Estimating Diffusion Models Using Repeated Cross-sections: Quantifying the Digital Divide*

Mainak Sarkar †
Department of Economics
Yale University

This draft: September 30, 2003

Abstract

This paper shows that a broad class of dynamic models of diffusion of new technology (usually estimated using duration models) can be estimated using data with extreme censoring i.e. when only discrete choice data is available in the form of *repeated cross-sections*. This finding should have broad relevance across numerous fields where such models are currently used, since it opens up new potential sources of data. For example expenditure surveys are commonly available for most countries in this format. We present Monte Carlo evidence to show that the maximum likelihood estimator is consistent and efficient in this context. We also apply this methodology to estimate a diffusion model of Internet services at home for U.S. consumers. We use publicly available *Current Population Survey (CPS)* data to estimate these models. Consequently we are able to test for the existence of and, estimate the dimensions of the so-called *Digital Divide.* These estimates also allow us to make forecasts of Internet penetration in the United States across various levels of aggregation (geographic and demographic).

*I would like to thank Patrick Bayer, Hanming Fang, Martin Pesendorfer, Gustav Ranis and Subrata Sen for many helpful discussions, seminar participants at Yale and also other IO folks at Yale who have all provided valuable suggestions and encouragement at different times.

†Corresponding Address: Department of Economics, Yale University. 37 Hillhouse Avenue, New Haven. CT 06520, e-mail:mainak.sarkar@yale.edu

1 Introduction

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 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 Waverman (2000) 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 (1999) reports similar findings for Internet access. Earlier studies such as Beckert (2000) also report a low

[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.

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

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 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

[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).

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 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 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)

[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.

[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.

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 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

[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.

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)).

3 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.

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

[9] A dynamic programming model with Bayesian learning process is usually assumed.
[10] Self-reported past date of adoption.

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

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

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 $\mathbf{X}$ 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 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

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

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

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

$$\max_t R(t) - C(t|x), \quad t \in [0, \infty) \tag{2}$$

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.* $\epsilon, \lambda(x) \geq 0$.

[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 $\epsilon \to 0$ in the limit.

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 $c > 0$, this is weakly convex and also increasing over time. Then the first order conditions from problem (2) implies the following:

$$\begin{aligned} R'(t^*) &= C'(t^*|x) \quad or, \\ a - bt^* &= c\lambda(x)\epsilon \quad or, \\ t^* &= \frac{a - c\lambda(x)\epsilon}{2b} \end{aligned} \tag{3}$$

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:

$$\begin{aligned} \Pr\left(t^* \leq T|x\right) &= \Pr\left(\frac{a - c\lambda(x)\epsilon}{2b} \leq T\right) \\ &= \Pr\left(\frac{a - 2bT}{c\lambda(x)} \leq \epsilon\right) \\ &= 1 - \mathbf{\Phi}\left((1/\sigma^2)\left\{\frac{a - 2bT}{c\lambda(x)} - 1\right\}\right) \end{aligned} \tag{4}$$

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:

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

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.

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

[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 $\epsilon = exp(x_u'\beta_u)$.

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

$$\frac{d\rho(t)}{dt} = \frac{d}{dt}[\log(R'/c')] = \frac{c'^2}{R'}\left(\frac{R''}{c''} - c'\right) \tag{5}$$

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:

$$\Pr(t \leq T|\mathbf{X} = x) = \Pr\left(\rho(T) + x'\beta) \leq \log(\epsilon)\right) \tag{6}$$

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

$$\begin{aligned} \Pr(t \leq T|\mathbf{X} = x) &= \int_{\rho(T)+x'\beta}^{\infty} g(\log \epsilon) d \log \epsilon \\ &= 1 - \mathbf{\Phi}\left(\frac{\rho(T) + x'\beta}{\sigma_{\log \epsilon}}\right) \end{aligned} \tag{7}$$

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.

[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 $E(\epsilon) = 0$.

3.1 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.

4 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.

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t} \tag{8}$$

and by definition this can be shown to be,

$$h(t) = \frac{f(t)}{S(t)} \quad where \quad S(t) = 1 - F(t) \tag{9}$$

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

$$S(t) = \Pr(T \geq t) \tag{10}$$

4.1 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 (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 (9) and using the Leibniz rule if we differentiate equation (7) we find

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{F'(t)}{1 - F(t)} = -\rho'(t) \left\{ \frac{g(\rho(t) + x'\beta)}{1 - F(t)} \right\} \tag{11}$$

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 (11) gives us the following hazard rate:

$$h(t) = G(\log \epsilon) = -\rho'(t)[1 + exp\{\rho(t) + x'\beta\}]^{-1} \tag{12}$$

We know that $\rho'(t) < 0$ (from (5) above), therefore if we assume $exp[\rho(t)] = t^{\alpha}$ where $\alpha < 0$. We can write the hazard rate as follows:

$$h(t) = \frac{(-\alpha)t^{\alpha-1}exp(x'\beta)}{1 + exp(x'\beta)t^{\alpha})} \tag{13}$$

which is simply the hazard rate for the log-logistic model (see table 1 below).

Also if we assume that $\alpha = -1$ then $\rho(t) = -\log t$, then by using the definition of the survivor function from (9) and (11) we can write

$$S(t) = 1 - F(t) = \mathbf{\Phi}\left(\frac{-\log t + x'\beta}{\sigma_{\log \epsilon}} \right) \tag{14}$$

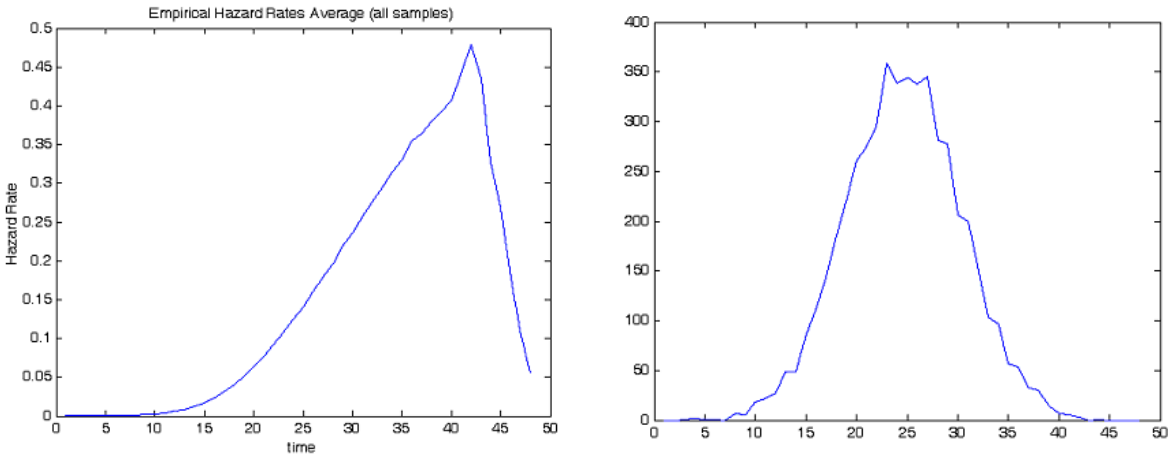
note that this survivor function is identical to the lognormal model where $\log t$ is distributed as

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

as defined below in table 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 (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]



Hazard Distribution $f(t)$

Figure 1: Simulation Results for the Adoption Model

4.2 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

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

[18] For examples see Lancaster (1990)

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).

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 1: Various Duration Models

The most commonly used models in this literature are summarized in table 1. Perhaps the most widely used model is the Weibull partly due to its simplicity, however 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 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.

4.3 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

[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.

which is updated as new information comes in every period (instantaneously in a continuous time setup). The Wiener process can be written as follows:

$$dz(t) = \mu dt + \eta(t)\sigma\sqrt{dt} \tag{15}$$

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]

$$f(t) = \frac{\alpha}{\sigma t^{3/2}} \phi\left(\frac{\alpha - \mu t}{\sigma\sqrt{t}}\right) \quad \forall\ t \geq 0 \tag{16}$$

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.

5 Estimation

5.1 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 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.

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

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

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.

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

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.

5.2 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

$$\mathcal{L} = \prod_{y_{i(t)}=1} Prob(\tau_i \leq t) \prod_{y_{i(t)}=0} Prob(\tau_i > t) \tag{17}$$

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 (9) above),

$$\mathcal{LL} = \sum_{t \in \{t_1, t_2, \ldots, t_M\}} \sum_{i=1}^{N_m} \left[y_{i(t)} \log(1 - S_i(t)) + (1 - y_{i(t)}) \log S_i(t) \right] \tag{18}$$

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 (1) we can write its log-likelihood as follows:

$$\mathcal{LL} = \sum_{t \in \{t_1, t_2, \ldots, t_M\}} \sum_{i=1}^{N_m} \left[y_{i(t)} \log(1 - exp\{(exp(-X'\beta)t^{\alpha})\}) - (1 - y_{i(t)})(exp(-X'\beta)t^{\alpha}) \right] \tag{19}$$

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

5.3 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 (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:

$$S(t) = \exp\left(-\int_0^t h(s)ds\right) \tag{20}$$

and using the definition of mixture models from table (1) (for the moment assume that ν is known and the survivor function conditional on ν is S_ν), we get

$$S_\nu(t) = \exp\left(-\int_0^t \nu \exp(x'\beta)h_0(s)ds\right) \tag{21}$$

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

$$S_\nu(t) = \exp\left(-\nu \exp(x'\beta)[\alpha_0 t + \alpha_1 t^2/2 + \alpha_2 t^3/3]\right) \tag{22}$$

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

Alternatively from (21)

$$S_\nu(t) = \exp\left(-\nu \exp(x'\beta) \int_0^t h_0(s)ds\right) \tag{23}$$

Then let us define the variable $\exp(\gamma(t)) = \int_0^t h_0(s)ds$ which when substituted in (23) gives us

$$S_\nu(t) = \exp\left(-\nu \exp[x'\beta + \gamma(t)]\right) \tag{24}$$

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.

$$S(t) = \int \exp\left(-\nu \exp[x'\beta + \gamma(t)]\right) f(\nu;\eta)d\nu \tag{25}$$

Similarly for the polynomial case. The log likelihood is obtained by substituting expression (25) in (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 (25), as follows:

$$S(t) = \left[1 + \sigma^2 \exp\{x'\beta + \gamma(t)\}\right]^{1/\sigma^2} \tag{26}$$

6 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. 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.

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 $N = 100$.

[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.

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.

Table 2: Weibull Hazards with Repeated Cross-sections

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 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,

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 3: Lognormal/Log-logistic Hazards with Repeated Cross-sections

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 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.

	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

Table 4: Proportional Hazards with Repeated Cross-sections

The proportional hazards results are reported in tables (4) and (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 5 we consider two alter-

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

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

native 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.

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

Table 5: Prop. Haz. with RCS II (Quasi-likelihood approach)

7 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.[31] 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.*[32] 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) and how much they paid for the connection.

For an overview of the methodology followed in designing the survey refer to *Technical Paper 63RV* (2002).[33] The survey uses a rotating panel, i.e. each household is interviewed for four months, then rested for eight months and then again interviewed for four months before they are retired permanently. This implies that the data is in the form of *repeated cross-sections* and not any sort of panel data. Other relevant demographic and economic data at various levels of geographic aggregation was also obtained from the 2000 census.

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

[32] 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.

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

7.1 Descriptive Analysis

Here we present some salient features of the data. We conduct our analysis at the household level since we believe the decision to have Internet service at home is a household decision. In the data a particular member of the household is designated as a primary householder or a reference person (generally the person who rents/owns the family home). All demographic variables like age, sex, marital status etc., refers to this person in the household. Table 6 reports the characteristics of the 2001 sample, the samples analyzed for the other years were found to be very similar in characteristics (not reported here). 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.

Variable	**Mean** 2001	**Variable**	**Mean** 2001
N	56,634	**Education**	
Male	0.537	Less than HS	0.151
Age	49.21	HS Dip. (or GED)	0.312
Income		College (or Ass. Deg.)	0.273
$\leq 25,000$	0.262	Bachelor's Degree	0.17
25,000-50,000	0.249	Advanced Degree	0.094
50,000-75,000	0.154	MSA	0.746
$\geq 75,000$	0.171	Central City	0.227
Ethnicity		Rural (Non-MSA)	0.249
White	0.853	Northeast	0.218
Black	0.102	South	0.294
Asian / Pac. Isle.	0.033	Midwest	0.257
Hispanic*	0.074	West	0.234

Source: 2001 Current Population Survey data.

Table 6: Some Descriptive Statistics

7.2 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[34] 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 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.[35]

[34] 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.

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

7.3 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 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:

$$P_{it} = \frac{K_i}{1 + e^{-(a_{it} + b_{it}t)}} \tag{27}$$

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:

$$log\left(\frac{P_{it}}{K_i - P_{it}}\right) = a_i + b_i t + \epsilon_{it} \tag{28}$$

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.

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.

Table 7: Diffusion Process

For our purposes we are only interested in aggregate diffusion for the whole country. In table 7 below we report the aggregate diffusion for the these two technologies, we find

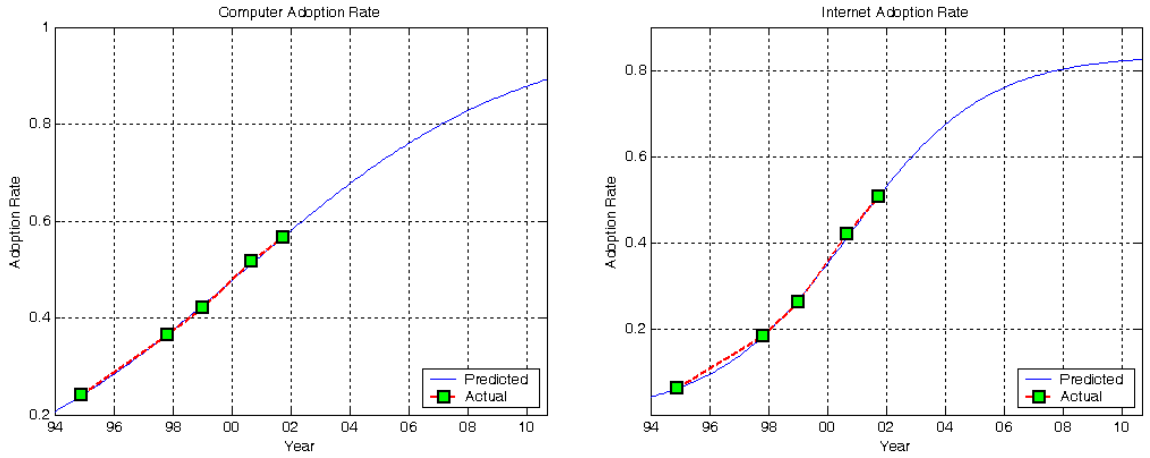
that computers which have been available for much longer had ownership of around 24% by the beginning of this study and increased to about 57% by the end (2001). Whereas the Internet which was available to the general public only in the early nineties had a usage level of about 6% by the beginning of 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 2 (a)-(b) below. Figure 2 (b) fits the logistic model for the trend in Internet usage and similarly figure 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.[36] 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.

Table 8 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.[37] The baseline technology used by most households to access the net remains the dialup modem with prices for access remaining roughly stationary over the period for which we have data

[36] 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 28 above.

[37] 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.



(a) Computer (b) Internet

Figure 2: Internet and Computer Ownership Trends

(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 increases 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.[38] This is inspite of the fact that cable was more expensive compared to DSL on an average.[39] 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.[40]

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

[39]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.

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

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 /			5.06	19.44	
Satellite				(11.55)	
d) Other (ISDN)			10.8	19.12	
				(12.51)	

Standard errors in parentheses

Table 8: Type of Access

8 Results

8.1 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 9. 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.

8.2 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 explosive growth. Each survey date is coded as months from this date. In table 10 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 11 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.

Table 9:

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 10:

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 11:

Duration Models II

Proportional Hazard cols.(3–4)

	Weibull	Log-logistic	Log-logistic QMLE	No Het.	Gamma 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 11:

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 12:

Duration Models III

Log-Logistic Model

Fixed Effects, Large Urban Sample#

	MSAs*	Counties*	10 Largest MSAs	Restricted 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.*

8.3 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 10 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 11 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.

8.4 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 11. 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.

8.5 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 11, 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:

$$-2[L_R - L_U] = -2[-92,851 + 92,212] = 1278 \quad \Rightarrow \quad \Pr(\chi_1^2 \geq 1278) \approx 1.0$$

In table 12 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.[41] 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

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

and we found given the highly non-linear nature of the log-likelihood it was not feasible to include all such 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 12. We do not find any significant differences from our baseline estimates in table 11, 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.[42] 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 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 12, 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.

$$-2[L_R - L_U] = -2[-13,190 + 13,180] = 20 = \Pr(\chi^2_9 \leq 20) = 0.982$$

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

8.6 Other Models

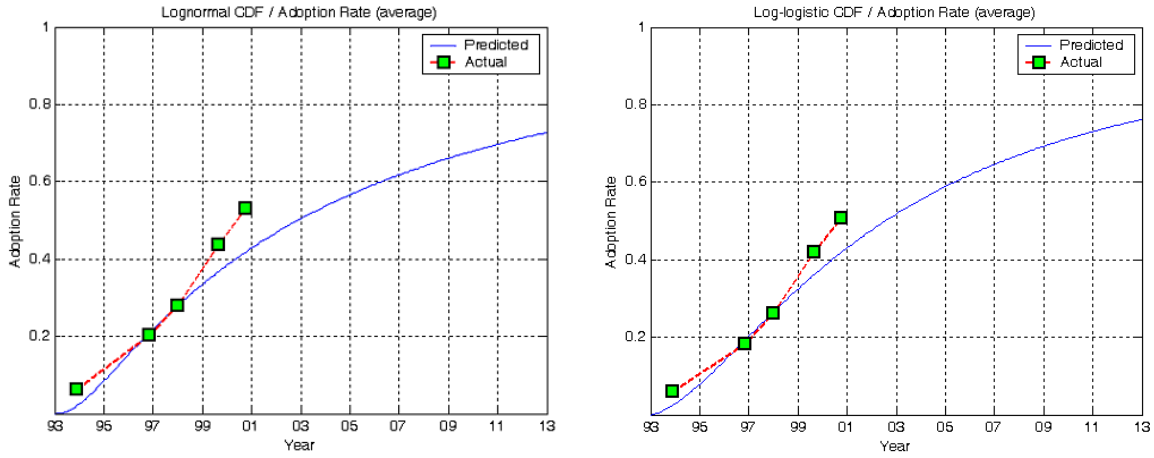
We considered by turns the non-parametric MLE suggested by Heckman and Singer (1984), however as noted by other authors (for example see Baker and Melino (2000)), we found the maximization routine failed to converge. In this situation others have arbitrarily assumed a binomial distribution and estimated 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.

9 Predictions for Individuals

9.1 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 3 (a)–(b).

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 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



(a) Lognormal (b) Log-logistic

Figure 3: Predicted vs. Actual Distribution function

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[43] 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 $110,085$.

The evidence is presented in table 13.[44] 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 13 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

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

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

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 mode. Estimation $N = 110,673$.

Validation sample $N = 110,085$

Table 13: Accuracy of Individual Predictions

and take the top 30% of the population and calculate this statistic as 34.3%.[45] 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.[46]

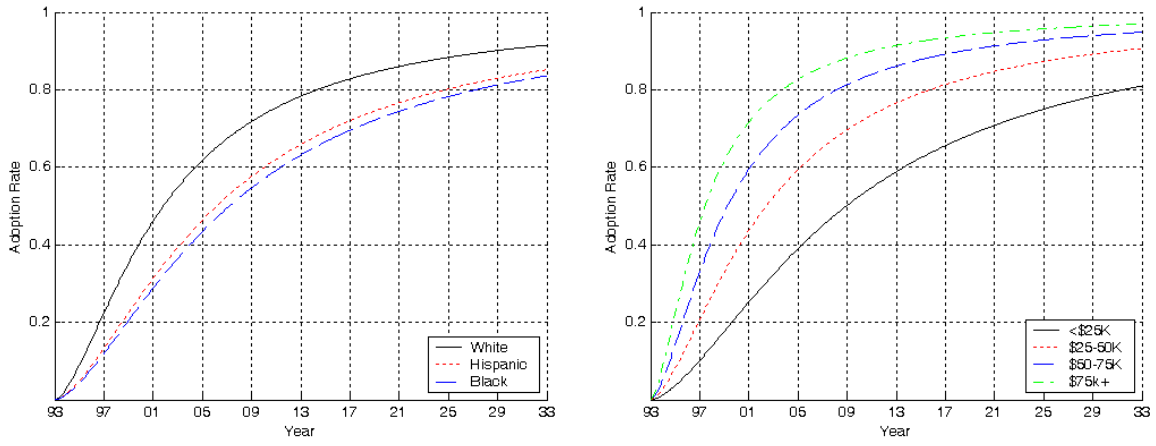
[45] Since from the table among the top 30% of the predicted adoptions 65.7% do and the rest don't.

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

9.2 Forecasting Diffusion Patterns

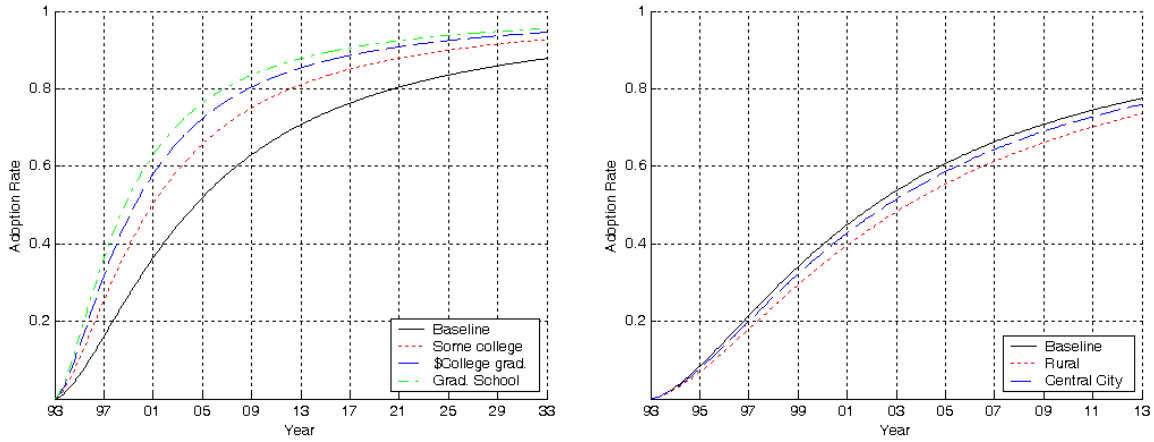
An attractive feature of duration models is that it allows us to forecast adoption rates at various levels of aggregation once the parameters of the underlying model has been estimated. We consider four dimensions of the digital divide over the next several years and plot the results implied by the full model in figures 4 and 5. The divide in terms of race has perhaps received the most attention, we find in figure 4(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 4(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 5(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 (once all other demographic variables are controlled for) either now or developing in the near future.



(a) Race (b) Income

Figure 4: Predicted Racial and Income Divides



(a) Education (b) Rural

Figure 5: Predicted Education and Rural Divide

10 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.

References

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

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.

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

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.

Crandall, R. W., and L. Waverman (2000): *Who Pays for Universal Service*. Brookings Institution Press, Washington, D.C.

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

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.

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.

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.

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.

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.

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.

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

Kridel, D. J., P. N. Rappoport, and L. D. Taylor (1999): "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.

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

Maddala, G. (1983): *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge, U.K.

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. 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.

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.

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

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