CA$
Low Computer Use	0.2699	0.2581	-0.0118
	(0.1196)	(0.1131)	(0.1646)
	$N = 2955$	$N = 4032$	
High Computer Use	0.5902	0.5687	-0.0215
	(0.1093)	(0.1321)	(0.1715)
	$N = 2698$	$N = 1646$	
Diff-in-diff			-0.0097
Estimates			(0.2377)

nicities with higher than median computer usage. Therefore taking the differences-in-differences we find that the interaction term is negative however not significant. There are several explanations for this phenomenon first, we do not have enough data to draw the right conclusions this leads to the next table. Alternatively we can think of their being a saturation point some ethnicities have reached a point in their diffusion process where they already have all the information they need to take the right decision, whereas for those ethnicities still on the rising part of the S–curve additional information matters therefore being in a high contact availability area helps. Third, this estimation procedure does not include any other controls in the regression and uses only ethnic and contact availability fixed effects and given the discussion on selection bias (see above) it is not surprising to find insignificant results. This gives us hope that adding more controls might partially offset this problem. Next in table 3.6 we present the evidence for a similar estimation procedure using data from four of the five waves in the CPS data[10]. We find here that even for people belonging to ethnicities with lower than median usage being in a high contact availability area hurts, this as before might imply two things either there are no learning effects or the selection bias is extreme to the point of obfuscating the learning effects, i.e. particular types of people (unlikely adopters due to socio-economic reasons) live in neighborhoods with a higher concentration of similar people so the additional information from higher contact availability does not help. However we do find in both tables that the coefficients are not significant by far, and also that using additional data lowers the coefficients (as one would expect with selection bias) towards zero and also makes them less significant.

[10] Data from 1994 was excluded as it was felt to be too early in the diffusion process to be comparable to the other waves.

3.5.2 Basic Results

Our primary results are summarized in table 3.7 below. Given our discussion from before if adoption of new technology does follow a diffusion process then observations from different survey waves might not be easily comparable, particularly so if the time interval between the observations are large and uneven. Therefore we drop the observations from the 1994 wave in our analysis below, since it was too early on in the diffusion process for personal computers. The PC was only just beginning to become popular with the coming of the Internet (1991) and the World Wide Web (1992). We therefore only use observations from the years 1997,1998, 2000 and 2001 in our analysis below, given our belief that these are roughly comparable since they are closer to each other in time and are roughly equi-spaced (with approximately a year between each survey wave). Note that with data that is equi-spaced in time it is relatively easier to control for the stage of the diffusion process using time dummies or a trend variable.[11] Below we consider the linear probability model of computer ownership at the household level discussed in the last section.

In steps we add dummy variables controlling for fixed effects by ethnicity, the year the survey was conducted (stage of the diffusion process), years since entry into the country and MSA level geographic dummies respectively. The implicit assumption being that there are unobserved factors at these levels that needs to be controlled for to get an unbiased estimate of the neighborhood effect. These factors are assumed to be the same for all individuals in the same cell. The variable that measures neighborhood effects is reported in the first line which is an interaction of the quality and quantity of the network available to an individual, the coefficients reported are mean-adjusted, meaning they actually report the coefficient for $CA*(\overline{comp}_{kt}-\overline{comp}_t)$

[11] Although the overall process might be non-linear in time over short intervals it might be linear provided the intervals in data are small relative to the overall time taken by the diffusion process.

which makes the interpretation of the coefficients easier for the other CA measure. We add CA in the regression to control for any additional selection bias and the coefficients are reported in the second line. We control for the sex, age, education of the householder and household income. Male is a dummy variable for the head of the household being male. We use three dummies for education with the baseline case being less than a high school degree. These are having a high school degree, some college and graduation from college or graduate studies respectively. We include three dummy variables for income with the baseline case being a family income of less than $\$25,000$.

We first report the coefficients with only the demographic controls mentioned before and dummies controlling for the ethnicity of the householder. Since we use data on individuals belonging to twenty ethnicities in the sample we use nineteen dummy variables with the baseline ethnicity being Puerto Rican. In column 1 we report the results for this specification, we find that the measure of social effects is positive and significant at one percent level of significance. As expected the CA variable separately has a negative sign and is highly significant, this signifies a negative selection bias with coefficients on the social effects biased towards zero. This is true of all the different models estimated although the significance level varies according to specification (it is significant at the five percent level most of the times). This provides some evidence that people sort to some extent according to some unobservable factor that is positively correlated with the availability of contacts and negatively to the use of computers. One possible candidate might be unobserved ability or human capital /skill etc. Later on we use industry level dummies to control for this selection bias, which works if individuals sort into professions according to this unobserved ability. If the preference ordering across industries are similar across individuals and employers observe this unobserved ability or skill then we can expect some industries

Table 3.7:
Linear Probability Model: Sample*

Computer (dummy)	1	2	3	4	5
$CA_{jk} * comp_k$	0.035	0.008	0.009	0.039	0.009
	(0.000)	(0.060)	(0.047)	(0.000)	(0.068)
CA_{jk}	-0.011	-0.003	-0.003	-0.013	-0.004
	(0.000)	(0.057)	(0.038)	(0.000)	(0.030)
Male	0.005	0.011	0.012	0.004	0.012
	(0.514)	(0.157)	(0.123)	(0.581)	(0.121)
Age	-0.003	-0.003	-0.004	-0.003	-0.004
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
High School Degree	0.059	0.06	0.06	0.056	0.056
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Some College	0.213	0.212	0.209	0.202	0.198
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
College and more	0.279	0.275	0.281	0.267	0.27
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Income					
$\$25,000 - 50,000$	0.176	0.168	0.158	0.176	0.158
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\$50,000 - 75,000$	0.316	0.303	0.288	0.315	0.287
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\$75,000$ and more	0.407	0.387	0.367	0.406	0.366
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ethnic dummies (19)	Yes	Yes	Yes	Yes	Yes
Year dummies (3)	No	Yes	Yes	No	Yes
Year of entry into US dummies (15)	No	No	Yes	No	Yes
MSA dummies (205)	No	No	No	Yes	Yes
R^2	0.329	0.341	0.346	0.347	0.363
N	11,329	11,329	11,329	11,329	11,329

*Sample: October 1997, December 1998, August 2000 and September 2001
$comp_k$ is the mean usage of computers in ethnic group k across the US
(P-values in parentheses, using robust standard errors)

Table 3.8:

Linear Probability Model (Cont.)

Computer (dummy)	$CA_{jk}*$ $comp_k$	p–val.	CA_{jk}	p–val.
(1) Baseline model with all dummies from before (all dummies)	0.009	0.068	-0.004	0.030
(2) Add primary industry dummy to (1) [50]	0.017	0.007	-0.007	0.002
(3) Add occupation dummy to (1) [44]	0.019	0.003	-0.007	0.001
(4) Use primary occupation group dummies in (1) instead: [12]	0.019	0.003	-0.007	0.001
(5) Use consolidated occupation group dummies instead in (1): [3]	0.019	0.002	-0.007	0.001
(6) Add interaction terms to (1) (ethnic and yr. of entry)	0.009	0.056	-0.004	0.024
(7) Add interaction term to (5) (ethnic and yr. of entry)	0.019	0.003	-0.007	0.002
(8) Add interaction term to (5) (yr. of entry and Occ. group)	0.019	0.003	-0.007	0.001
(9) Add interaction term to (5) (Occ. group and ethnicity)	0.017	0.007	-0.007	0.002

Pooled sample, all years, $N = 8,231$.

*Robust Standard Errors, [n]–# of dummies

All dummies refer to the head of the household to have predominantly high ability workers who might be able to afford to live in better neighborhoods with a lower level of contact availability or few people from a similar ethnic background (based on our discussion on residential segregation from before). Then controlling for which industry the person works in controls for this unobservable factor in the regression estimates. We do find some evidence as reported below that this might indeed be the case.

We also find that it does not make a difference whether the head of the household is male or female since the coefficient is positive but insignificant always. Also households with older heads are less likely to own computers although we find the magnitude of this to be small. The results on education of the householder is intuitive the more educated he or she is the more likely it is that the household will have a computer. We obtain a similar result for income, families with higher family incomes are more likely to own computers. We also find that most of the ethnic dummy variables are highly significant (not reported here).

Next we add more controls to our baseline model, first we consider whether there is a difference between the years by adding dummy variables that control for any year specific effects τ_t in equation (3.3). We find that these effects are significant (as one would expect with a diffusion process), note that this provides further evidence against the approach taken by Goolsbee and Klenow (2002). Our primary variable for social effects is still positive but only significant at the ten percent level of significance (p-value of 6%). There are two explanations first maybe due to the paucity of data standard errors are not being estimated efficiently. Also there is the possibility that we need to add more controls. In column 3 we further add dummy variables controlling for years in the U.S. (altogether sixteen dummy variables are added). The intuition for this is that the longer a family has been in this country the more time they have had to assimilate with the mainstream, learn the language and culture of the place and generally feel more comfortable and therefore are likely to interact more with the general population and less with people of similar ethnicity, therefore the impact of social effects might not be as pronounced as for fresh immigrants to the country. We find that the coefficient on the social effects variable does not change but becomes significant at the 5% level, with this change in the model.

Given the limited amount of data we have (only a few observations for each MSA) it is highly problematic to estimate fixed effects for each MSA. So in the next column (4) we only consider the ethnic dummies and add all the MSA dummies, all 205 of them into the model. Surprisingly we find that this does not make a difference to our basic model reported in column 1 since the coefficients are very close and also highly significant (at 1% level). Thus emboldened we try to add all dummy variables we have considered this far i.e. for ethnicity, year of survey, years in the country and also for every MSA. We get more or less the same result that we got for our second and third case above, however now the coefficient is less significant (p–value of 6.8%) which we attribute to the large number of additional variables we estimate in this case. We find this equation is significant i.e. F-test that all variables are zeros fails and with this specification almost 37% of the variation is explained which is fairly high.

As discussed above our primary concern with the results is a selection bias such that people are sorting based on some unobservable that is also correlated with adoption of new technology. The most likely sorting variable after income, education and age (which we have already included before) is occupation of the head of the household since that is usually one of the most important deciding factors in a family's decision to settle in a particular location, since some jobs might be more easily available at certain locations. Since the CPS is primarily a labor study we have detailed data on a fairly disaggregate level about the occupation or primary industry of the householder. Accordingly we use these dummy variables by turn, the results are reported in table 3.8 below. Apart from being a sorting variable there are other reasons to include these variables, since we also need to control for the fact that we consider a general purpose technology (PCs) that individuals are required to learn and use in certain professions and not in others. This makes some people more likely to have computers at home compared to others from a similar background (same income and education etc.) since they already know the utility of it and do not need to learn from others. So ideally we would like to exclude such people from our analysis though we need to take them into consideration when calculating $\overline{comp}_{kt}$ since they are the ones driving the quality of the network.

In this table we only report the coefficients of our two main variable the social effects variable and CA_{jk} which measures the selection bias. In the first row we report the results from before (our baseline model with all dummies) for ease of comparison. In the second row we include fairly detailed classified information about the primary industry that the head of the household works for. There are 51 groups and therefore we add a total of 50 dummy variables to equation (3.3). We find that this dramatically improves our estimates of learning effects which is now higher by a factor of two and also significant at 1% level compared to 10% before. We also find that our estimate of the bias is still negative and more significant (at 1% level compared to 5% before). Next in steps to move to a more aggregated level of this measure of occupation, in the next row we report the estimates using the primary occupation of the householder dummy variables (44 of them total), neither the coefficients nor the standard errors of our main variables change much. Then we include the primary occupation group aggregated at two levels in rows (4) and (5), first with a more detailed classification (12 dummy variables) and then a very rudimentary classification into four categories (3 dummy variables). We find surprisingly that even with a very parsimonious classification of occupation groups our estimates stay the same.

As mentioned before data limitations prevent us from including all necessary interaction terms in the regressions, however we can include some of them selectively and only set of interactions at a time. We now report these results. First we add what we found to be the most significant of all the interactions; the one between year of entry into the U.S. and ethnicity, this is intuitive since successive waves of immigrants are not identical as discussed before, the longer time a person spends in this country the more time he/she has to assimilate into the mainstream. We add these interaction terms to our baseline equation first (reported in row 6) from table 3.7, we find that the magnitude of the estimated do not change however it does make the coefficients more significant. This also reiterates a running theme throughout this analysis, that in the presence of a negative selection bias the more controls we add that controls for the unobservables the better (i.e. more significant) our estimates are for the social effects. Here we find that the inclusion of these terms actually makes it almost significant at the 5% level. Similarly it also lowers the estimates for the selection bias variable. We next add these interaction terms to (5) which is our most parsimonious representation with the fewest occupation dummies that makes the coefficient significant. We find that this does not have an impact on the point estimates although it increases the standard errors of the estimates.

In row 8 we consider to the same equation interaction terms controlling for the year of entry into the country and occupation groups, i.e. controlling for the fact that those who came before might have shifted to more profitable professions with better language skills. We find that this makes no difference to our estimates from before (same as row 7 and row 5). Next we add the interaction terms between ethnic dummies and occupation groups, again as before the reasoning being if some ethnicities are predominantly in certain professions this might be driving the results. We find that although the estimates of the standard error increases from before (row 8) the estimates declines by a small magnitude. The problem with limited data is that it makes the standard error estimated highly inaccurate even when the underlying statistical model (data generating process) is correctly specified. In the next section we try to use interval estimates instead of point estimates which are likely to be informative.

3.6 Robustness Checks

3.6.1 Bootstrapping the confidence interval

The problem of small sample sizes can potentially lead to insignificant point estimates. The alternative approach, more in tune with Bayesian methods, would be to consider the *interval* estimates of the parameters instead. As a convenient byproduct this also allows us to examine the robustness of our findings. For example if the 95% confidence interval lies entirely on the positive axis that could be interpreted as strong evidence in favor of social effects, even if the point estimates are not statistically significant. Therefore for level of testing α, we are interested in constructing a $(1-\alpha)100$ confidence interval for the parameters of interest, ie. with probability $(1-\alpha)$ the true parameter lies in this interval. The *standard* interval estimates in this context (reported by most statistical packages), is constructed using the standard normal deviate and the estimated coefficients as reported in table 3.7. By definition this is $\hat{\theta} \pm z^{(\alpha)}\hat{\sigma}$, where $\hat{\theta}$ is the point estimate, $\hat{\sigma}$ is an estimate of the standard deviation of $\hat{\theta}$ and $z^{(\alpha)}$ is the $100\alpha^{th}$ percentile of the standard normal deviate. This is correct under the following assumption, from asymptotic theory;

$$\hat{\theta} \sim N(\theta, \sigma^2) \tag{3.5}$$

with σ^2 constant. The problem with the standard interval is that it is based on asymptotic approximations that might be quite inaccurate in practice, for examples see DiCiccio and Efron (1996). It has been documented by numerous authors that the standard interval can deviate from the exact interval (calculated analytically) widely, more so in small samples.

Therefore instead of using the estimates of standard error from earlier we choose to *bootstrap* the confidence interval instead. Very briefly the bootstrapping is a computational tool that is extremely useful for estimating standard errors of coefficients from small samples, to the extent that the algorithms used are automatic it replaces the analytical effort of obtaining exact intervals (which might not be possible for all but the simplest of problems) with computational effort via simulation. For example if the parameter of interest θ is a $k \times 1$ vector of coefficients and X is a $n \times k$ matrix of data generated by the data generating process, $x \sim F(\theta)$ and we are interested in obtaining the confidence interval for some point estimate $\hat{\theta}(X)$. Then the bootstrap algorithm repeatedly obtains random samples x^* from X of size $m \leq n$ with replacement and estimates the coefficient vector $\hat{\theta}(x^*)$ for each sample. The simplest non–parametric estimate of the cumulative distribution function $\hat{G}(c)$ with B replications and denoting each bootstrap replication as $\hat{\theta}^*(b)$ is as follows:

$$\hat{G}(c) = \#\{\hat{\theta}^*(b) < c\}/B \tag{3.6}$$

The bootstrap intervals reported below use the methodology known as BC_a which stands for *bias–corrected* and *accelerated.* This is a highly accurate though computationally intensive method for estimating confidence intervals from bootstrap distributions. References include Hall (1988), DiCiccio and Romano (1995), and Efron and Tibshirani (1993). For a general survey of the recent developments in this fields as well as the asymptotic properties and derivations of these estimators the reader is referred to DiCiccio and Efron (1996). The following argument motivates the BC_a method. Suppose there exists a monotone increasing transformation $\phi = m(\theta)$ such that $\hat{\phi} = m(\hat{\theta})$ is normally distributed for every choice of θ, but possibly with a bias and a non-constant variance,

$$\hat{\phi} \sim N(\phi - z_0\sigma_\phi, \sigma^2), \qquad \sigma_\phi = 1 + a\phi \tag{3.7}$$

where z_0 is a bias correction parameter and a is the acceleration parameter estimated from the data. Then the BC_a interval of level $(1-\alpha)$ is defined as $(\hat{\theta}_{BC_a}[\alpha/2], \hat{\theta}_{BC_a}[1-\alpha/2])$, where the endpoints are defined as follows,

$$\hat{\theta}_{BC_a}[\alpha/2] = \hat{G}^{-1}\Phi\left(z_0 + \frac{z_0 + z^{(\alpha/2)}}{1 - a(z_0 + z^{(\alpha/2)})}\right) \tag{3.8}$$

where as before $z^{(\alpha)}$ is the 100αth percentile of a standard normal deviate and Φ is the standard normal c.d.f. Note that with $z_0 = 0$ and $a = 0$ the BC_a interval converges to the standard non–parametric c.d.f. given in 3.6 above. Hall (1988) shows that the assumption in 3.7 is second–order accurate i.e.

$$Prob\{\theta < \hat{\theta}_{BC_a}[\alpha]\} = \alpha + O(1/n) \tag{3.9}$$

whereas for standard assumption 3.5 is only first order accurate i.e.

$$Prob\{\theta < \hat{\theta}_{STAN}[\alpha]\} = \alpha + O(1/\sqrt{n}) \tag{3.10}$$

Table 3.9:
Bootstrap Confidence Intervals*

Bootstrap	Coeff.	Samples drawn	S.E.	95% Interval	
Basic Model					
$CA_{jk} * comp_k$	0.009	100	0.005	0.002	0.025
CA_{jk}	-0.004	100	0.002	-0.007	-0.0001
$CA_{jk} * comp_k$	0.009	300	0.005	0.0002	0.02
CA_{jk}	-0.004	300	0.002	-0.007	-0.0001
Add industry dummies [50]					
$CA_{jk} * comp_k$	0.008	100	0.005	-0.0003	0.023
CA_{jk}	-0.004	100	0.002	-0.007	-0.00003
Add interaction terms ethnicity and yr. of entry					
$CA_{jk} * comp_k$	0.009	100	0.006	-0.004	0.019
CA_{jk}	-0.004	100	0.002	-0.009	-0.000
(a) Allow clustering of SE via ethnicity					
$CA_{jk} * comp_k$	0.008	100	0.008	0.00001	0.029
(a) Allow clustering of SE via location (MSA)					
$CA_{jk} * comp_k$	0.008	100	0.011	-0.0004	0.032
(a) Allow clustering of SE via industry					
$CA_{jk} * comp_k$	0.008	100	0.006	0.0014	0.025

Pooled sample all years.
*Using Robust Standard Errors, Bias–corrected confidence intervals

We report the confidence intervals estimated in table 3.9 below. We start of with our basic model reported in table 3.7 before, with all the various dummy variables included for ethnicity, year of survey, year of entry into the U.S. and MSA. The variables we are interested in is our estimate of social effects $CA_{jk} * comp_k$ and the measure of the selection bias CA_{jk} in this equation. We start of by running 100 repetitions i.e. 100 random samples are drawn using the original data each of the same size as the original data.[12] A major concern in any computer simulation is that of convergence of the parameter estimates, i.e. how many repetitions to do before one gets stable estimates of the parameters that cannot be improved upon substantially by increasing the number of samples drawn or repetitions of the algorithm. In this context the accuracy of the estimates is often dictated by the computational power available to the researcher and feasibility.[13] We started with a number of different seeds fo r the random number generator and obtained different samples with n repetitions each. The rule of thumb used in such studies is that if the coefficients change substantially from one run to another then the number of repetitions needs to be increased. For our study we settled on 100 repetitions as the coefficients seemed reasonably stable. We also report the case where the same basic equation is estimated with 300 repetitions. We find that the bias–corrected confidence interval in both cases are positive lending strong support to the hypotheses of positive social effects, and the coefficients are significant at the ten percent level of testing as before. Also the fact that the interval estimates for CA_{jk} are always negative lends strong credence to our claim of negative selection bias. Note that the coefficient estimates do not change substantially when repetitions are tripled, therefore we stick with 100 repetitions for the other models considered next.

As reported earlier our baseline model improves with the addition of dummy variables controlling for the primary industry that the head of the household works for since this might be controlling for the unobservables affecting location decision as well as the decision variable. With this specification we find that the upper

[12] To convince yourself that this indeed leads to different and unique random samples consider the following example, if the data consists of four points (1,2,3,4) then one can obtain numerous random samples using this data each of size four as follows: $(1,2,2,4)$, $(1,3,4,1)$ $(4,2,1,3)\ldots$

[13] Highly optimized code for conducting such simulations are available for publicly available statistical software like STATA, and is used by this study.

bound for the interval of social effects stays the same whereas the lower bound now becomes negative although by a very small order of magnitude. As before the interval estimates of the selection bias stays negative. Next we consider the model with interaction terms added to control for the fact that immigrants across cohorts may not be identical this adds a number of new variables and therefore increases the standard errors, although the intervals for social effects is still mostly positive. Next we consider this model allowing for the clustering of standard error across various levels (variables). The bootstrap then estimates separate standard errors for each value of the cluster variable. We report the three cases where clustering or heteroscedasticity is allowed for at the level of ethnicity (each ethnic group has a different s.d. σ_ϵ), location (MSA) and at the industry level. Except for clustering at the MSA level for the other two cases we find the interval for social effects is entirely positive although the point estimates of the coefficients are not significant. Whereas if clustering is allowed for at the MSA level then the coefficients are not significant and also the interval becomes negative at its lower bound although again the order of magnitude is small. The likely explanation for these facts might be the multicollinearity problem resulting from the addition of so many new variables to the equation.

3.6.2 Specification Checks

Next we consider how sensitive are our results to model specification and sample choice. In table 3.10 below we report the results of our specification tests for functional forms. For ease of comparison we have included in the first row the results from our baseline model with all fixed effects and controls included, this is the same as column 5 in table 3.7. The coefficients reported are for our primary variable of interest measuring social effects $CA * (\overline{comp}_{kt} - \overline{comp}_k)$. As discussed above an endemic problem of studies estimating social effects is whether the model is identified or not, Brock and Durlauf (2001b) shows that one can get around many of the problems using non-linear models. Therefore we start of with a logit specification reported in the second row. The logit coefficients are not directly comparable to the linear probability model considered earlier however we do find that the coefficient is positive and highly significant. In the third row we add the year specific dummy variables to control for the diffusion process and as before we find that the coefficients for social effects are generally lower but still significant at the five percent level of testing. Alternatively we can specify that the latent variable is distributed as a normal variable which leads to the probit model, again we report the case without year dummies in row 5 and with all year dummy variables in row 6. The results are similar we get positive and significant coefficients for our estimate of social effects, adding the year dummies lowers the coefficient estimate and also makes it somewhat less significant[14]. Note that the latent variable models discussed logit and probit are estimated using maximum likelihood methods. This leads to a computational limit since we cannot add the MSA level fixed effects since this makes the number of coefficients to be estimated too many and we had trouble with convergence of the maximization algorithm.

Next we take our baseline model (row 1) and use different measures for the social effects. First we use the logarithm of the mean computer usage (reported in row 7) we find that the coefficients change in magnitude and become insignificant. Similarly when we replace our measure of contact availability with the logarithm we find that the coefficient is still positive but insignificant. One explanation may be that we do not have enough data to calculate so many coefficients therefore next we drop the MSA fixed effects and we find that this improves things substantially. Our

[14] Its still almost significant at the 5% level

Table 3.10:
Functional Form Checks

Change in functional form	Coeff.	P-Val.*
(1) Specification as before (baseline model includes all F.E)	0.009	0.075
(2) Logit without year or MSA F.E	0.236	0.000
(3) Logit with year but no MSA F.E	0.069	0.019
(4) Probit without year or MSA F.E	0.134	0.000
(5) Probit with year but no MSA F.E	0.037	0.030
(6) As in (1) but without MSA F.E	0.009	0.052
(7) As (1) but mean computer is replaced by log mean computer in the interaction term	0.002	0.126
(8) As (1) but CA is measured in logs rather than levels	0.006	0.62
(9) As (8) but no MSA F.E	0.024	0.013
(10) As (1) but CA measured as C_{jk}/A_j	1.769	0.005
(11) As (1) but CA measured as $ln(C_{jk}/A_j)$	-0.036	0.000
(12) As (1) but CA measured as $ln(C_{jk})$	0.021	0.011
(13) As (1) but a quartic polynomial in CA is included as control	0.006	0.164

*1. Reported coeff. of CA*Mean Usage of ethnic group.*

2. Dependent variable is computer ownership.

3. Pooled sample consists of all MSA households, all years.

**All robust standard errors used.*

estimates of the social effect is positive and significant. We use various measures of contact availability next and we find that in general it does not affect our findings. First if CA is measured only as C_{jk}/A_j i.e. we drop the normalizing factor in the denominator we find that this makes our estimates larger and also more significant (as reported in row 10). However when we take the logarithm of the same values we find a negative and significant coefficient in row 11. The simplest way to measure contact availability is just to count how many people of similar ethnicity live in the same area C_{jk}, using the logarithm of this value gives us a positive and significant coefficient. Next we consider the case that computer ownership might depend on the size of the network in a non-linear fashion therefore we include a quartic polynomial in the basic equation and we find a positive but insignificant coefficient for social effects. We conclude by saying most of findings above show that our findings are robust to alternative model specifications.

3.6.3 Impact of Controls

We claimed above that compared to the BLM study this one suffers from a negative selection bias. If this is true then the more controls we add the better the estimates should be both in terms of magnitude and significance provided the controls are correlated with the unobservables. We do find this is the case and our findings are reported in table 3.11 below. We find that ethnicity is often the strongest explanatory variables in this case. Although there is no simple correlation between adding controls the sign and magnitude of the relevant estimates, we do find that eventually when all controls have been added a better estimate is obtained compared to the earlier ones.

3.6.4 Sample Choice

We also investigate whether the choice of sample has an impact on the results. First in table (3.12) the same linear probability model is estimated for separate demographic groups. Certain salient features emerge such as neighborhood effects are not as strong for relatively wealthier households with higher incomes, although the coefficients are not statistically significant for either of the income groups. For the lower income group the sample is much larger by a factor of four which might explain a lower standard error and relatively lower p-value.

For education there is a dramatic difference between the two subgroups considered, those with college education versus those without even a high school one. A priori we would expect neighborhood effects to be particularly strong for the latter and not the former. It turns out to be just the reverse with the estimates being highly significant at 5% level of testing for those with college degrees. On the same note a similar finding is obtained for different age groups, given that we expect that those in the country longer would have had a better chance to integrate with the mainstream neighborhood effects should be lower. However we find just the reverse with it being very strong and significant for the younger immigrants. This result might be due to different age compositions for different ethnicities. This is further investigated in table (3.13), where the same model is estimated for different ethnic groups. We find network effects are not existent or weak for people migrating from countries with English as the first language such as England and Canada. The effect is conversely found to be very strong for the Asian sample. Since given the recent wave of Indian techies immigrating here leaving them out does not significantly change the conclusions either. Also surprisingly the neighborhood effect is estimated to be strong and significant for the groups with higher than median computer usage, which implies a threshold beyond which households have enough information to make the network effect significant.

Table 3.11:
Impact of Controls

Controls	Coefficient	P-Value (Robust S.E)
(1) Only ethnic F.E	0.034	0.000
(2) Ethnic and year F.E	-0.003	0.591
(3) Ethnic, year and MSA F.E	-0.0001	0.979
(4) (3) + male and income F.E and age of householder	0.005	0.265
(5) All controls	0.009	0.075

*1.Reported coefficient of the interaction term CA *mean computer usage of ethnic group.*
2.Dependent variable computer ownership at home.
1.Pooled sample consists of all MSA households, all years.

3.6.5 Selection Bias

As mentioned earlier the major problem associated with studies of social networks is *selection bias* arising due to unobserved variables are not controlled for satisfactorily in the estimation process. We do find strong evidence of such selection as reported in table (3.14). Here contact availability is regressed on the other demographic variables that were found to have been significant earlier. The columns from left to right start with the baseline linear regression and add stepwise year dummies, ethnic dummies and MSA controls respectively. With the addition of the MSA level fixed effects we find that the explanatory power reaches around 80% which suggest a very

Table 3.12:
Sensitivity to Sample Choice

Change in Sample: (by demographics)	Coeff.	P-Val.*	Sample Size
(8a) Family income > $50,000	0.006	0.620	3,013
(8b) Family income ≤ $50,000	0.008	0.127	8,318
(9a) Education, HS degree or less	0.001	0.870	6,602
(9b) Education, some college or more	0.026	0.002	4,729
(10a) Age of householder ≤ 45	0.019	0.013	6,412
(10b) Age of householder > 45	0.007	0.304	4,919

**Robust standard errors used.*

strong sorting going on, i.e. unobserved variables that decide residential choice (and concomitant social networks) are also likely to be correlated with technology choice because of this. Given this strong bias working against us making the coefficients only partially significant, when combined with the small sample problems.

3.7 Conclusion

In recent times there has been considerable concern expressed regarding the so-called digital divide. The policies followed to bridge this divide has been entirely based on prices with subsidies provided for Internet access. We argue in this paper that if learning from others is a strong factor then such policies are not likely to be very effective. A more effective policy can perhaps be designed using the so-called social multiplier phenomenon. A strong step in this direction has been made by the E-rate

Table 3.13:
Sensitivity to Sample Choice

Change in Sample: (by ethnicity)	Coeff.	P-Val.**	Sample Size
(1) Original sample	0.009	0.075	11,331
(2a) Lower than median group computer usage	0.0087	0.14	6,987
(2b) Higher than median group computer usage	0.0484	0.003	4,344
(3a) Hispanic sample*	0.005	0.407	5,514
(3b) Non- Hispanic sample	0.033	0.002	5,817
(4) Exclude India from sample	0.009	0.067	10,774
(5a) Asian sample*	0.087	0.000	2,661
(5b) Non-Asian sample	0.009	0.097	8,670
(6a) European sample + Canada*	-0.022	0.654	2,327
(6b) Non-European sample	0.0096	0.059	9,004
(7) Drop England and Canada*	0.009	0.067	10,494

*1. Reported coefficient of the interaction term CA*mean computer usage of ethnic group*

2. Dependent variable computer ownership at home

**All ethnic samples defined and discussed in text.*

***Robust standard errors used.*

Table 3.14:
Residential Choice Decision

Contact Availability CA_{jk}	1	2	3	4
Male	0.136	0.138	0.16	0.128
	(0.419)	(0.416)	(0.203)	(0.142)
Age	0.3	0.3	0.153	0.054
	(0.000)	(0.000)	(0.000)	(0.000)
Age^2	-0.002	-0.002	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Income				
$\$25,000 - 50,000$	-1.192	-1.195	-0.498	-0.169
	(0.000)	(0.000)	(0.001)	(0.099)
$\$50,000 - 75,000$	-1.331	-1.328	-0.433	-0.126
	(0.000)	(0.000)	(0.001)	(0.371)
$\$75,000$ and above	-1.864	-1.865	-0.424	-0.072
	(0.000)	(0.000)	(0.054)	(0.622)
High School Degree	-0.109	-0.1	-0.455	-0.235
	(0.654)	(0.682)	(0.012)	(0.045)
Some College	-0.496	-0.494	-0.88	-0.686
	(0.048)	(0.048)	(0.000)	(0.000)
College and more	-1.558	-1.548	-1.557	-1.012
	(0.000)	(0.000)	(0.000)	(0.000)
Year dummies (3)	No	Yes*	Yes	Yes
Ethnic dummies (19)	No	No	Yes	Yes
MSA dummies (272)	No	No	No	Yes
R^2	0.033	0.034	0.515	0.789
N	11,331	11,331	11,331	11,331

Dependent variable is the contact availability measure CA_{jk}
(P-values in parentheses)
**Insignificant at 5% level of testing.*

program initiated by the federal government to provide funding to rural and poorer, schools and libraries.

An earlier attempt by Goolsbee and Klenow (2002) suffered from serious methodological problems making their results harder to interpret. The approach taken here more typical of the social network literature is to explicitly define the social networks of the individuals and measuring their impact. We do find strong evidence in favor of network effects in the diffusion of new technology particularly the personal computer considered here. However a few caveats need to be mentioned, first, since the sample considered ethnic minorities this behavior may or may not carry over to the entire population. Second, the sample sizes led to highly noisy estimates which are not as unequivocal as would have been hoped however the broad trend is clear from this study.

Chapter 4

Estimating Demand for the Internet

The telecommunications sector has always been heavily regulated in the United States, as well as in most other developed countries. Regulations have been typically justified through one or more of the following arguments, a) presence of strong *network effects*[1], b) high *fixed/ sunk costs* of entry leading to a natural monopoly in most markets and, c) *universal access* to telecommunications services at reasonable rates has long been enshrined as an official goal of the FCC through legislation by the U.S. Congress (starting with the Telecom Act of 1934). It has also been characterized by a rapid pace of technological change. Opponents therefore have long argued that such regulations end up being counterproductive. Even correct policies can lead to severe losses in consumer welfare due to bureaucratic delays in the adoption of a superior technology, for example see Hausman (1999) (voice messaging services).

Whether universal access is a valid policy goal has long been a subject of debate (Crandall and Alleman 2002) among economists. Some have also noted that subsi-

[1] Utility derived by a consumer depends strongly on the number of other users of the technology, for example telephones.

dies may not be the best way to achieve this goal given the low price elasticity of access estimated for telephones, as well as the associated deadweight losses (Taylor 1994). The price elasticity of access is defined as the percentage change in usage (penetration) rates at the market level for a one percent change in average prices. This goal of universal access has more recently been extended to the Internet (the digital divide debate). For instance the telecommunications Act of 1996 set up a system of subsidies for schools, hospitals and libraries to be financed through taxes on toll calls (called the E-rate program). Hausman (1998) finds that this program caused a deadweight loss of around 2.25 billion dollars (in 1997) which is roughly equal to the amount disbursed by the program. The high deadweight loss results from a high price elasticity of demand for toll calls. The general public also enjoys implicit subsidies from the *Internet Tax Freedom Act* of 1998 which initially placed a three year moratorium on all taxes on Internet access and has since then been extended for an additional three years.

Much of this debate has been qualitative in nature since empirical studies have been rare in this context, Kridel, Rappoport, and Taylor (1999b) being the notable exception. Their study applied the standard method used in earlier studies of telephone demand. Kridel, Rappoport, and Taylor (2002) has applied a similar technique in estimating the demand for cable modems, a proxy for broadband technologies. The recent debate regarding the potential for broadband in improving consumer welfare and the role of government regulations is discussed in Crandall and Alleman (2002).[2]

This study hopes to make the following contributions, first, we point out certain serious flaws with the earlier studies and correct their errors. Second, we consider a structural model of consumer choice which allows us to explicitly take into account the differentiated nature of Internet services (as opposed to telephones), and incor-

[2] Much controversy has surrounded the asymmetric regulation of cable and DSL for instance.

porate the wide variation in prices observed in the data. We also account for the censored nature of the price data available. Third, using household level micro data we obtain new estimates for the price elasticity of access as well as consumer surplus from this new technology. We find the price elasticity to be substantially higher compared to earlier studies. This finding is significant since it provides some justification for the policies undertaken in this context. Fourth, we use data on utilization of the Internet to control for unobserved household tastes.

**The rest of the paper is laid out as follows, in section 2 we discuss related studies, as well as point out some of their flaws. Following which in section 3 we present some descriptive analysis of the data used for this study. In section 4 an econometric model of demand for the Internet is set up. Section 5 presents our main results and certain extensions such as measures of consumer surplus and price elasticities are discussed in section 6 and finally section 7 concludes.

4.1 Related Work

There has been several studies in this context in recent times, perhaps fueled by the controversy surrounding the regulation of 'broadband' by the FCC. Goolsbee and Klenow (2002) estimates the demand for broadband. Craverman provides an excellent survey of this debate surrounding the regulation of broadband. The econometric model considered here is a Type II tobit model which has been used recently by Scott and Garen (1994) to study the incidence of the lottery tax as well as Min and Kim (2003) to study credit card borrowing decisions. However this study is closer in spirit to the original application by Gronau (1973) which estimated the value of time for housewives and its impact on participation in the labor market.

The study by (Kridel, Rappoport, and Taylor 1999b)(henceforth KRT) used a binary logit approach to estimate the price elasticity of demand for the Internet. A second study by the same authors ((Kridel, Rappoport, and Taylor 2002)) applied the same technique to study the demand for cable modems, a proxy for broadband connections. Since the technique and the sources of data are identical much of the following critique applies to both studies. The data was obtained from a national survey of households by a private firm, for the two years 1997 and 1998. The sample sizes were quite large similar to ones considered here.[3] The methodology and the variables used for their study, replicates earlier studies on household demand for telephones (for example (Perl 1978)). Taylor (1994) provides an excellent survey of this literature. The binary variable of access is regressed on price and demographic variables. The price variables is constructed as either the price actually paid for service, for current subscribers, or the average price that the household can expect to pay for service, which is defined as the average price paid by all subscribers in that location (MSA). Our study is closer in spirit to (Taylor and Kridel 1990), which estimated a structural demand model.

We note that these studies suffer from a couple of serious methodological flaws as follows, KRT implicitly assumes, a) homogeneous good and, b) exogeneity of prices. Both of which might be more appropriate for studying the demand for telephones, compared to the Internet. Telephone services are relatively more homogeneous, and is usually provided in most locations by a local (natural) monopoly, i.e. consumers usually have very little choice in terms of services.[4] Also a credible case can be made for the exogeneity of prices, since prices are heavily regulated by the FCC based on local cost and technological factors.[5] On the other hand each household

[3] The only additional information available for their study was information regarding the ownership of other new technology goods by the household. They utilized this information to control for household level unobserved factors such as learning costs etc.

[4] Some choices involve opting for a fixed versus measured (fee depends on number of calls made) service, and also options like voice messaging and caller-id etc.

[5] Another critical difference is that for the telephone studies prices were obtained directly from

had considerable choice in terms of Internet services over the sampling period, with numerous ISPs offering services differentiated along numerous dimensions such as the amount of online storage space provided for hosting personal webpages, or the number of e-mail accounts provided etc., apart from the usual unlimited or limited services.[6] Therefore if early adopters and/or heavier users opt for a more expensive service, an upward sloping demand curve is estimated unless this *quality of service* aspect is explicitly controlled for. Indeed in our analysis we found that in a similar setup (emulating KRT), the price coefficient to be either not significant or to be positive (and sometimes significant).[7] In reality *quality* is difficult to control for since details regarding the features/ characteristics of the Internet service that each household subscribes to is not available in the data considered by KRT or here.

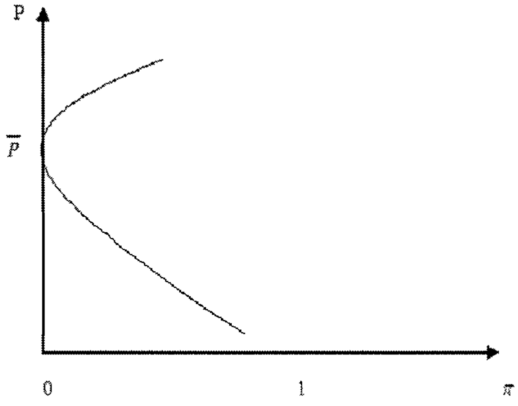

Figure 4.1: Logit Model Demand Curve

Secondly, even if we assume that prices are purely exogenous their approach is still fatally flawed since it ignores the censored nature of the price data. Assume that each individual obtains a random draw from the independent and identical distribution

service providers, which implies that measurement error in observed prices is likely to be significantly less.

[6] Unlimited hours versus a fixed number of hours.

[7] In separate estimation not reported here.

of prices, and those with lower draws adopting the Internet. This implies that the unobserved price draws (for non-subscribers) were higher on average compared to the observed ones. Therefore taking the average across those with subscriptions seriously underestimates the true mean of the distribution. Given that the mean of any censored distribution is never equal to the mean of the original (for all non-degenerate probability distributions). Gronau (1973) made a similar argument in the case of wage draws for working and non-working women. Figure (4.1) shows an example of the curve fitted by their procedure, controlling for all other demographic variables, the average price leads to no access and other prices both higher and lower are associated with a positive probability of access. Therefore a linear demand specification can lead to a spurious positive and significant estimates of the price coefficient (π is the probability of having an Internet service at home). We found this always to be the case in our effort to duplicate their results using the data at hand.

In this setup a negative and significant estimate of the price coefficient (as obtained by (Kridel, Rappoport, and Taylor 1999b)) can only arise if the distribution of prices are positively skewed. Since they use arithmetic mean to calculate expected prices for non-consumers, and it is known that in the presence of a few outliers in the data, the arithmetic mean can be substantially different from the median of the distribution. For instance this can happen if a few subscribers pay very high monthly fees compared to the majority of subscribers. This will lead to a (artificially) high computed (expected) price for the non-subscribers and (on average) much lower observed prices for the subscribers, i.e. a negative correlation between prices and adoption probabilities.[8]

[8] In our data we found no correlation between a price variable constructed as in KRT and the dummy for Internet access.

In the next section we specify an econometric model that specifically addresses some of these issues.

4.2 Data

The data used for this study was discussed earlier in chapter 1 in details and will not be repeated here. We briefly discuss the variables used for this study here.

Basic variables: Internet is a dummy variable for access to the Internet at home through either PC or Web TV, price is the monthly subscription fee paid for service as well as any long distance or toll fee paid for each call (average for each month). This is truncated at \$90 per month.[9]. Apart from age most others are dummy variables constructed as follows. Male is a dummy for the head of the household being male. Five dummies for family income is constructed with the baseline (excluded) being the lowest (upto $\$20,000$) level. Education is divided into four categories with the excluded being the first (No HS/GED). Note that the next category (HS/GED/Some College) is fairly broad however since education is highly correlated with income we found that a finer categorization leads to problems of multicollinearity.[10] College refers to only four year college degree and all other professional degrees are classified under advanced degrees. Black and Hispanic are two race dummies used for the analysis. Note that by census definition Hispanic is an ethnicity and black is a race as are Asian-Americans etc.[11] Therefore Hispanic and black are not mutually exclusive. It is also customary to control for the household size and the total number of children (under 15 years of age) in the household.

[9] We do not consider this as a serious problem since very few data points even come close to this level.

[10] It includes vocational training, associate two year degrees as well as HS graduates and GED certificates.

[11] Other categories such as Asian were not found to be significantly different from the general population. This may also be due to small sample sizes.

Geographic variables: Rural is a dummy for households not living in designated metropolitan areas (MSAs decided by the census). Since MSAs can be geographically very large areas, central city is a dummy for living in the central city of the MSA and not in the suburbs. Since the size of the MSA can be a crucial factor, large is a dummy variable denoting households in MSAs with population greater than a million. The median income is obtained from SAIPE, with non-MSAs being assigned the state median income.

Technology variables: Technological background or sophistication is an unobserved variable that is likely to affect the search and learning costs and therefore reservation prices strongly. We therefore construct several variables to control for this, computer is a dummy for the household owning more than one computer. Also given the rapid pace of technological change more savvy households are expected to own late model computers. Thus a series of year dummies are constructed from the survey question, 'when was the latest computer in the household bought?' (yr00 is 2000 and so on). The excluded variable being no computer or vintage earlier than 1997 for the 2000 sample (*Earlier* dummy).

Fixed effects: Outside of MSAs the location of other households are not reported by the CPS, therefore to control for local fixed effects we construct a MSA only sample. Other authors (Goolsbee and Klenow 2002) have noted that local city/MSA specific unobserved factors are likely to play a large role in deciding the level of Internet penetration. We seek to control for this through a series of MSA level variables as reported in table (1.2). We consider the median income of each MSA (more accurate than state level considered earlier), population as well as the proportion of households with a computer, with more than one computer and, owning the latest year computer respectively.

The utilization variables are self-explanatory.

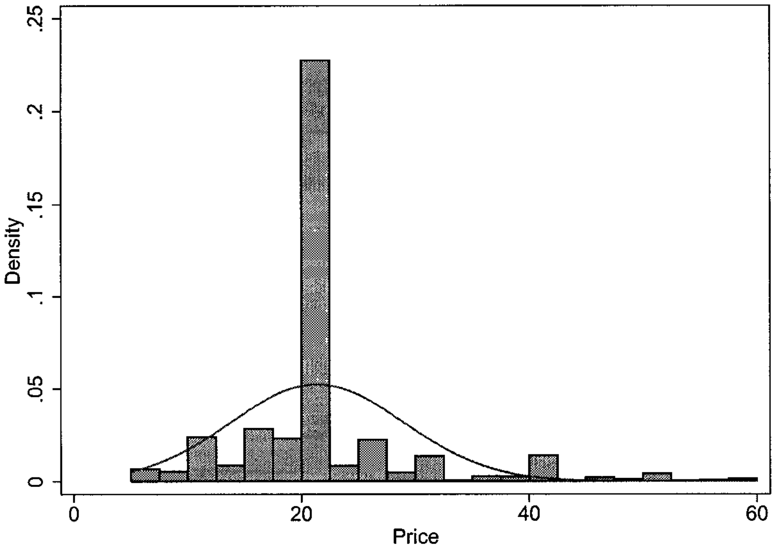

Figure 4.2: Price Distribution (all) 2000

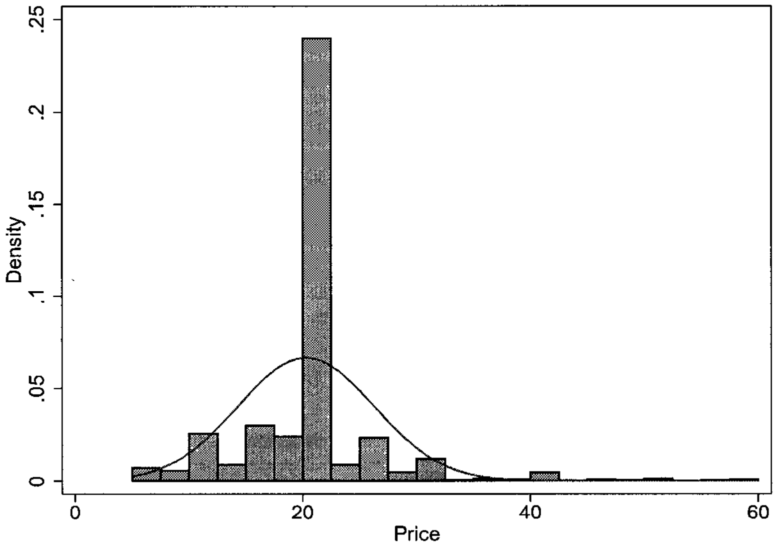

Figure 4.3: Price Distribution (dialup only) 2000

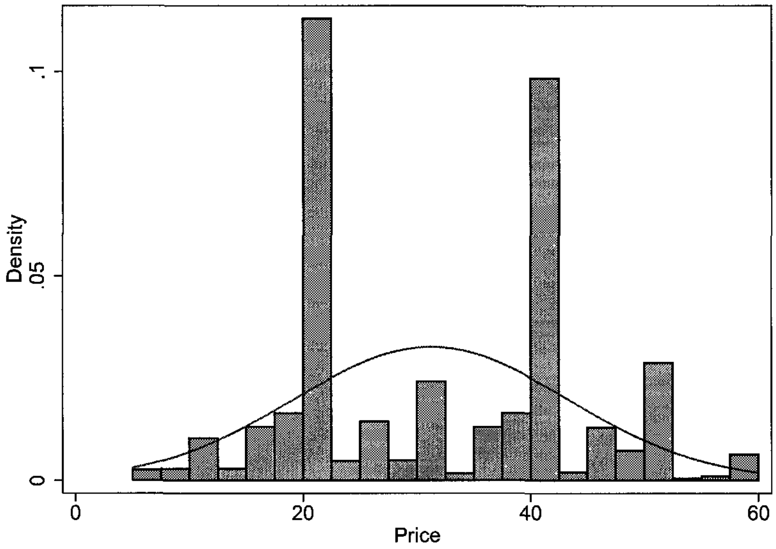

Figure 4.4: Price Distribution (broadband only) 2000

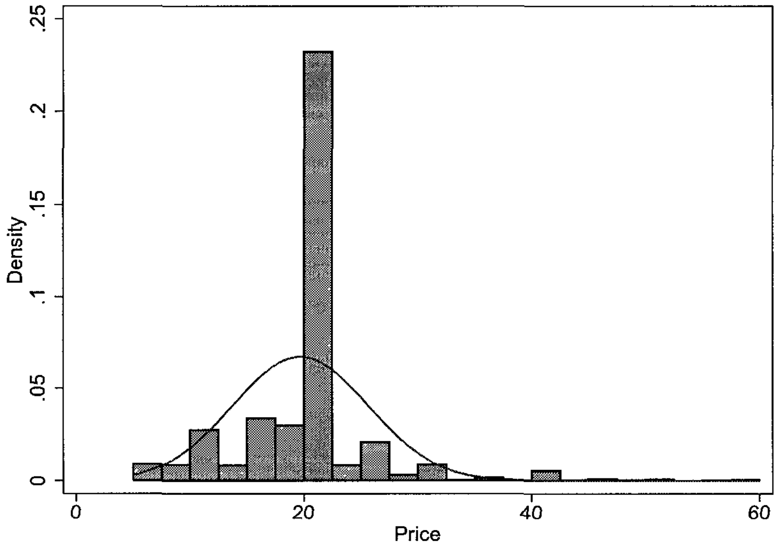

Figure 4.5: Price Distribution (all) 1998

4.2.1 Prices

Our basic premise in this paper is that a simple one size fits all, uniform random pricing model (simple probit) does not work in this context due to the actual variation in prices observed in the data. Therefore it is instructive to show some figures of actual distribution of prices obtained from the data. Figures (4.2) and (4.5), show the overall distribution of prices for the years 2000 and 1998 respectively. A normal curve is overlayed for reference in the figures. For the 2000 sample, the availability of broadband in some locations and their higher prices (compared to dialup) may have caused a high standard deviation in the estimated price distribution. Therefore we break the prices down by method of access, only dialup prices are shown in figure (4.3) and broadband is shown in figure (4.4) respectively.

Note that the vast majority of dialup subscribers pay around $20 - 22.5 per month as subscription fees for Internet access however there is also substantial variation around that mean for both the years 2000 and 1998, even after controlling for broadband. We also find that the distribution of broadband prices are bimodal, this might be due to two reasons, first there may be promotional pricing available to new adopters (such as lower prices for the first six months). We find only 50 MSAs with a significant number of subscribers, which implies that in 2000, broadband was still being launched across much of the country. Second, it could be also due to two alternative types or qualities of services were available in most locations. The twin peaks did not correspond to the different methods of access (cable vs. DSL), either or both were found to be bimodal in most locations in the sample.

4.3 Econometric Models

In this section we consider a simple structural model of adoption of the Internet at home. We seek to correct the fatal flaws of earlier studies noted before, by controlling for both the *endogeneity* of prices and *censoring.* We show that under certain simplifying assumptions this model leads to a standard censored regression setup, Amemiya (1984) refers to this as a *Type II* Tobit model. This allows us to utilize standard estimators with well studied properties of consistency and asymptotic normality (Amemiya (1984) provides an excellent survey of this literature). Gronau (1973) originally used a similar setup to study women's participation in the labor market, since then it has been applied in a number of other contexts as well (see Amemiya (1984)). Examples include studies of automobile demand, other durable goods expenditure etc. (Scott and Garen 1994) studied the incidence of the lottery tax using a standard type II Tobit model setup.

The market for Internet access at home is specified as follows: on the demand side, households follow a simple decision rule whereby they adopt the Internet when actual price falls below their reservation price. Each period households obtain the actual price by a random draw from the current price distribution (to be specified). Also there is a one-time fixed cost of adoption for associated equipment such as a PC or Web TV etc., which is necessary for access. For simplicity we assume that this fixed cost F, which is either the price amortized over the life of the equipment or the rental rate. This is assumed to be the same for all households in a given period.[12] Households are assumed to differ along both observed and unobserved dimensions, i.e. certain socio-demographic variables such as income, education etc. are observed

[12] Think of this as the minimum cost of such an equipment. However if there is geographic dispersion in prices a complete treatment would also include a draw from the fixed cost distribution (for each region/MSA), each period, however given the data at hand this is well beyond the scope of this paper. A search for additional data on PC prices across cities for the years in question, was not successful.

for each household. Whereas other relevant factors such as tastes, technological inclination and/or sophistication, existing knowledge of computers and access to technical help and support (through informal networks), are unknown. Earlier studies had used ownership of other relatively new technology products and/or questions regarding attitudes of households towards such gadgets as control for some of these. Since such ownership or attitudes data is not available to us we use the ex post utilization data to control for some of these same factors. Hence the utility derived from the Internet and reservation prices follow a probability distribution across the population.

Note that this setup yields a standard S-shaped adoption curve which is almost always observed for new goods. Most studies of diffusion of new technologies have used variations of this framework (see Geroski (2000) for a survey). For example, if the reservation price P^r is distributed as a normal random variable and price follows the time path $P_t = P_0 - \alpha * t$, then the decision rule, adopt if $P_t \leq P^r$, leads to a standard S-shaped market penetration curve.

We assume that the Internet can be used for K activities $(a_1, \ldots, a_K)$, such as e-mail, searching for information etc. Let $a_k = 1$ if the household uses this particular service and zero otherwise. Each household demands a portfolio of services $(a_{i1}, \ldots, a_{iK})$. Given household characteristics X_i, let,

$$a_{ik} = X'_{ik}\delta_k + \zeta_{ik} \quad \forall k = 1, \ldots, K \tag{4.1}$$

where ζ_{ik} is a mean zero unobserved taste parameter, and X_{ik} are selected columns of X_i. Note that $\{\zeta_{i1}, \ldots, \zeta_{iK}\}$ is drawn from a multivariate distribution and they are assumed not to be independent of each other, i.e. unobserved tastes across activities maybe correlated. For simplicity assume that utility derived from each service is identical for all households,[13]. Then we define the reservation price as,

$$P^r = \alpha_0 + \alpha_1 E(a_1|X) + \ldots + \alpha_K E(a_K|X) \tag{4.2}$$

where the i subscript for households has been dropped, and α_k is the monetary value attached by the household for service k. $E(a_k|X)$ is the probability of using the kth service given household characteristics X. Without loss of generality this can be rewritten as:

$$P_i^r = X_{1i}'\beta_1 + u_{1i} \tag{4.3}$$

where $X_{1i} \subset X_i$. What we have in mind is a situation where someone who wants to use the Internet for only e-mail is willing to pay a different price, possibly lower, compared to someone who also wants to read the news online.

We observe considerable variation in prices paid for monthly services by households in our sample. Thus we assume that Internet services are differentiated, along unobserved dimensions. Since we do not have details of the service chosen by the household certain assumptions need to be made. Specifically we assume that Internet services differ in quality along a continuum and each consumer given her optimal portfolio of services $(a_{i1}, \ldots, a_{iK})$ chooses a particular quality of service, which in turn implies a particular price distribution from which she draws the actual price. A more general model would start by noting that Internet service in most markets are offered competitively by a multitude of firms (given low entry barriers), which implies that there are no profits, price equals average costs. If there is a continuum of services available, for instance if the number of firms goes to infinity and each offers a particular portfolio of services. We can specify the supply side equivalent of

[13] Note that a general model where valuation of services as well as probability of using the service varies across households, and depends on the same set of variables such as income, education etc. is not identified.

equation (4.2), as follows:

$$P^o = \gamma_0 + \gamma_1 E(a_1|X) + \ldots + \gamma_K E(a_K|X) \tag{4.4}$$

where P_o refers to the offer price of this contract. This is the supply price of offering the particular portfolio of services demanded by household i, this can be similarly simplified as,

$$P_i^o = X_{2i}'\beta_2 + u_{2i} \tag{4.5}$$

as before $X_{2i} \subset X_i$, also it is possible that $X_{1i} = X_{2i}$, i.e. all variables that affect the reservation price for the household also affects the offer price of the contract it wants to buy. Also we assume that $\{u_{1i}, u_{2i}\}$ are distributed bivariate normal with mean zero and variance:

$$V(u_{1i}, u_{2i}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ & \sigma_2^2 \end{bmatrix}$$

Then the selection equation can be written as:

$$P_i = \begin{cases} P_i^o & if \quad P_i^o \leq P_i^r \\ 0 & if \quad P_i^o > P_i^r \end{cases} \tag{4.6}$$

Also let us define the dummy variable for Internet access $y_i = 1$ when household i has adopted the Internet and zero otherwise. The data consists of dependent variables (y_i, P_i) for all households and exogenous variables X_i for all households. Similarly the portfolio of services consumed $(a_{i1}, \ldots, a_{iK})$ is only observed when $P_i^o \leq P_i^r$. Note that the reservation price is never explicitly observed by the researcher but only the offer price when the household has Internet service.

The model we estimate consists of equations (4.3), (4.5) and (4.6), for the reservation price, the offered price and the selection equation respectively. This formulation leads to the standard censored regression model referred to by Amemiya (1984) as *Type II* Tobit model.[14] Note that a standard Tobit model could also be used in this setup, with a single equation for participation and the service chosen conditional on purchase. However a well known shortcoming of the standard Tobit model ((Cragg 1971)) is that if the probability of adoption is less than half then it implies that cheaper services are more likely to be observed compared to more expensive services.[15] i.e. the conditional pdf declines with a rising price. However for Internet access (as for consumer durables), conditional on adoption the most likely price is a positive integer and not zero (around $20 here), i.e the mode of the conditional price distribution is a positive integer, which makes the Type I Tobit an inappropriate choice in this context. The extension of the standard tobit model considered here separates the participation and purchase decisions, and is flexible enough to accommodate this phenomenon. The likelihood function for this model can be written as follows,

$$\mathcal{L} = \prod_{y_i=0} Prob(P_i^o > P_i^r) \prod_{y_i=1} Prob(P_i^o \leq P_i^r) f(P_i^o | P_i^o \leq P_i^r) \tag{4.7}$$

where $f(.|.)$ is the conditional density of the offer price.

Maddala (1983) discusses identification conditions and estimation procedures for this model. This model can be estimated by either maximizing the likelihood in (4.7), alternatively the Heckman two-step estimator may also be used for consistent estimates. Since implementing the Heckman estimator is significantly simpler it has been widely used in empirical studies, for instance see Scott and Garen (1994)(lottery

[14] Note that there is a subtle difference with the standard type II models where there are two equations for selection and consumption respectively. Whereas in this model (also in (Gronau 1973)) the two structural equations are combined to obtain the selection equation so certain variables necessarily enter both equations. Which implies that the impact of these variables are not identified without certain assumptions.

[15] This is due to the assumption of normality and if it is left censored above the mean.

tax).We briefly outline the estimation procedure here, start by defining,

$$Z_i'\delta = \frac{X_{1i}'\beta_1 - X_{2i}'\beta_2}{\sigma} \quad (4.8)$$

$$u = \frac{u_{2i} - u_{1i}}{\sigma} \quad (4.9)$$

where

$$\sigma^2 = Var(u_{2i} - u_{1i}) = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} \quad (4.10)$$

and $Z_i = X_{1i} \bigcup X_{2i}$. This implies that the participation equation (4.6) can be written as, (i.e. price is observed if the offer price is less than the reservation price),

$$\begin{aligned} Prob(P_i^o \leq P_i^r) &= Prob(X_{2i}'\beta_2 + u_{2i} \leq X_{1i}\beta_1 + u_{1i}) && (4.11) \\ &= Prob(Z_i'\delta \geq u) && (4.12) \end{aligned}$$

Then we can rewrite the likelihood as,

$$\mathcal{L} = \mathcal{L}* \prod_{y_i=1} f(P_i^o | P_i^o \leq P_i^r) \quad (4.13)$$

where we define:

$$\begin{aligned} \mathcal{L}* &= \prod_{y_i=0} Prob(P_i^o > P_i^r) \prod_{y_i=1} Prob(P_i^o \leq P_i^r) \\ &= \prod_{y_i=0} Prob(Z_i'\delta < u) \prod_{y_i=1} Prob(Z_i'\delta \geq u) && (4.14) \\ &= \prod_{y_i=0} \Phi(Z_i'\delta) \prod_{y_i=1} (1 - \Phi(Z_i'\delta)) \end{aligned}$$

For those with the Internet i.e. $y_i = 1$, rewrite equation (4.5) as

$$P_i^o = X_{2i}'\beta_2 + \sigma_{2u}\lambda(Z_i'\delta) + \epsilon_i \quad (4.15)$$

where

$$\lambda(Z_i'\delta) = \frac{\phi(Z_i'\delta)}{[1-\Phi(Z_i'\delta)]} \tag{4.16}$$

$$\epsilon_i = u_{2i} - \sigma_{2u}\lambda(Z_i'\delta) \tag{4.17}$$

Note that $\phi(.)$ and $\Phi(.)$ are the density and cumulative density functions of the standard normal variable. Also note that ϵ_i can be shown to be distributed with zero conditional mean. The model to be estimated consists of the participation equation (4.23) and conditional on performance the observed price paid for service (4.5) respectively, which is done by a two step procedure as follows:

Step 1 Estimates $\hat{\delta}$ are obtained from a probit regression of y_i on the set of independent variables Z_i. This step provides consistent estimates of β_{1j}/σ for variables in X_{1i} but not in X_{2i}, and also estimates β_{2j}/σ for variables in X_{2i} but not in X_{1i}. Also we get estimates $(\beta_{1j} - \beta_{2j})/\sigma$ for variables in both X_{1i} and X_{2i} respectively.

Step 2 Using these estimates of $\hat{\delta}$ estimate equation (4.15) for only those with Internet service, i.e. $y_i = 1$ using simple least squares method. Heckman had shown that this provides consistent estimates of β_2 and $\sigma_{2u} = (\sigma_{12} - \sigma_2^2)/\sigma$.

Maddala (1983) notes that for identification of this model either of these conditions are required to hold *a priori*:

(i) u_{1i} is distributed independently of u_{2i}, i.e. $\sigma_{12} = 0$, or

(ii) There is at least one variable in X_{2i} which is not present in X_{1i}.

Note that under either conditions all parameters can be consistently estimated.

In order to estimate price elasticity we need an estimate of σ. The estimation procedure varies somewhat depending on the identifying assumptions made.

*Case I:*If there is any element in X_2 not present in X_1 (say the j^{th}), the estimation of σ is straightforward. Given estimates of β_{2j}/σ obtained from the first step and β_2 from the second step, we can estimate σ.

Case II: Assume $\sigma_{12} = 0$, but all variables appear in both equations. Given estimates of σ_{2u} from the second step and since it is known that $\sigma_{2u} = (\sigma_{12} - \sigma_2^2)/\sigma$, since $\sigma_{12} = 0$ we only need an estimate of σ_2^2 to be able to estimate $\hat{\sigma}$. σ_2^2 is obtained by first calculating the residuals from the second step regression,

$$\hat{u}_{2i} = P_i - X'_{2i}\hat{\beta}_2 \quad if \quad y_i = 1 \tag{4.18}$$

and using them these to obtain,

$$\hat{\sigma}_2^2 = \frac{1}{N_1}\sum_{i=1}^{N_1}[\hat{u}_{2i}^2 + \hat{\sigma}_{2u}^2(Z_i'\hat{\delta})\lambda(Z_i'\hat{\delta})] \tag{4.19}$$

The model specification is completed by noting that the offer price is always constrained to be positive by economic theory, although this has not always been imposed in empirical studies (Gronau 1973). There are two relatively simple ways to incorporate the non-negativity of prices (see Cragg (1971)), a) first the offer price can be truncated at zero[16] and the density function scaled up. b) A less arbitrary way is to consider the log of the offer price, i.e. instead of assuming that P^o is normal, we assume that,

$$\log(P^o) \sim N(X'_{2i}\beta_2, \sigma_2^2) \tag{4.20}$$

Note that reservation price can be negative in this context since it includes the

[16] It is truncated above by the reservation price P^r* which is unobserved

fixed cost of adoption such as equipment costs, learning and search costs etc. For comparability we would like to consider the logarithm of reservation prices as well. Therefore to avoid taking the logarithm of zero or negative numbers are not defined we need to define the new variable $\underline{\mathrm{P}}^r = \max[0+\tau, P^r]$, where τ is a sufficiently small number. Without loss in generality we can define,

$$\log\left(\underline{\mathrm{P}}^r\right) \sim N(X_{1i}'\beta_1, \sigma_1^2) \tag{4.21}$$

Since the reservation prices are unobserved censoring them at zero does not affect the estimates however it makes the estimation easier. Since it implies that the participation equation retains its character (as in 4.23 above),

$$\begin{aligned} \Pr(\log P_i^o \leq \log \underline{\mathrm{P}}_i^r) &= Prob(X_{2i}'\beta_2 + u_{2i} \leq X_{1i}\beta_1 + u_{1i}) && (4.22) \\ \Pr\left(y_i = 1\right) &= Prob(Z_i'\delta \geq u) && (4.23) \end{aligned}$$

Lastly note that since utilization is observed only for current users the data is essentially censored. Presumably the utilization decision is taken simultaneously with the adoption decision. However a model allowing for censoring of multiple variables leads to an intractable likelihood. We therefore need to assume that if the utilization data contains additional information on unobserved household characteristics (not already controlled for), then this impacts both the offer price and reservation prices equally, i.e. $\beta_{j1} = \beta_{j2}$ for all utilization variables. If utilization depends on demographic variables already controlled for then they will have no additional explanatory power and this assumption affects nothing. We believe this admittedly extreme assumption is still better than ignoring this additional data (see results below).

4.4 Results

4.4.1 Aggregate Model

Hausman (1998),(1999) has pioneered a method of estimating price elasticity and consumer surplus using only market level data. He applied this method for instance to estimate the impact of regulatory delay in introduction of voice messaging services (1998) and also cell phones (1999). A standard log-log model of demand is specified at the market level, i.e the logarithm of subscriptions is regressed on the logarithm of average prices along with per capita income, population etc. An estimate of price elasticity is directly obtained (coefficient of log price), which is used to obtain an aggregate measure of consumer surplus, by integrating the area under the aggregate demand curve from current prices to the maximum price at which demand becomes zero. The intuition being the non-availability of the new good in previous periods is economically equivalent to it being available at a *virtual* price which sets demand equal to zero.

There are several shortcomings of this approach, first, the estimates are first approximations of the actual figures since for the log-log model the virtual price is infinite. Second, aggregation leads to high measurement errors, and third, it is difficult to control for the endogeneity of prices. However the strong argument in its favor is that it can almost always be obtained given the limited requirement of data. Given these concerns we concentrate on discrete choice household models for the remainder of this study. We briefly mention some of the estimates obtained for Internet access here in table (4.4.1).

For 1998 we find that both the OLS and instrumental variables (IV) are close and the former is significant at 5% level of testing. The instruments used are prices and subscriptions for 2000. The instruments are valid under the assumption that there

Table 4.1:
Aggregate Demand Estimates for MSAs

	1998		2000
	OLS*	IV	OLS
Log of monthly price	-0.697	-0.945	-0.099
	(0.307)	(1.059)	(0.148)
Log of income	1.008	0.984	0.644
	(0.353)	(0.367)	(0.181)
Log of population	0.781	0.786	0.96
	(0.073))	(0.076)	(0.038)
Intercept	-6.941	-6.046	-6.787
	(3.522)	(5.087)	(1.728)
Number of Observations	105	105	105
R^2	0.623		0.892

**Dependent variable is log of subscriptions (dialup only) 1998.*
Note: Standard errors (robust) in parentheses.

are location specific cost factors that affect prices, then present and future prices will be correlated but there is no reason to expect current subscriptions to depend on future prices. The Hausman test rejects the null hypothesis of no systemic difference between the coefficients. However the IV estimates are not significant with p-value of 0.37. Note that these estimates particularly the IV estimates are very close to ones from the discrete analysis reported below. All regressions reported are highly significant. Note that the relatively few number of observations for MSAs is due to the fact that we use a separate source for the per capita incomes, the CPS data being top coded cannot be used to calculate this crucial variable.[17] Only those MSAs were selected where the geographic areas matched exactly.[18]

[17] The sample selected affects the estimates considerably which might be either due to measurement error introduced from using the wrong per capita income (when more MSAs are added). We also cannot rule out a selection bias, for instance if the census definitions changed for areas with high growth which are also likely to have a high subscription say.

[18] The income estimates are obtained from the *Small Area Income and Poverty Estimates (SAIPE).*

The estimates for 2000 however are never significant and deviate widely from the estimates of price elasticity obtained below. The IV estimates are not reported but are also never significant and sometimes positive. There are several potential explanations for this, a) there are no cost difference across MSAs, variations in average prices are entirely caused by measurement errors, b) endogeneity of prices, as noted before quality of service affects prices, therfore if locations with higher levels of adoption also has people with more expensive services can lead to a positive and significant estimate of elasticity. c) Simultaneity bias, prices are subscriptions are determined jointly by unobserved location specific fixed effects, such as industrial composition etc. If people are sorted across locations based on some unobserved ability such as technological sophistication then also a spurious positive relationship might arise.

4.4.2 Discrete Model

One advantage of the model outlined above is that the decision to adopt the Internet can be broken down into two parts, first there is the participation constraint (eq. 4.23) and, second, there is the contract or monthly service that is purchased. Both these decisions can potentially depend on different sets of variables. However *a priori* economic theory does not suggest that any of the variables considered here will affect only participation and not the contract purchased or vice versa.

The first set of estimates reported in table (4.5.1) assume that $\sigma_{12} = 0$, i.e. that the error term for the offer price and the reservation price are not correlated. This can be justified by noting that the whereas the error term in the offer price (u_{2i}) is interpreted as cost shifters that vary in unobserved ways across various locations, the error term included for reservation prices (u_{1i}) is due to unobserved variations in tastes or household characteristics. We briefly discuss the estimated coefficients here.

Table 4.2:
Discrete Choice Demand Estimates I

	2000		1998	
	Probit	OLS	Probit	OLS
Age	-0.011	0.0001*	-0.014	-0.001*
	(0.001)	(0.001)	(0.001)	(0.001)
Male	0.005*	0.015	0.077	-0.004*
	(0.021)	(0.007)	(0.019)	(0.009)
$20,000–35,000	0.260	0.005*	0.224	-0.008*
	(0.030)	(0.016)	(0.029)	(0.016)
$35,000–50,000	0.413	-0.025*	0.495	-0.001*
	(0.033)	(0.019)	(0.031)	(0.018)
$50,000–75,000	0.613	-0.041**	0.691	-0.024*
	(0.034)	(0.024)	(0.032)	(0.022)
$75,000+	0.827	-0.023*	0.846	-0.008*
	(0.038)	(0.031)	(0.035)	(0.026)
HS/GED/Some College	0.203	-0.011*	0.261	0.018*
	(0.029)	(0.014)	(0.027)	(0.016)
College Degree	0.019	-0.012*	0.127	0.015*
	(0.033)	(0.012)	(0.028)	(0.012)
Advanced Degree	0.415	-0.065	0.535	-0.049
	(0.036)	(0.189)	(0.032)	(0.021)
Black	-0.379	0.045	-0.479	0.029*
	(0.034)	(0.020)	(0.035)	(0.020)
Hispanic	-0.421	0.052	-0.404	0.059
	(0.038)	(0.021)	(0.037)	(0.019)
Married	0.101	-0.06	0.033*	-0.042
	(0.028)	(0.011)	(0.026)	(0.012)
Single	-0.109	-0.031	-0.049*	-0.030
	(0.033)	(0.014)	(0.031)	(0.015)
Employed	-0.013*	-0.019	-0.01*	-0.002*
	(0.025)	(0.009)	(0.024)	(0.011)
Household size	0.056	0.02	0.096	0.025
	(0.012)	(0.005)	(0.011)	(0.005)
No. of children	-0.119	-0.021	-0.137	-0.022**
	(0.017)	(0.007)	(0.015)	(0.007)

Robust standard errors in parentheses.

**Not significant at 10% level of testing.*

***Significant at 10% level only, not 5%.*

Table 4.2: Cont..

Discrete Choice Demand Estimates I

	2000		1998	
	Probit	OLS	Probit	OLS
Rural	-0.193	-0.054	-0.215	-0.023
	(0.030)	(0.012)	(0.027)	(0.013)
Central City	-0.001*	0.016*	0.037**	0.020
	(0.025)	(0.009)	(0.022)	(0.010)
MSA (large)	-0.059	0.018	-0.016*	-0.036
	(0.024)	(0.008)	(0.020)	(0.008)
South	-0.024*	0.022	-0.009*	0.007*
	(0.022)	(0.007)	(0.019)	(0.008)
Computers($\geq$ 2)	0.678	-0.036*	1.233	-0.036*
	(0.039)	(0.027)	(0.029)	(0.041)
Leased	0.037*	-0.015*	0.348	-0.033*
	(0.130)	(0.049)	(0.121)	(0.045)
Comp. bought 00	1.949	-0.087*		
	(0.032)	(0.058)		
Comp. bought 99	1.997	-0.108**		
	(0.029)	(0.060)		
Comp. bought 98	1.839	-0.113	0.834	-0.037*
	(0.029)	(0.054)	(0.030)	(0.027)
Comp. bought 97	1.740	-0.112	0.659	-0.040**
	(0.038)	(0.051)	(0.039)	(0.023)
Use outside home	-0.118	0.011*	0.022*	-0.021
	(0.026)	(0.009)	(0.023)	(0.010)
Median Income	0.048	0.012	0.035	0.015
	(0.015)	(0.005)	(0.013)	(0.005)
Heckman Coefficient***		0.121		0.036
		(0.047)		(0.049)
Constant	-1.433	2.889	-1.226	2.882
	(0.083)	(0.033)	(0.076)	(0.037)
N	40,011	14,635	41,618	10,236

Robust standard errors in parentheses.

Dep. var. is Internet access and monthly fee, for probit and OLS resp.

**Not significant at 10% level of testing.*

***Significant at 10% level only, not 5%.*

****Calculated from first step probit estimates.*

Table 4.2: Cont..

Utilization variables

Demand Estimates (Cont..)

Uses	2000	1998
Email	0.003*	0.024
	(0.009)	(0.01)
Online Courses	0.002*	-0.004*
	(0.009)	(0.010)
News	0.024	0.016
	(0.007)	(0.008)
Phone	0.071	0.037
	(0.015)	(0.014)
Information Searches	-.003*	-0.007*
	(0.008)	(0.009)
Job Search	0.006*	0.001*
	(0.009)	(0.010)
Job Related	0.014**	0.035
	(0.008)	(0.009)
Shopping	0.010*	0.007*
	(0.008)	(0.009)
Games/Entertainment	0.048	0.032
	(0.015)	(0.016)

Robust standard errors in parentheses.

**Not significant at 10% level of testing.*

***Significant at 10% level only, not 5%.*

Note that only qualitative conclusions can be drawn from the coefficients reported.

For the 2000 participation equation we find as expected households headed by older individuals are less likely to have the Internet, similarly lower income leads to lower probability of adoption. However we do not find any gender specific differences in adoption rates, higher education leads to a greater chance of adoption, however the surprisingly the college dummy is not significant. The race dummies are as expected with blacks and Hispanics having lower adoption rates. Married i.e. traditional families are more likely to have the Internet just as being single (never married) implies a lower adoption probability. Larger households are also more likely to have the Internet, however as number of children increase the adoption rate declines which might be because larger families are also on average poorer. The biggest surprise is that the employed dummy is never significant, alternatively an unemployment dummy as well a not in labor force was not significant (not reported), this may be due to extremely low rates of unemployment over the sampling period which led to small sampling bias.

Rural areas lag behind significantly in adoption rates, however there is no difference between city and suburban locations, the dummy for south was not significant (unlike previous studies). Larger MSAs actually on average have lower adoption rates. The technological variables behave as expected with households having more than one computers substantially more likely to have the Internet, as well as households with more recent computer purchases are likely to be users. Whether the computer is leased or bought does not affect the outcome. The outside usage dummy is insignificant this is partly since a distinction cannot be made whether it is used at work or in a public library etc. Richer MSAs naturally have higher adoption rates.

The results for 1998 are very similar, we briefly note the differences, first gender does affect adoption rates, this implies households with male heads did adopt early.

The college dummy is significant, however married or single dummy does not affect adoption rates unlike in 2000. Also living in a large MSA (with population ¿ million) does not affect the adoption rate. Lastly households with leased computers are actually more likely to adopt.

For the second step price of service regressions we find for 2000, neither age nor income or education matters, except for advanced degrees which lowers the price paid. Minority households on average pay higher prices which offers some support to the claim that adoption can be affected by higher price draws (it can also be caused by unobserved factors). Both married and single households pay less, so does surprisingly employed people, and households with more children but otherwise household size increase prices. Geographically rural consumers opt for less expensive services, and prices in larger MSAs as well as in the South are on average higher. People with older computers pay less for service and surprisingly none of the other technological controls affects prices. The coefficients are qualitatively very similar for 1998, however more variables are insignificant which might be due to the smaller size of the sample.

We do not find that after controlling for the demographic factors the utilization variables do not proxy for significant unobserved tastes. For 2000 we find that those who use the Internet for news or phone calls or playing games pay on average more for their services. None of the other variables are significant, however overall we can reject the hypothesis that all these coefficients are zero, so they do have some explanatory power albeit not substantial. The results are very similar for 1998 except those who use it for e-mail or job related uses are also more likely to pay more, this maybe because the early adopters were primarily those who used it for their work.

4.5 Extensions

4.5.1 Elasticity Estimates

A prerequisite for any discussion of policies based on prices (such as subsidies), is the *price elasticity of Internet access*, which is defined as the percentage change in overall Internet adoption rate caused by a one percent change in prices, $\frac{\partial E(y)}{\partial p}\frac{p}{E(y)}$, where $y_i = 1$ denotes access. Unfortunately for the two stage model estimated above there is no direct way to derive this from the estimated coefficients. Instead we estimate this indirectly through simulation (or sample enumeration).[19] Elasticity is estimated by considering the counterfactual situation where average prices of all types of contracts are 1% higher and predicting the access probabilities for everyone in the sample.

This is achieved as follows; first, the predicted probabilities of adoption are calculated from the probit estimates of the participation equation (4.23) $(\hat{\pi}_i)$. Then the mean of the price offer curve is shifted, and new predicted probabilities calculated for all individuals $(\tilde{\pi}_i)$. The price elasticity of access is obtained from the (weighted) average change in participation probabilities, ((4.24) below).[20] For example let the estimated constant from equation (4.15) be $\hat{\beta}_2^0$ then if all prices increase proportionately say by $\theta\%$, the mean of the distribution of $\log P^r$ (given $\log P^r \sim N(X'_{2i}\beta_2, \sigma_2^2)$ increases by $\Delta\beta_2^0 = \log\{(1 + \theta/10) * \exp(\hat{\beta}_2^0)\}$. Since the estimated constant for the probit model (from (4.23) above) is actually $\hat{\gamma}_0 = (\hat{\beta}_1^0 - \hat{\beta}_2^0)/\hat{\sigma}$, a $\theta\%$ change in offered prices is equivalent to a change in the constant term of $\Delta\gamma_0 = -\Delta\beta_2^0/\hat{\sigma}$. Therefore, $\hat{\gamma}_0$ is replaced by this new constant $(\gamma_0 + \Delta\gamma_0)$ and the corresponding new

[19] A similar method is followed by (Kridel, Rappoport, and Taylor 1999b) although their model is relatively more straightforward (probit).

[20] Where the weights are the sample weights, for the CPS this is interpreted as the inverse of the probability of selection or the number of households in the population that the i^{th} household is supposed to represent.

predicted probabilities of adoption $\hat{Pr}_i$ are calculated. Finally the elasticity estimate is obtained as follows, (in percentage terms)[21]

$$\eta = \sum_{i=1}^{N} \omega_i \left\{ \frac{\tilde{\pi}_i - \hat{\pi}_i}{\hat{\pi}_i} * \frac{100}{\theta} \right\} \tag{4.24}$$

where ω_i is the sample weight for household i.

The estimated price elasticities are reported in table (??). By turns we consider the whole sample and only MSAs. MSAs are considered as a separate sample because of two reasons, first, since like most new products the Internet was initially available only in the cities, considering only MSA data avoids the questions of availability. Also since they are the smallest relatively homogeneous geographical area that can be identified from the data. Thus we can construct a set of variables which seek to control for MSA level fixed effects.

The estimates are presented in sequence starting with the baseline model controlling for standard demographic variables such as income, education, age and race. Subsequently we add a number of other controls to check for the robustness of the estimates. We find that the estimates do not change significantly with the addition of further controls which we take as evidence confirming the estimates. In order we first add household characteristics such as size and number of children etc, then we add geographic variables such as controls for rural areas, large MSAs etc. Consequently we control for the level of technological savvy through controls for number of computers, the year when the latest computer was bought etc. Income effect refers to median income which is added separately since it's relative importance has been noted in other studies as *wealth* effects in adoption, also for non-MSAs this is hard to define and we substitute the state median income instead. Finally we add the

[21] In practice the elasticity was obtained by taking the average of the changes in predicted probabilities for a 10% increase and decrease in prices respectively.

Table 4.3:
Estimates of Price Elasticity

Model	2000	1998
(1) Baseline model	0.716	0.765
(2) Add household characteristics to (1)	0.434	0.041
(3) Add geographic controls to (2)	0.399	0.134
(4) Add household tech. vars. to (3)	0.416	0.834
(5) Control for income effect in (4)	0.418	0.847
(6) Add utilization variables *[in step 2]*	0.419	0.804
Sample: MSAs only		
(7) Complete model *[same as (5)]**	0.361	0.873
(8) Add controls for MSA fixed effects to (7)	0.369	0.874
(9) Add utilization variables in (8)	0.367	0.847
(10) Only dialup option	0.534	
(11) Add broadband price in (10)	0.463	

See text for model specifications.
**Poplulation of MSA also added.*

utilization variables in the second stage regressions and do not find any significant changes.

However given that there is no discernible trend in the estimates as more controls are added this raises questions regarding the efficacy of the controls. As reported earlier we find that all the controls used here are significant at 5% level of testing in the participation equation. For the 2000 sample we find that elasticity ranges from a low of 0.36 to a high of 0.43 with the only exception being the baseline model with only the basic demographic controls ($\eta = 0.72$). This means that a one percent increase in average prices leads to around a third of one percent increase in adoption rates, which is much higher when compared to the earlier studies and needs to be justified. We note that a higher elasticity compared to telephones is expected since telephones had long back reached saturation levels of adoption and therefore any further increments are unlikely to be caused by a fall in prices, whereas the Internet being a relatively new technology with adoption rates at around 44% implies a significant opportunity for growth. This hypothesis is also supported by the change in the elasticity over time, as we see from the table the average elasticity in 1998 was significantly higher (around 0.87%), when adoption rates were also much lower (28%). The (Kridel, Rappoport, and Taylor 1999b) study also obtained a significantly lower estimates however as noted before their study is marred by serious technical shortcomings.

We also did not find any significant differences between the estimates obtained for the different samples. The MSA sample which adds a number of MSA level controls does not alter the estimated elasticities significantly. Although on average they are lower for 2000 which is expected given higher existing penetration rates in cities. It is however on average higher for 1998 compared to the whole sample, which can be explained by bias introduced by availability, i.e Internet connections may not have been easily obtainable in non-MSA regions in 1998.

4.5.2 Exclusion Restrictions

Alternatively instead of assuming $\sigma_{12} = 0$ this model is also fully identified if we assume that certain variables that affect the offer price do not affect the reservation price i.e. some columns in X_2 are not present in X_1.[22] Note that conceptually this is very similar to using cost shifters as instruments in standard demand estimation. Although intuitive this procedure is usually harder to implement since the choice of instruments is not usually known a priori. Particularly for the Internet very little or no data is available in the public domain for the supply side i.e. for Internet service providers. Gronau (1973) also notes the essential arbitrariness of such exclusions.

One can argue that $\sigma_{12} \neq 0$ if there are unobserved factors that affect both reservation prices and offer prices. For instance there can be unobserved location specific fixed effects (which we tried to control for earlier), that can potentially affect both. For example the presence of a high tech industry (imagine Silicon Valley) implies a more technologically savvy population which can lead to higher reservation prices as well as higher prices as more sophisticated services are demanded (it can also presumably lead to lower prices due to more discerning or well-informed consumers).

Since theory does not suggest any such variable the intuitive approach would be to look for such variables in the earlier demand estimates, specifically we are looking for variables that are significant in the second step regression (OLS) but not in the first step probit. From table (4.5.1) we find that the two dummy variables for gender (male) and region (south) are significant in the price estimates but not in the adoption equation. Using male as the excluded variable gives an estimate of

[22] Scott and Garen (1994) uses religion and location as exogenous variables that affect participation but not the amount of purchase of lotteries in their study.

$\sigma = 2.837$ and elasticity of $\eta = 0.42$, this is very close to our earlier estimates obtained assuming $\sigma_{12} = 0$. A more appropriate choice perhaps would be the dummy for south since based on the model outlined the coefficients should have opposite signs in the two estimates. Using it gives an estimate of $\sigma = 0.919$ and corresponding elasticity estimate of $\eta = 1.307$. This is too high to be realistic.

Given the arbitrary nature of choosing instruments from demand side data and the closeness of the estimates obtained we use the earlier assumption of $\sigma_{12} = 0$ for further analysis reported below.

4.5.3 Consumer Surplus

The model estimated above can also be used to obtain estimates of consumer surplus. The surplus for the i^{th} consumer is defined as the difference between what she is willing to pay, her reservation price and what she ends up paying, the offer price, i.e.

$$CS_i = P_i^r - P_i^o$$

Given that tastes or unobserved factors are distributed normally over the population we need to take the average of this value, i.e. the consumer surplus for the average person with characteristics X_i is,

$$E(CS_i) = E_{u_1}(P_i^r) - E_{u_2}(P_i^o) \tag{4.25}$$

where $E(.)$ is the expectations operator defined over the variable in subscript. Then total consumer surplus is defined as,

$$E(CS) = \sum_{i=1}^{N} \omega_i [E_{u_1}(P_i^r) - E_{u_2}(P_i^o)] \tag{4.26}$$

where ω_i is the population weight. However since the price variables are defined in logs we obtain the predicted values for the mean of the logarithmic distribution, i.e. $E(\log P^r)$ and $E(\log P^o)$ respectively. Assuming $\sigma_{12} = 0$ (independence) we can obtain $E(P^r)$ and $E(P^o)$ using the standard transformation,

$$E(x) = \exp\{E(\log x) + Var(\log x)/2\} \tag{4.27}$$

Then equation (4.26) is used to obtain the aggregate consumer surplus.

4.5.4 Marginal Effects

The demand estimates report earlier were only of qualitative value i.e. they showed whether a higher income raised or lowered the probability of adoption of the Internet. We need to calculate the marginal effects of the exogenous variables to quantify their impacts on both participation and the price paid for Internet access. For the participation equation this is straightforwardly obtained as follows,

$$\frac{\partial \Pr(y_i = 1)}{\partial x_k} = \frac{\partial \Phi(Z_i'\delta)}{\partial x_k} = \delta_k \phi(Z_i'\delta) \tag{4.28}$$

where ϕ and Φ are the pdf and cdf of the standard normal, and x_k is the k^{th} column of X_i. The weighted mean of which is reported in table (??) below. From (Maddala 1983) it is known that,

$$E(\log P^o | y_i = 1) = X_{2i}'\beta_2 + \sigma_{2u}\frac{\phi(Z_i'\delta)}{\Phi(Z_i'\delta)} \tag{4.29}$$

The marginal effect in this case can be computed by differentiating this w.r.t x_k,

$$\frac{\partial E(\log P^o)}{\partial x_k} = \beta_k + \sigma_{2u}\delta_k \frac{\phi_i}{\Phi_i}\left[Z_i'\delta - \frac{\phi_i}{\Phi_i}\right] \tag{4.30}$$

where $\phi_i = \phi(Z_i'\delta)$ and $\Phi_i = \Phi(Z_i'\delta)$ respectively. Then the actual impact of the variable can be estimated in percentages as $\exp\{\frac{\partial E(\log P^o)}{\partial x_k}\} - 1$.

4.5.5 A Note on Broadband

A similar model was estimated for broadband which has been at the center of much discussion in recent times. There has been several papers in recent times noting the importance of broadband (Crandall and Alleman 2002). However empirical studies in this context remain few with the studies by Goolsbee and Klenow (2002) and Kridel, Rappoport, and Taylor (2002) being the notable exceptions. Broadband is defined as high-speed connections to the Internet, specifically we considered only cable modems and digital subscriber lines (DSL) although the survey designates a number of other forms of access such as wireless etc as also high-speed. However we concluded that given the limited data at our disposal accurate estimates of the price elasticity of broadband access cannot be obtained. We briefly discuss the negative results obtained since it raises serious questions regarding accuracy of the other estimates as well.

Kridel, Rappoport, and Taylor (2002) selected an arbitrary figure as the average price for non-subscribers and as before used the actual price paid for subscribers to broadband. There are several shortcomings to their approach which raises serious questions about the estimates obtained there, first, we found that for prices to be calculated accurately we chose MSAs with at least ten observations for broadband, surprisingly only fifty MSAs were left, which suggests that broadband availability across all locations is a serious concern, even partial deployment of broadband technologies voids any measures of elasticity obtained for the entire population. Since the data used by them is from a much earlier date availability is likely to be a more serious issue. Second we found the distribution of prices for broadband to be bimodal i.e. with twin peaks as can be seen from figure (ref), this can be explained by either two types of services being available in most areas or the more likely explanation is promotional pricing was highly prevalent for the period in question since broadband was being launched across much of the country in 2000. Therefore we found average prices in most locations to be substantially lower than considered by KRT.

The distribution of broadband prices leads to the belief that in most locations only a limited number of options existed for broadband, since most cable franchises enjoy a monopoly and DSL at least in 2000 was being provided only by the local telephone company.[23] Then the two step tobit model used previously fails since the basic assumption of differentiated product is violated. An alternative approach using an ordered probit setup was also used, which assumes a step system whereby individuals with the highest reservation prices adopting broadband early. However we found the estimates to be not significant.[24] This leads us to suspect that the data available is too sparse to estimate price elasticities. Given the bimodal nature of the price distribution an average high (normal) and low (introductory) price was also calculated for each location where broadband was widely available (at least ten subscribers in the sample). We did not find these to be significant either, the minimum price of broadband in each location does prove to be significant but this is not a reliable indicator given the strong endogeneity of prices (presumably locations where broadband was available early and more widely, also had more attractive/lower prices offered to lure more consumers).

[23] Regulations would make it mandatory for phone companies to allow competitors to also offer DSL connections over their wires, which led to the current situation where consumers do have a number of choices for broadband providers.

[24] These results are not reported to conserve space.

4.5.6 Tests

The alternative to the Type II Tobit model considered here is the standard Tobit model where the same set of variables decide participation and price paid.

4.6 Conclusion

This study obtained preliminary estimates of the price elasticity of access for Internet usage at home, using very limited public data from the CPS. Although we believe this to be a significant improvement over earlier studies which were marred by serious methodological errors, we acknowledge the scope for further study in this context.

The main caveat to the results obtained is that the data used for this study did not contain any product characteristics (for Internet subscriptions), although we did find strong evidence in favor of a differentiated product assumption and/or endogeneity of prices. Given this restriction this paper sought to control for this phenomenon indirectly, however future studies containing such characteristics would be necessarily more accurate. More detailed data is also essential for obtaining more accurate estimates of both static (current period) and dynamic (over the lifetime of the product) estimates of consumer welfare derived from this product. However the technique used here is robust in terms of accounting for the censoring of price data as well as allowing a more complete analysis by considering the participation and purchase decision separately.

Bibliography

AMEMIYA, T. (1984): "Tobit Models: A Survey," *Journal of Econometrics*, 24(1-2), 3–61.

ARTHUR, B. (1989): "Competing Technologies, Increasing Returns and Lock-in by Historical Events," *Economic Journal*, 99, 116–131.

AUTOR, D. H. (2001): "Wiring the labor market," *Journal of Economic Perspectives*, 15(1).

BAILEY, J. (1998): "Electronic Commerce: Prices and Consumer Issues for Three Products: Books, Compact Discs, and Software," *Organization for Economic Cooperation and Development OCDE/GD*, 98(4).

BAKER, M., AND A. MELINO (2000): "Duration Dependence and Nonparametric Heterogeneity: A Monte Carlo Study," *Journal of Econometrics*, 96, 357–393.

BALA, V., AND S. GOYAL (1995): "A Theory of Learning with Hetergeneous Agents," *International Economic Review*, 36(2), 303–323.

——— (1998): "Learning from Neighbours," *The Review of Economic Studies*, 65(3), 595–621.

BANERJEE, A. (1992): "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107, 797–817.

Barber, B. M., and T. Odean (2001): "The Internet and the Investor," *Journal of Economic Perspectives*, 15(1).

Bass, F. M. (1969): "A New Product Growth Model for Consumer Durables," *Management Science*, 15, 215–227.

Beckert, W. (2000): "Estimation of Stochastic Preferences: An Empirical Analysis of Demand for Internet Services," mimeo, University of Florida.

Berg, G. J. V. D. (2000): "Duration Models: Specification, Identification, and Multiple Durations," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. V. North-Holland.

Berndt, E. R., R. S. Pindyck, and P. Azoulay (2000): "Consumption Externalities and Diffusion in Pharmaceutical Markets: Anti-ulcer Drugs," *National Bureau of Economic Research Working Paper*, No. 7772.

Bertrand, M., E. F. P. Luttmer, and S. Mullainathan (2000): "Network Effects and Welfare Cultures," *Quarterly Journal of Economics*, 115(3), 1019–55.

Besley, T., and A. Case (1993): "Modeling Technology Adoption in Developing Countries," *American Economic Review*, 83(2), 396–402.

Bikhchandani, S., J. Hirschleifer, and I. Welch (1992): "A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, 992–1023.

Boozer, M., and S. Cacciola (2001): "Inside the 'Black Box' of Project STAR: Estimation of Peer Effects Using Experimental Data," mimeo, Economic Growth Center, Yale University.

Borenstein, S., and G. Saloner (2001): "Economics and Electronic Commerce," *Journal of Economic Perspectives*, 15(1).

Borjas, G. J. (1995): "Assimilation and Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s," *Journal of Labor Economics*, 13(2), 201–245.

Brock, W., and S. Durlauf (2001b): "Interactions–Based Models," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 5. North–Holland, Amsterdam.

Brynjolfsson, E., and M. Smith (2000): "Frictionless Commerce, A Comparison of Internet and Convetional Retailers," *Management Science*, 46(4).

Cameron, S. V., and J. J. Heckman (1998): "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy*, 106(2), 262–333.

Chatterjee, R., and J. Eliashberg (1990): "The Innovation Diffusion Process in a Heterogeneous Population: A Micromodeling Approach," *Management Science*, 36(9), 1057–74.

Compaine, B. M. (2001): "Information Gap: Myth or Reality," in *The Digital Divide: Facing a Crisis or Creating a Myth*, ed. by B. M. Compaine, pp. 105–18. MIT Press, Boston.

CPS (2002): "Design and Methodology," Technical report 63rv, U.S. Department of Labor, Bureau of Labor Statistics, available at www.bls.census.gov/cps/tp/tp63.

CRAGG, J. G. (1971): "Some Statistical Models For Limited Dependent Variables With Appliation To The Demand For Durable Goods," *Econometrica*, 39(5), 829–844.

CRANDALL, R. W., AND J. H. ALLEMAN (2002): *Broadband: Should We Regulate high-Speed Internet Access?* AEI-Brookings Joint Center for Regulatory Studies, Washington, D.C.

DAVIES, S. (1979): *The Diffusion of Process Innovations.* Cambridge University Press, Cambridge.

DICICCIO, T. J., AND B. EFRON (1996): "Bootstrap Confidence Intervals," *Statistical Science*, 11(3), 189–228.

DICICCIO, T. J., AND J. P. ROMANO (1995): "On bootstrap procedures for second-order accurate confidence limits in parametric models," *Statis. Sinica*, 5, 141–160.

EFRON, B., AND R. TIBSHIRANI (1993): *An Introduction to the the Bootstrap.* Chapman and Hill, New York.

ELBERS, C., AND G. RIDDER (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, 49(3), 403–410.

ERDEM, T., AND M. P. KEANE (1996): "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15(1), 1–20.

EVANS, W. N., W. E. OATES, AND R. M. SCHWAB (1992): "Measuring Peer Group Effects: A Study of Teenage Behavior," *Journal of Political Economy*, 100(5), 966–991.

FINKELSTEIN, D. M. (1986): "A Proportional Hazards Model for Interval-Censored Failure Time Data," *Biometrics*, 42, 845–854.

GANDAL, N., M. KENDE, AND R. ROB (2000): "The Dynamics of Technological Adoption in Hardware/Software Systems: The Case of Compact Disk Players," *RAND Journal of Economics*, 31(1), 43–61.

GEROSKI, P. A. (2000): "Models of technology diffusion," *Research Policy*, 29, 603–625.

GLAESER, E. L., B. SACERDOTE, AND J. A. SCHEINKMAN (1996): "Crime and Social Interactions," *Quarterly Journal of Economics*, pp. 508–548.

GLICK, R., AND A. K. ROSE (1999): "Contagion and Trade: Why are Currency Crises Regional," *International Journal of Money and Finance*, 18(4), 603–617.

GOOLSBEE, A., AND P. J. KLENOW (2002): "Evidence of Learning and Network Externalities in the Diffusion of Home Comupters," *Journal of Law and Economics*, 45(2), 317–343.

GRILICHES, Z. (1957): "An Exploration of Technological Change," *Econometrica*, 25(4), 501–522.

——— (1996): "The Discovery of the Residual: A Historical Note," *Journal of Economic Literature*, 34, 1324–1330.

GRONAU, R. (1973): "The Effect of Children on the Housewife's Value of Time," *Journal of Political Economy*, 81(2), S168–S199.

GRUBER, H., AND F. VERBOVEN (2001): "The Diffusion of Mobile Telecommunications Services in the European Union," *European Economic Review*, 45, 577–88.

HALL, B. H., AND B. KHAN (2003): "Adoption of New Technology," mimeo, National Bureau of Economic Research.

HALL, P. (1988): "Theoretical comparison of bootstrap confidence intervals (with discussion)," *Annals of Statistics*, 16, 1431–1452.

HANNAN, T., AND J. MCDOWELL (1984): "The Determinants of Technology Adoption: The Case of the Banking Firm," *Rand Journal of Economics*, 15(3), 328–35.

HAUSMAN, J. (1998): "Taxation by Telecommunications Regulation," *Tax Policy and the Economy*, 12, 29–48.

——— (1999): "Cellular Telephone, New Products, and the CPI," *Journal of Business and Economic Statistics*, 17(2), 188–94.

HAUSMAN, J., AND D. A. WISE (1981): "Stratification on an Endogenous Variable and Estimation: The Gary Income Maintenance Experiment," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. F. Manski, and D. McFadden, pp. 365–391. MIT Press, Cambridge, MA.

HECKMAN, J. J., AND B. SINGER (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models of Duration Data," *Econometrica*, 52, 271–320.

HENDRY, D. F. (1984): "Monte Carlo Experimentation in Econometrics," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. II, pp. 937–976. North-Holland, Amsterdam.

HENEBRY, K. L. (1996): "Do Cash Flow Variables Improve the Predictive Accuracy of a Cox Proportional Hazards Model for Bank Failure?," *Quarterly Review of Economics and Finance*, 36(3), 395–409.

HOFFMAN, D. L., T. P. NOVAK, AND A. E. SCHLOSSER (2001): "The Evolution of the Digital Divide: Examining the Relationship of Race to Internet Access and Usage over Time," in *The Digital Divide: Facing a Crisis or Creating a Myth*, ed. by B. M. Compaine, pp. 47–98. MIT Press, Boston.

HOROWITZ, J. L. (1996): "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable," *Econometrica*, 64(1), 103–137.

——— (1999): "Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity," *Econometrica*, 67(5), 1001–1028.

HUANG, J. (1996): "Efficient Estimation for the Proportional Hazards Model with Interval Censoring," *The Annals of Statistics*, 24(2), 540–568.

HUANG, J., AND A. J. ROSSINI (1997): "Sieve Estimation for the Proportional-Odds Failure-Time Regression Model with Interval Censoring," *Journal of the American Statistical Association*, 92(439), 960–967.

JORGENSON, D. W. (2001): "Information Technology and the U.S. Economy," *American Economic Review*, 91(1).

JOVANOVIC, B. (1979): "Job Matching and The Theory of Turnover," *Journal of Political Economy*, 87(5), 972–990.

KRIDEL, D., P. RAPPOPORT, AND L. TAYLOR (2002): "The Demand for High-Speed Access to the Internet," in *Forecasting the Internet, Understanding the Explosive Growth of Data Communications*, ed. by D. G. Loomis, and L. D. Taylor. Kluwer Academic Publishers, Boston.

KRIDEL, D. J., P. N. RAPPOPORT, AND L. D. TAYLOR (1999a): "An Econometric Study of the Demand for Access to the Internet," in *The Future of the Telecommunications Industry: Forecasting and Demand Analysis*, ed. by D. G. Loomis, and L. D. Taylor, pp. 21–42. Kluwer Academic Press, Boston.

——— (1999b): "An Econometric Study of the Demand for Access to the Internet," in *The Future of the Telecommunications Industry, Forecasting and Demand Analysis*, ed. by D. G. Loomis, and L. D. Taylor. Kluwer Academic Publishers, Boston.

LANCASTER, T. (1990): *The Econometric Analysis of Transition Data.* Cambridge University Press, New York, NY, 1st edn.

LITAN, R. E., AND A. M. RIVLIN (2001): *Beyond the Dot.coms: The Economic Promise of the Internet.* Brookings Institution Press, Washington D.C.

MACKIE-MASON, J., AND H. VARIAN (1995): "Pricing Congestible Network Resources," *IEEE Journal on Selected Areas in Communications*, 13(7), 1141–1149.

MADDALA, G. (1983): *Limited-Dependent and Qualitative Variables n Econometrics.* Cambridge University Press, Cambridge, UK.

MAHAJAN, V., E. MULLER, AND F. M. BASS (1991): "New Product Diffusion Models in Marketing: A Review and Directions for Research," in *Diffusion of Technologies and Social Behavior*, ed. by N. Nakicenovic, and A. Grubler, pp. 125–77. Springer, New York.

MANSKI, C. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531–542.

MANSKI, C. F. (1988): "Identification of Binary Response Models," *Journal of American Statistical Association*, 83, 729–38.

MELNIKOV, O. (2000): "Demand for Differentiated Durable Products: The Case of the U.S. Computer Printer Market," mimeo, Yale University.

MEYER, B. D. (1990): "Unemployment Insurance and Unemployment Spells," *Econometrica*, 58(4), 757–782.

MIN, I., AND J.-H. KIM (2003): "Modeling Credit Card Borrowing: A Comparison of Typle I and Type II Tobit Approaches," *Southern Economic Journal*, 70(1), 128–143.

MOFFITT, R. (2001): "Policy Interventions, Low-level Equilibria, and Social Interactions," in *Social Dynamics*, ed. by S. Durlauf, and H. Young, pp. 45–82. MIT Press, Cambridge.

MOKYR, J. (1990): *The Lever of Riches: Technological Creativity and Economic Progress.* Oxford University Press, Oxford, UK.

NTIA (2000): "Falling Through the Net, Toward Digital Inclusion," Technical report, National Telecommunications and Information Agency, available at www.ntia.doc.gov/ntiahome/digitaldivide/.

PERL, L. J. (1978): "Economic and Demographic Determinants of Residential Demand for Basic Telephone Service," Discussion paper, National Economic Research Associates, Inc.

ROBERTS, J. H., AND J. M. LATTIN (2000): "Disaggregate-Level Diffusion Models," in *New-Product Diffusion Models*, ed. by V. Mahajan, E. Muller, and Y. Wind, pp. 207–236. Kluwer Academic Publishers, Boston.

ROGERS, E. (1995): *Diffusion of Innovations.* The Free Press, New York, 4th edn.

ROSE, N., AND P. JOSKOW (1990): "The diffusion of new technologies: evidence from the electric utility industry," *Rand Journal of Economics*, 21, 354–373.

SCHMIDT, P., AND A. D. WITTE (1989): "Predicitng Criminal Recidivism Using Split Population Survival Time Models," *Journal of Econometrics*, 40, 141–159.

SCOTT, F., AND J. GAREN (1994): "Probability of Purchase, Amount of Purchase and the Demographic Incidence of the Lottery Tax," *Journal of Public Economics*, 54, 121–143.

SOLOW, R. M. (1957): "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, 39, 312–320.

STONEMAN, P., AND P. DAVID (1986): "Adoption Subsidies Vs. Information Provision as Instruments of Technology Policy," *Economic Journal*, Supplement 96, 142–50.

TAYLOR, L. D. (1994): *Telecommunications Demand in Theory and Practice.* Kluwer Academic Publishers, Boston.

TAYLOR, L. D., AND D. J. KRIDEL (1990): "Residential Demand for Access to the Telephone Network," in *Telecommunications Demand Modelling*, ed. by A. de Fontanay, M. Shugard, and D. Sibley. North–Holland, Amsterdam.

WOOLDRIDGE, J. M. (2001): "Asymptotic Properties of Standard M-Estimators for Standard Stratified Samples," *Econometric Theory*, 17, 451–470.