



HAL
open science

Difference between Adoption and Access Frequency to Internet and Consumer Surplus

Walid Hadhri, Mohamed Ayadi, Adel Ben Youssef

► **To cite this version:**

Walid Hadhri, Mohamed Ayadi, Adel Ben Youssef. Difference between Adoption and Access Frequency to Internet and Consumer Surplus. *Internet econometrics*, Palgrave Macmillan, pp.107-129, 2012, 9780230362925. halshs-00937177

HAL Id: halshs-00937177

<https://shs.hal.science/halshs-00937177>

Submitted on 28 Jan 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Difference between adoption and access frequency to Internet and consumer's surplus

Walid HADHRI¹

ADIS and UAQUAP -Univ. of Paris Sud and Univ. of Tunis

Mohamed AYADI²

UAQUAP - Univ. of Tunis

Adel BEN YOUSSEF³

GREDEG - Univ. of Nice Sophia-Antipolis

Abstract

The objective of this paper is twofold: first we use the methodology developed by Goolsbee and Klenow (2006) (G&K) that estimates the consumer's surplus of Internet's connection for France in 2005. Second, our paper challenges the initial methodology developed by G & K by using two complementary hypotheses and a Heckman's two stage estimation method. The first hypothesis take into account the concavity of the demand function of Internet and the second adds more realistic non monetary variables. We also make some differentiation between Internet adoption and access frequency patterns using Heckman's (1976) correction procedure to resolves the selection problem. We find that French time opportunity cost is three times more important than connection cost. We find also that high-income people were more able to adopt Internet, but they spend less time online than low-income ones. In 2005, the French consumer surplus ranged between 1240\$ and 3126\$ if we use the G& K methodology, between 1679\$ and 3126\$ if we use our concave demand function, but between 2107\$ and 2651\$ if we use our two stage estimation method.

Key words: Consumer Surplus, Internet, ICT uses, time opportunity cost.

JEL Classification: D6, L0

¹ walid.hadhri@u-psud.fr

² ayadimed2001@yahoo.fr

³ adel.benyoussef@gmail.com

Introduction

New products and services have significantly affected how households live, consume and allocate their time between different activities. Their adoption has grown rapidly last decade. These new goods and services have a significant impact on consumer's welfare. Many studies have tried to quantify these economic impacts by measuring consumer's surplus (Nevo 2001; Petrin, 2002, Goolsbee & Petrin, 2001; Greenwood & Kopecky, 2007)⁴.

The Internet is a new service whose adoption and access frequencies are growing rapidly. Modeling the Internet demand function and quantifying its impacts on consumer's welfare is an important analytical and empirical challenge (Hausman, 2002; Brynjolfsson et al. 2003; Gentzkow, 2006; Goolsbee and Klenow, 2006). Goolsbee and Klenow 2006 (G & K) have estimated this Internet contribution to the consumer's surplus during 2005 and they found that it is ranging between 2500 and 2800 US\$ in the United States. They use a specific two arguments utility function modeling a link between the time opportunity cost of time dedicated to Internet and income. Since this initial work, according to our knowledge little empirical findings were developed in order to compare their results and methodologies within other countries.

The aims of this paper are twofold: first, we estimate a demand function and a consumer surplus of Internet for French households using the same methodology than G&K applied on the French Household Survey Data for 2005. Second we challenge this methodology by considering more realistic hypothesis. We suppose a concave demand function for Internet instead of the G&K's linear demand function. We also assume that the intensity of Internet use could be affected by some specific households' characteristics and some other activities like TV watching or computer gaming. We use Heckman methodology estimating separately, Internet adoption and access intensity patterns to resolve the selection bias problem.

Many households, for various reasons, do not use the Internet. They cannot afford access, or they aren't aware about Internet services. They may reject Internet uses or are simply unable to use it (this is the case for the poorly educated or the disabled people). There are significant differences in Internet adoption decision and access. Adoption alone is not necessarily the only measure to evaluate Internet usage effect on consumer welfare. Heckman has developed two-steps method to correct selection bias (identify factors contributing to the selection bias). The Heckman method attempts to control for the effect of non-random selection by incorporating both the observed and unobserved factors that affect non-response.

Our results show that French households have found Internet to be a valuable addition to their welfare levels. In 2005, the French consumer surplus ranged between 1240\$ and

⁴ Nevo (2001) has calculated the consumer gains from the introduction of the Ready-to-eat cereal industry. Petrin (2002) finds large consumer effects from the introduction of the Minivan. Goolsbee and Petrin (2001) have calculated the gains from direct broadcast satellites and the competition with cable television. Greenwood and Kopecky (2007) measure the welfare gains from Personal Computers in Canada. Another way of studying these impacts is to consider the Price Index.

3126\$ if we use the G& K methodology, between 1679\$ and 3126\$ if we use our concave demand function, but between 2107\$ and 2651\$ if we use our two stage estimation method.

The paper is organized as follows. Section 1 discusses the econometric specifications for the basic Internet demand function. Section 2 depicts the Internet adoption and frequency differentiation for Internet demand analysis. Section 3 presents the data and the variables used. Section 4 discusses the econometric results. Section 5 concludes.

1. Econometric specification of the Internet demand function

Goolsbee and Klenow (2006) identify a link between the opportunity cost of time (time devoted to other uses compared to time devoted to Internet) and the wage rate. They consider that Internet users maximize the following separable utility function:

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

Where (L_I) and (L_O) are fraction of time spent using Internet services and on other goods and services, respectively. (C_I) and (C_O) are respectively the consumption time devoted to Internet services and the consumption of other goods and services. θ is a weight allocated to Internet sub utility. Finally, $(1 - \alpha_I)$ and $(1 - \alpha_O)$ are time intensities of Internet and other goods and services respectively.

Starting from a standard model of leisure/work trade off, Internet users have to consider the following budget constraint:

$$P_I C_I + F_I + P_O C_O = W(1 - L_I - L_O) \quad (2)$$

P_I and P_O are prices of the Internet services and of other goods respectively. F_I is the fixed cost of the Internet access. P_I the marginal cost of using Internet services. We consider that the marginal cost is zero because Internet access is priced as a flat monthly fee. W is the average monthly wage.

The combined Cobb-Douglas bundles are denoted as $U_I = C_I^{\alpha_I} L_I^{1-\alpha_I}$ and $U_O = C_O^{\alpha_O} L_O^{1-\alpha_O}$.

ρ_I and ρ_O are the weighted average of the market price and the price of time (i.e. the wage). This allow us to write the following prices:

$$\rho_I = \left(\frac{P_I}{\alpha_I}\right)^{\alpha_I} \left(\frac{W}{1-\alpha_I}\right)^{1-\alpha_I} \quad \text{and} \quad \rho_O = \left(\frac{P_O}{\alpha_O}\right)^{\alpha_O} \left(\frac{W}{1-\alpha_O}\right)^{1-\alpha_O}$$

Then, the optimal choices for the bundles become

$$U_I = \frac{W-F_I}{\rho_I(1+\Delta)} \quad \text{and} \quad U_O = \frac{W-F_I}{\rho_O(1+\Delta)}$$

Where,

$$\Delta = \left(\frac{\rho_I}{\rho_O}\right)^{\sigma-1} \left(\frac{1-\theta}{\theta}\right)^\sigma$$

The optimal choices the bundles into their consumption and time counterparts are:

$$\begin{aligned} C_I &= \frac{\alpha_I \rho_I U_I}{P_I} & C_O &= \frac{\alpha_O \rho_O U_O}{P_O} \\ L_I &= \frac{(1-\alpha_I) \rho_I U_I}{W} & L_A &= \frac{(1-\alpha_O) \rho_O U_O}{W} \end{aligned}$$

Using these optimal choices we can get the following expression for Δ :

$$\Delta = \frac{(1-\alpha_I)(1-F_I/W) - L_I}{L_I}$$

The fixed cost F_I is typically very small relative to full income (0.3 per cent for our sample). The Internet access costs are small flat fees so that $F_I/W = 0$. Consequently, we adopt the assumption that there is no marginal use pricing. This allow us to write $\alpha_I = 0$. Thus, the above expression of Δ becomes approximately equal to the time spent on other activities $(1 - L_I)$ relative to the time spent on Internet activities.

$$\Delta = \frac{1 - L_I}{L_I}$$

Using the prices ρ_I and ρ_O in the following expression can be rewritten Δ (according of the wages) as follows:

$$\Delta = AW^{(\alpha_O - \alpha_I)(\sigma-1)} \left(\frac{1-\theta}{\theta}\right)^\sigma$$

Where $A = \left[\frac{(P_I/\alpha_I)^{\alpha_I(1-\alpha_O)^{1-\alpha_O}}}{(P_O/\alpha_O)^{\alpha_O(1-\alpha_I)^{1-\alpha_I}}} \right]^{(\sigma-1)}$

Consumer maximizes an utility function (eq. 1) subject to his budget constraint (eq. 2) Giving the fact that G&K internet demand function which have two arguments: time opportunity cost (time devoted to other uses compared to time devoted to Internet) and wage rate⁵, we can write the expression (3) as follows:

$$\ln \left[\frac{1-L_I}{L_I} \right] = \ln(A) + (\alpha_A - \alpha_I)(\sigma - 1) \ln(W) + \sigma \ln \left(\frac{1-\theta}{\theta} \right) \quad (3)$$

The left hand side of equation (3) is the log of ratio of time spent on non Internet activities $(1-L_I)$ relative to time spent on the internet L_I . $\ln(A)$ is a constant term across

⁵ For more details see Goolsbee and Klenow (2006).

consumers. $(\alpha_0 - \alpha_I)$ measures the difference between time intensity of the Internet and other goods. σ is the substitution elasticity between the two bundles. As θ may change from one Internet user to another, G&K consider $\sigma \ln\left(\frac{1-\theta}{\theta}\right)$ as an error term, which will be denoted by ε in the following equations.

In order to compare our results to those obtained by G&K we consider the same specification defined by G&K applied for the French households' survey data. In a second time, we consider alternative specifications adding two more realistic components (concavity of wage effect and Non Income Variables).

Blundell and al. (1993) consider that consumer demand patterns typically found in micro data sets vary considerably across households with different household characteristics and with different levels of income. We model this variability by allowing for nonlinear $\ln(W)$ terms. To describe individual household preferences, we first abstract from all over explanatory variables except income and write our demand function as follows:

$$\ln\left[\frac{1-L_I}{L_I}\right] = \alpha + \sum_{j=1}^L \beta_j g_j(\ln(W))$$

$g_j(\ln(W))$ are known polynomials in W .

This form is sufficiently general to cover many of the popular forms for Engel curves. To illustrate these points more explicitly, we consider a quadratic extension of our demand equation. In this model $L=2$ and the g_j 's are simply polynomial logarithmic terms, so that our demand equation may be written as

$$\ln\left[\frac{1-L_I}{L_I}\right] = \alpha + \beta_1 \ln(W) + \beta_2 [\ln(W)]^2 \quad (4)$$

Internet demand function may increase when wage increases, however, over a certain level of wage the impact seems to be smaller and people do not increase their connection time at the same level. We have some kind of concavity of Internet demand function.

Young people and well-educated persons are using more intensively Internet than older people. Seniority, which represents the Internet experience in Internet use, modify Internet preferences. On the other hand, Internet is considered to be in competition with other activities like playing computer's games. Those two sets of variables must be added to the Internet demand function as Non Income variables (NINC). We can write the expression (4) like the followings:

$$\ln\left[\frac{1-L_I}{L_I}\right] = \alpha + \beta_1 \ln(W) + \beta_2 [\ln(W)]^2 + \gamma \text{NINC} + \varepsilon \quad (5)$$

2. Internet adoption and access frequency differentiation for Internet usage analyses

In our sample, data are censored. Indeed, we will be able to identify factors underlying access frequency (or Internet demand) only if Internet access frequency is higher than zero.

Internet demand function of Internet user does not, in general affords a reliable estimate of what non users would have demand if they have become users. Internet demand function estimated on selected samples does not in general, estimate population demand functions. Comparisons of Internet demand of present users with demand of non users result in a biased estimate of the effect of the random treatment of users.

Characterization of the selection bias problem (Heckman 1979)

Consider a random sample of I observations. Equations for individual i are like the followings:

$$Y_{1i} = X_{1i}\beta_1 + U_{1i} \quad (6a)$$

$$Y_{2i} = X_{2i}\beta_2 + U_{2i} \quad (6b)$$

$i=1, \dots, I$

Where X_{1i} is a $1 \times K_j$ vector of exogenous regressors, β_i is a $K_j \times 1$ vector of parameters, and

$$E(U_{ji}) = 0, \quad E(U_{ji}U_{j'i''}) = \sigma_{jj'} \text{ if } i = i'' \\ = 0 \text{ if } i \neq i''$$

The final assumption is a consequence of a random sampling scheme. The joint density of U_{1i} , U_{2i} is $h(U_{1i}, U_{2i})$.

Suppose that one seeks to estimate equation (6a) but that data are missing on Y_1 for certain observations. The population regression function for equation (6a) may be written as

$$E(Y_{1i}|X_{1i}) = X_{1i}\beta_1$$

However, the regression function for the subsample of available data is

$$E(Y_{1i}|X_{1i}, \text{sample selection rule}) = X_{1i}\beta_1 + E(U_{1i}|\text{sample selection rule})$$

If the conditional expectation of U_{1i} is zero, the regression function for the selected subsample is the same as the population regression function. Least squares estimators may be used to estimate β_1 on the selected subsample. The only cost of having an incomplete sample is a loss in efficiency. However in the general case, the sample selection rule that determines the availability of data has more serious consequences.

Suppose that data are available on Y_{1i} if $Y_{2i} > 0$, while if $Y_{2i} < 0$, there are no observations on Y_{1i} . In the general case

$$\begin{aligned} E(U_{1i} | \text{sample selection rule}) &= E(U_{1i} | X_{1i}, Y_{2i} \geq 0) \\ &= E(U_{1i} | X_{1i}, Y_{2i} \geq -X_{2i}\beta_2) \end{aligned}$$

In the case of independence between U_{1i} and U_{2i} , so that the data on Y_{1i} are missing randomly, the conditional mean of U_{1i} is zero. However in the general case, it is nonzero and the subsample regression function is

$$E(Y_{1i} | X_{1i}, Y_{2i} \geq 0) = X_{1i}\beta_1 + E(U_{1i} | U_{2i} \geq -X_{2i}\beta_2) \quad (7)$$

The selected sample regression function depends on X_{1i} and X_{2i} . Regression estimators of the parameters of equation (1a) fit on the selected sample omit the final term of equation (7) as a regressor, so that the bias that results from using non-randomly selected samples to estimate behavioral relationships is seen to arise from the ordinary problem of omitted variables.

Heckman's estimator and its properties

Assume that $h(U_{1i}, U_{2i})$ is a bivariate normal density. Using well known results of Johnson and Kotz (1972) :

$$\begin{aligned} E(U_{1i} | U_{2i} \geq -X_{2i}\beta_2) &= \frac{\sigma_{12}}{(\sigma_{12})^{\frac{1}{2}}} \lambda_i \\ E(U_{2i} | U_{2i} \geq -X_{2i}\beta_2) &= \frac{\sigma_{22}}{(\sigma_{22})^{\frac{1}{2}}} \lambda_i \end{aligned}$$

Where

$$\lambda_i = \frac{(Z_i)}{1 - \Phi(Z_i)} = \frac{(Z_i)}{\Phi(-Z_i)}$$

Where ϕ and Φ are the density and distribution function for a standard normal variable respectively.

" λ_i " is the inverse of Mill's ratio which is a monotone decreasing function of the probability that an observation is selected into the sample, $\Phi(-Z_i) (= 1 - \Phi(Z_i))$, where $Z_i = -\frac{X_{2i}\beta_2}{(\sigma_{22})^{1/2}}$

The full statistical model for normal population disturbances can now be developed. The conditional regression function for selected samples may be written as

$$E(Y_{1i}|X_{1i}, Y_{2i} \geq 0) = X_{1i}\beta_1 + \frac{\sigma_{22}}{(\sigma_{22})^{1/2}} \lambda_i,$$

$$Y_{1i} = E(U_{1i}|X_{1i}, Y_{2i} \geq 0) + V_{1i} \quad (8)$$

Where

$$E(V_{1i}|X_{1i}, \lambda_i, U_{2i} \geq -X_{2i}\beta_2) = 0$$

$$E(V_{1i}^2|X_{1i}, \lambda_i, U_{2i} \geq -X_{2i}\beta_2) = \sigma_{11} \left((1 - \rho^2) + \rho^2(1 + Z_i\lambda_i - \lambda_i^2) \right)$$

Where

$$\rho^2 = \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}}$$

and

$$0 \leq 1 + \lambda_i Z_i - \lambda_i^2 \leq 1$$

If one knew Z_i and hence λ_i , one could enter λ_i as a regressor in equation (8) and estimate that equation by ordinary least squares. The least squares estimators of β_1 and $\sigma_{12} / (\sigma_{22})^{1/2}$ are unbiased but inefficient.

In practice, one does not know λ_i . But in the case of a censored sample, in which one does not have information on Y_{1i} if $Y_{2i} < 0$, but one does know X_{2i} for observations with $Y_{2i} < 0$, one can estimate λ_i by the following procedure (Heckman, 1979):

(1) Estimate the parameters of the probability that $Y_{2i} > 0$ (i.e., $\beta_2/(\sigma_{22})^{1/2}$) using *probit* analysis for the full sample.

(2) From this estimator of $\beta_2/(\sigma_{22})^{1/2}$ ($= \beta_2^*$) one can estimate Z_i and hence λ_i . All of these estimators are consistent.

(3) The estimated value of λ_i may be used as a regressor in equation (8) fit on the selected subsample. Regression estimators of equation (8) are consistent for β_1 and $\sigma_{12} / (\sigma_{22})^{1/2}$ (coefficients of X_{1i} and λ_i , respectively).

Internet demand function with censored sample

We define a variable $Adopt_i$ such that:

$Adop_i = 1$ if individual i adopt Internet;

$Adop_i = 0$ if individual i do not adopt Internet.

We note Y_{1i} the log of ratio of time spent on non Internet activities ($1-L_i$) relative to time spent on the internet L_i . However

$$Y_{1i} = \alpha + \beta_1 \ln(W) + \beta_2 [\ln(W)]^2 + \gamma NINC \quad \text{if } Adop_i = 1$$

$$Y_{1i} \quad \text{unobserved} \quad \text{if } Adop_i = 0$$

Using Heckman's two step selection method we will consider two separate equations: a selection equation, estimating the probability of Internet adoption which help us to evaluate inverse of Mill's ratio, and the Internet demand equation adjusted for selection bias, used to estimate the determinants of Internet access frequency.

(1): Selection equation: Probability of Internet adoption

We consider a latent variable $Y_{2i} = X_{2i}\beta_2 + \varepsilon_{2i}$ (9a)

where X_{2i} is a vector of individual-level controls, including demographics and wages we suppose ε_{2i} has normal distribution $N(0; \sigma_2)$.

We suppose that

$$Adopt_i = 1 \text{ if } Y_{2i} > 0 ; \quad (9b)$$

$$Adopt_i = 0 \text{ if } Y_{2i} \leq 0$$

(2): Internet demand equation adjusted for selection bias:

Internet access frequency is observed only if individuals adopt Internet. From equation 9, we can write:

$$adop_i = 1 \text{ if } \varepsilon_{2i} > (-X_{2i}\beta_2)$$

$$adop_i = 0 \text{ if not}$$

According to Heckman (1979), we can write the equation of access frequency:

$$E(Y_{1i}/W_i, NINC_i, adopt_i = 1) = \alpha + \gamma NINC_i\beta_2 + \beta_1 W + \beta_2 W^2 + E(\varepsilon_{2i} / adopt_i = 1)$$

Thus the access frequency equation on the selected sample depends at the same time on X_{2i} , W_i and $NINC_i$.

$$E(Y_{1i}/W_i, NINC_i, adopt_i = 1) = \alpha + \gamma VSD_i + \beta_1 W + \beta_2 W^2 + E(\varepsilon_{2i} > -X_{2i}\beta_2)$$

Indeed,

$$E(Y_{1i}/W_i, NINC_i, adopt_i = 1) = \alpha + \gamma VSD_i + \beta_1 W + \beta_2 W^2 + \rho \sigma_2 \frac{(X_{2i}\beta_2)}{\Phi(X_{2i}\beta_2)} \quad (10)$$

where, ρ is the coefficient of correlation of the errors terms ε_{1i} and ε_{2i} . ϕ and Φ are, respectively, the density and distribution function for a standard normal $N(0; \sigma_2)$.

We follow Heckman two stages estimate procedure giving consistent estimators for the parameters of our model.

On the first stage: we estimate the parameter $\beta_2/\sigma_2 \equiv \beta_2$ (where $\sigma_2 = 1$) using maximum likelihood for probit model (9a, 9b):

On the second stage: we apply the Ordinary Least Squares (OLS) to the access frequency equation under the assumption of normality of the residuals:

$$E(Y_{1i}/adopt_i = 1) = \alpha + \gamma NINC_i + \beta_1 W + \beta_2 W^2 + \rho \sigma_2 \hat{\lambda}_i + \vartheta_i \quad (11)$$

Where $\hat{\lambda}_i = \frac{(X_{1i}\hat{\beta}_1)}{\Phi(X_{1i}\hat{\beta}_1)}$ is a consistent estimator of the inverse of Mill's ratio $\lambda_i = \frac{(X_{2i}\beta_2)}{\Phi(X_{2i}\beta_2)}$.

3. Data and variables definitions

Data

In order to estimate the parameters of Internet demand function, we have used the French Household survey data of October 2005. 5603 representative households were questioned about their Internet preferences and consumption patterns.

In order to carry out the approximation of the intensity of Internet use (the dependent variable of our model), we consider the number of days of connection per individual. Thus, following Goolsbee and Klenow (2006) assumptions, we compute the frequency of access in the form of classes and we approximate each variable by its average as Table 1 shows.

Number of times per month	Type of frequency	Day average numbers per stage	Percentage of population
All days or almost	Strong	30	26%
At least four times per month	Average	15	13%
From one to three times per month	Very weak	2	8%
Not user	Null	0	53%

Table1. Access time of the Internet by stages

On average, in 2005, each French was connected to Internet for 4.64 hours per week (IDATE, 2005) against 7.7 hours per week in the United States during the same period, which corresponds to 4.1 per cent of his non-sleeping time⁶. The monthly Internet subscription cost in France is 25 euros for the DSL connection and 9.46 euros for the low broadband (IDATE, 2005). In our sample, 71 per cent individuals access to DSL connection against 29 per cent access low-level connections. Thus, the average cost of connection (FI) is 20.49 euros per month, which represents 0.38 per cent of the average income⁷.

Variables

There are a number of determinants affecting households adoption and Internet use such as: Demographic Factors, Geographic location and housing, Main location of Internet access, types of application or service used, Level of Educational, Internet and computer skills and peoples lifestyle.

The following aspects of claimed Internet adoption and frequency access behavior were measured:

Demographic Factors

Traditionally, demographic variables have always been examined as the initial predictors of Internet adoption and use. We control demographic factors such as level of income, age, number of children at home, marital status of respondent (married or not) and owning or not of housing.

Consistent with our reasoning from our theoretical model and the literature, we develop four hypotheses for empirical tests.

Income ($\ln(W)$) is a very strong determinant of household Internet adoption and access. It is expected to have significant and positive impact related to the intention to adopt Internet, and a significant and negative impact on how time is spent individually on line. The variable age (AGE) is expected to have the same effect (negative) on the Internet adoption and frequency access, i.e., younger Internet users spent more time online. For a larger number of studies,

⁶ We consider that a person may sleep about 8 hours per day. Thus the non-sleeping available time may be 112 hours per week. This time is allocated between Internet connection (4.1per cent), other leisure (60.2 per cent) and work (35.7 per cent).

⁷ For the United States Goolsbee and Klenow obtain 0.33 per cent.

number of children in a household also becomes a significant factor of Internet adoption and Internet frequency access.

We enlarge our analysis beyond effect of the usual demographics and geographic location on Internet frequency access to the effect location of Internet access and reasons for using the Internet.

Main location of Internet access

An Internet user is defined as someone who has had access to the Internet in the last month. On INSEE survey respondents were required to indicate the present access location to the Internet from a list of possible locations. The survey asked individuals about usage of Internet at friends or at other family members (*Internet Friends/Family*), at a cybercafés (*Internet café*) and at school and university (*Interne University/School*).

Furthermore we analyze also the effect of access to Internet on a Laptop (*Internet Laptop*). Those variables are expected to have positive effect on Internet frequency access.

Those five variables are both dummies variables.

Internet services used

There are now thousands of Internet 'home pages' which serve as information sources for individuals, institutions and organizations.

Most administrations, banks, universities, public and private organizations provides information through the Internet. Prospective people can access information on their banking account, playing games, listen music... The World Wide Web also provides very easy access to some government documents and legislative materials.

Respondents who have access to the Internet were asked to indicate their specific uses of the Internet. We analyze the effect of many Internet services used on the time spent on line. More specifically, the use of the Internet for home banking (*Home Banking*), playing games on line (*Playing Games*), listening and downloading music (*Music Online*), accessing administrative information (*Administrative Information*) and to online shopping (*Online Shopping*). Those services are used generally by active users.

Those five variables were measured as a dummy variable relative to 1 if respondent uses each of these Internet services and 0 if not.

The correlation between decision to adopt Internet and education, ICT skills and lifestyle is well established in the literature. One of our aims is to analyze the relation between those three factors and Internet adoption decision.

Level of Education

In many publications researchers consider education as another major factor affecting Internet adoption decision. They find that people with a high level of education are more likely to adopt Internet. Indeed, higher levels of education increase Internet adoption decision.

For the education variables, we make the distinction between two levels high school diploma (*High school graduate*) and university graduate (*University/college graduate*).

Computer and Internet Skills

Our study examines the role of computer skills (*Computer Skills*) and Internet skills (*Internet Skills*) on Internet adoption. Skills were considered as individual's ability to use the computer and the Internet efficiently and effectively.

Many studies raise that the question of Internet adoption inequalities does not relate to income, education, age, location... but it is more related to ICT skills. People with higher levels of e-skills are more likely to adopt Internet.

Responders were asked to answer seven computer skills such as, can: copy or delete a file, use the copy/delete tool to move information in a document, install/uninstall a software, install a new material (printer, modem, etc), use basic arithmetic formulas in a spreadsheet software (Excel, Quattro, Lotus, etc), compress or decompress files (by using Winzip, Winrar, Winace for example) and write a computer program by using a specific data-processing language (ex: visual BASIC, FORTRAN, java, C++)... and five Internet skills such as: use a search engine (yahoo, google, here...), send emails with attached files (document, photograph), visualize the history of the visited pages, remove temporary files and cookies and to create or update a web site.

Those answers are summed on two scores (Computer skills score and Internet skills score).

Lifestyle

ICT equipment, as mobile phone (*Mobile Phone*), laptop (*Laptop*), DVD (*DVD*) and numerical camera (*Numerical Camera*), influences positively the probability of Internet adoption. They bring to light the taste of the individuals for the technologies. Those variables were measured by dummy variables which took 1 if household owning each of these ICT tools and 0 if not.

4. Econometric Results

We estimated first the decision or the probability of the Internet adoption. This analysis consists of a simple probit model in which the dependant variable is probability of the Internet adoption decision. It's a dummy variable that is equal to one for the adopters and zero for the non-adopters. The independent variables are various factors which would have an effect on this probability, such as age, educational level, income, number of children in the family...

Secondly, we estimated the Internet use, based on the time that the individual spends on line. The dependant variable represents the number of hours per week that the individual spends connected to the Internet. Several explanatory variables are the observed factors that are supposed to have an influence on the results, such as income, age, Internet connection type (low or high band), location of Internet use...The variable for the correction of selection problem”(inverse Mills ratio) is obtained in the first stage.

It was necessary to control for bias due to heteroscedasticity problems involved by correction of the selection problem. Stata uses the Huber-White estimator to control for the bias due to clustering. This technique deflates the standard errors of the parameter estimates, in this case the coefficients, correcting the inference statistics.

The output from this estimation is displayed in Table 2. Estimated elasticity was computed. Table 2 reports basic G&K model, Internet access frequency models (without and with Non INCOME variables) and Internet adoption equation estimated by Probit model.

Table 2 estimation results indicates a statistical significance of all the covariates. One can conclude that all the expected theoretically signs of our coefficients are verified.

Variables	Basic Model		Heckman Access Frequency Model				Internet Adoption Model	
	Coefficients	T-stat	Basic Regression		Model with NINC		Coefficients	T-stat
			Coefficients	T-stat	Coefficients	T-stat		
Socio-Economic Factors								
ln(W)	0.2064***	4.92	0.1502***	3.91	0.0777**	2.12	0.0466***	7.71
[ln(W)] ²	-0.0288***	-5.20	-0.0195***	-3.92	-0.0106**	-2.26	-	-
Age	-	-	-	-	-0.0030*	-1.87	-0.0087***	-5.10
# Children in household	-	-	-	-	0.0251	1.46	0.0380*	1.91
Married	-	-	-	-	0.0733*	1.84	0.0797*	1.87
Owner of housing	-	-	-	-	-0.0012	-0.03	0.0915**	2.13
Main Location of use								
Internet Friends/Family	-	-	-	-	0.1564***	4.05	-	-
Internet café	-	-	-	-	0.1126*	1.78	-	-
Internet University/School	-	-	-	-	-0.1302**	-2.02	-	-
Internet Laptop	-	-	-	-	-0.1035***	-2.92	-	-
Services used								
Home Banking	-	-	-	-	-0.2422***	-6.97	-	-
Playing Games	-	-	-	-	-0.1331***	-3.27	-	-
Music Online	-	-	-	-	-0.1092***	-2.92	-	-
Administrative information	-	-	-	-	-0.1460***	-3.93	-	-
Online shopping	-	-	-	-	-0.2187***	-6.74	-	-
Level of Education								
High school graduate	-	-	-	-	-	-	0.1744***	3.85
University/college graduate	-	-	-	-	-	-	0.6159***	11.96

Skills								
Computer Skills	-	-	-	-	-	-	0.0369***	2.89
Internet Skills	-	-	-	-	-	-	0.6401***	43.2
Lifestyle								
Mobile Phone	-	-	-	-	-	-	0.0369*	1.80
Laptop	-	-	-	-	-	-	0.4416***	8.22
DVD	-	-	-	-	-	-	0.2370***	5.01
Numerical Camera	-	-	-	-	-	-	0.1892***	4.84
Constant	3.7142***	113.8	3.4824***	104.04	4.1028***	41.42	-1.7783***	-13.9
P			0.6512			0.4774		
Σ			0.9594			0.8585		
λ (Mills)			0.6247***			0.4099***		
# obs	2462		5603		5603		2462	
Prob>F	0.0000				-			
Prob>Chi2					0.0000			
Elasticity	0.40		0.62		0.78		-	
*: Significant at 10%, **: Significant at 5%, ***: Significant at 1%								

Table2. Internet Adoption and Heckman Access Frequency

Internet adoption model

By observing all the explanatory variables of the Internet adoption model, it can be outlined that almost all variables have a statistically significant effect on Internet adoption and our results confirm most of the theoretical expected effects.

Firstly, we found a strong relationship between Internet use and income level. Higher income means greater affordability and higher usage levels of Internet and thus we would expect a positive association between higher income and higher probability of Internet usage.

Our results confirm that higher incomes people have higher probability to adopt Internet (Table 2, column eighth and column nine). Furthermore, income is statistically significant factor that increases the probability of Internet adoption.

On the other hand, many studies have found that younger people are more able to use Internet. Our results confirm this assestion. Age seems to be one of the most important determinants of Internet adoption. The older are the individuals, the lower are the Internet usage. For instance, a one per cent increase in the individuals' age yields a significant decrease of around 0.8 points of percentage of probability of Internet adoption.

Higher levels of income and younger people are more willing to use Internet.

Household size and type is an important determinant in Internet adoption. This can be seen using both the measure of household type and the presence of children. Big families and (families with children) have higher probability of Internet adoption. Our results confirm that households with a single family have higher probability to adopt Internet. This result confirms that younger people are often asking for Internet access in particular for communication uses

(chat, forum, email...) or P2P. We have found significant and positive coefficients of the variables “*Married*” and “*Children in household*”.

Our econometric results reported in Table 2 show that housing statute occupation has a significant and positive effect on Internet adoption. Our results demonstrate that households owning a house had a higher probability to use Internet, compared to households renting their house.

There is a strong link between education and the use of Internet services. Our results confirm earlier findings that a higher level of education is associated with a higher level of household Internet use. Indeed, we have a correlation between higher school or university graduate and the probability of adopting the Internet. The more the level of education increases the more the probability to use the Internet increases.

Our results related to income and education effect confirm the assertion of Chaudhuri and al. (2005) according to which the income and the level of education have a significant positive effect on Internet adoption.

In order to explain the usage differences of Internet, many studies show that the level of ICT skills term appears to be one of the most important factors. Internet adoption is positively correlated with computer and Internet skills. Our econometric estimation demonstrate that computer and Internet skills increase the probability of Internet access or adoption. The coefficients of the variables "Computer Skills" and "Internet Skills" are positively and statistically significant. This confirms the idea that skilled people have a higher probability to adopt Internet.

Finally, we obtain an interesting result is the weak link between ICT equipment and uses and the probability of Internet adoption. Our results confirm that lifestyle, which indicates ICT and electronic tool uses positively correlate with Internet use. All the other variables have a positive and statistically significant coefficient. Indeed, using an ICT or electronic tool, as Mobile Phone, Laptop, DVD or digital camera influences positively the probability of adoption of Internet. We are in the presence of "technophiles households" or wired lifestyle (Hoffman and al., 1998).

Our findings confirm the fact that Higher education level, computer and Internet skills and lifestyle have a positive effect on Internet adoption.

Heckman Access Frequency Model

As we have seen, the estimates generally have the expected sign and are statistically significant. Many studies have found that Internet usage frequency is directly related to various socioeconomic factors such as age, income, marital status... Our purpose is to add interesting factors affecting this frequency such as location of use, which take into account whether the Internet is used alone or with other people and the kind of Internet services used.

The Mills ratio is significant, and indicates that the two-stage estimation procedure is appropriate.

As wages increase, time devoted to Internet access decreases. This result, already found by Goolsbee & Klenow (2006), confirms the seminal Becker (1965) theory on time allocation. We ask for more leisure activities as income increases. High-income groups have more disposable income to spend on these technologies than lower income groups. The opportunity cost of time is more important for high-income, and thus as income increases we spend less time on the Internet (Goldfarb and Prince, 2008). Moreover, if wage raises, time-intensive activities, such as the Internet, become more expensive, and wage rises will lead to a shift in consumption.

However, the relation between wage and Internet use is nonlinear, as the effect of $\ln(W)$ on Internet use increases but remains less than proportionally. The estimation results confirm this nonlinearity as the sign of $[\ln(w)]^2$ is negative and statistically significant. We conclude that the Internet intensity may be an increasing but concave function.

The econometric estimation shows that the coefficient of regression between the logarithm of wages and time devoted to the Internet is positive and is equal to 0.07.

According to this value, we can conclude that the more the person is rich the less he surfs on the web, therefore the more significant the income, the more the time of connection decreases. This result is primarily related to the opportunity cost of time. Indeed, the incomes lost during one hour are more important for the rich person than for the poor one and the individuals on low level of income have a cost of the time of weak leisure which is due at the reduced level of the wages. Also, this is due to the Internet utility. i.e. the individuals on low level of income find the Internet more useful than the others, because it their badly does not provide services which they do not find elsewhere. The individuals on low level of income have a leisure time more important than the others.

All non income variables (NINC) have the expected signs. The age of the household head (*Age*) has a positive effect, i.e., the older the household head the higher is his intensity of Internet use. This variable can resume the seniority effect. Thus, persons having greater experience in the Internet uses may have a more intense Internet use due to this experience and spend more time in order to reach several services (such as: email, home banking, information...). Furthermore, as the Internet is a network technology, perhaps at this stage of diffusion (2005), the offline network is still the social dominant network. In fact, seniority has a positive impact on building social networks.

Household income level has a negative effect on time devoted to Internet connection. However age has a positive effect on time spent online.

The coefficient for marital status of respondent (*Married*) is positive and significant at the 10 percent level of significance. Our results suggest a negative and significant relationship between marital status and time spent online. One argument is that the individual time devoted to the family reduces time devoted to Internet connection. Thus, married people may be disposed to spend less time online.

The presence of children and home ownership do not have an effect over time on connection to the Internet. Coefficients of these two variables are statistically insignificant.

Number of Internet users increased strongly. Indeed, location of Internet use outside home increased. Our results show that location of use is a strong predictor of time spent online. Access to Internet at friends home or at other family members home (*Internet Friends/Family*) and at a cybercafés (*Internet café*) has significant and negative effect on time devoted to Internet connection. However, access to Internet at school and university (*Internet University/School*) is correlated positively with Internet frequency use. This result can be explained by the offline network.

Types of Internet connection and access to Internet on a laptop are positive correlation with time devoted to Internet connection.

Finally, our econometric results reported in Table 2 (column six and column seven) show that purpose of Internet use has a positive effect on Internet frequency use. The coefficient for home banking (*Home Banking*), playing games on line (*Playing Games*), listening and downloading music (*Music Online*), accessing administrative information (*Administrative Information*) and online shopping (*Online Shopping*) are all positives and significant at the one percent level of significance. Using those services, give us an indication of the effect of the experience in the use of Internet. Different activities reflect their needs, personality characteristics, and tastes of Internet users.

Home banking, playing games on line, listening and downloading music, accessing administrative information and online shopping have positive and significant effects on time spent online.

Difference between Internet adoption and access frequency

In our econometric estimations we note a difference between the effects of various variables on the adoption and the frequency of access to Internet, such as: income level, age and marital status. Besides, we note differences in the statistical significance in other variables, such as presence of children and home ownership. Thus, our empirical results are consistent with our hypotheses:

- (1) Difference in household income effect: Income level has a positive effect on Internet adoption. However, conditional on adoption, it has negative effect on time spent online.

- (2) Difference in household age effect: age has a negative effect on Internet adoption. However, conditional on adoption, it has positive effect on time spent online
- (3) Location of Internet access has a significant impact on the time spent on line.
- (4) Purpose of Internet use has a positive and significant effect on time spent online.
- (5) The higher the education and skills levels of head of household, the higher the chance of Internet adoption, and
- (6) Lifestyle has positive effect on Internet adoption.

5. Internet's Consumers surplus

In order to approximate the consumer's surplus we use the Equivalent Variation Value (EV) associated to our Internet demand function. This approach was developed and used firstly by Hausman (1981, 1997, 1999) and Hausman and Newey (1995). They consider the formula (12) as an estimation of the Equivalent Variation Value.

$$EV = 0,5 \times L_I / [\sigma(1 - L_I(1 - F_I/W))] \quad (12)$$

Where F_I is the average cost of Internet connection, L_I time spent on the Internet activities, and σ is the price elasticity.

Results reported in Table 3 show that on average the consumer surplus obtained in the United States is higher than the one obtained in France for the same period for all our estimates. This consumer's surplus gap may be associated to three complementary explanations. On the one hand, in 2005, diffusion delays still between France and the United States regarding Internet connection. Internet adoption and usage were less significant in France than in the United States. Also, the time devoted to Internet in the United States (7.7 hours per week) is higher than the French one (4.64 hours per week). This delay is narrowing nowadays, the consumer's surplus gap will be reduced. The consumer's surplus gap may be affected also by the income differences between the US and France. GDP per capita in France is lower by 23.68 per cent than the American one. Finally the surplus gap could be explained by cultural differences which act on the collective consumer's preferences, as the Internet demand elasticity between the two countries is a little bit different.

However, one can see the difference between our specification and Goolsbee and Klenow one's. Our estimation, taking into account the concavity of the demand function of Internet demand leads to lower level of consumer surplus. It seems for us more appropriate to take into account these facts and to consider this relationship even in order to compare with the results obtained in the United States.

Variables	United States	France1	France2	France3
Time of connection (hours/week)	7.7	4.64	4.64	4.64

Time devoted to the Internet	6.9%	4.1%	4.1%	4.1%
Part of the budget devoted to the Internet	0.33%	0.38%	0.38%	0.38%
Part of the budget by associating the leisure cost	0.12%	0.14%	0.14%	0.14%
Surplus fraction in the income	2.9%	2.6%	2.6%	2.07%
Elasticity	1.32	0.40	0.62	0.78
Consumer's surplus (\$)	2500-3800 \$	1786 \$	2651 \$	2107 \$

a: France1: *Basic Model*,

b: France2: *Heckman Access Frequency Model (Basic regression)*,

c: France3: *Heckman Access Frequency Model (Model with NINC variables)*

Table3. Internet consumer's surplus comparison between the United States and France

6. Conclusion

The objective of this article was to measure in a simple way the consumer's surplus when the service consumption has a strong component in terms of time devoted to their use. From this point of view the Internet is illustrative of these new services. We carried out calculations of the opportunity cost of this activity in France by setting realistic assumptions starting from the statistics provided by the survey of INSEE 2005 and the available statistics at the IDATE and the ARCEP. To solve the problem of selection, we are based on the method of estimate in two stages of James Heckman (1979). We noted difference between the effects of various variables on the adoption and the frequency of access to Internet. Our estimation shows that on average the French have a surplus which varies between 2107 and 2651 dollars. This estimation was made under a strong hypothesis that needs to be relaxed. We suppose that leisure time is evaluated at the same rate as the working time and that the labor market is able to provide the necessary working hours. On the other hand, we suppose that the time spent on Internet is a leisure time but this is not always the case. The dividends in terms of business opportunities, transactions, work, training, better information for better actions are not considered here while they are the basic motivation of the Internet navigation nowadays. Thirdly, the estimation of the intensity of use is based upon the frequency of access. A better measurement of the intensity of usage allows us to have more realistic demand function and then consumer's surplus estimation.

REFERENCES

- ARCEP (Autorité de Régulation des Communications Electroniques et des Postes) (2005) ‘*Rapport d'activité de l'Autorité de Régulation des Communications Electroniques et Postales 2005*’.
- G.S. Becker (1964) ‘*Human Capital: A Theoretical and Empirical Analysis*’ with Special Reference to Education.. Chicago, University of Chicago Press.
- R. Blundell, P. Pashardes and G.Weber (1993) ‘What do we learn about Consumer Demand Patterns from Micro-Data?’ *The American Economic Review*, 83(3), 570-597.
- T. S. Bresnahan and R. J. Gordon (1997) ‘*The Economics of News Goods*’ The University of Chicago Press, Chicago, IL.
- E. Brynjolfsson, M. D. Smith and Y. J. Hu (2003) ‘Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers’ Working papers 4305-03, *Massachusetts Institute of Technology (MIT)*, Sloan School of Management.
- K. Chaudhuri, A. K. Flamm and J. Horigan (2005) ‘An analysis of the determinants of internet access’ *Telecommunications Policy*, 9(10), October.
- M. Gentzkow (2006) ‘Valuing New Goods in a Model with Complementarities: Online Newspapers’ *NBER Working Papers Series N° 12562*, October 2006.
- A. Goldfarb and J. Prince (2008) ‘Internet adoption and usage patterns are different: Implications for the digital divide’ *Information Economics and Policy* 20(1), 2-15.
- A. Goolsbee and P. J. Klenow (2006), ‘Valuing Consumer Products By the Time Spent Using Them: An Application to the Internet’ *American Economic Review*, 96(2), 108-113.
- A. Goolsbee and A. Petrin (2001) ‘The consumer gains from direct broadcast satellites and the competition with cable television’ *National Bureau of Economic Research Working paper*, W8317, Cambridge, MA.
- J. Greenwood and K. A. Kopecky (2007) ‘Measuring the Welfare Gain from Personal Computers’ *Economie d'Avant Garde Research Reports* 15.
- J. Hausman (1981) ‘Exact consumer surplus and deadweight loss’ *American Economic Review* 71, 662-676.
- J. Hausman (1997) ‘*Valuing of News Goods Under Perfect and Imperfect Competition*’ Bresnahan, T. and Gordon, R. (eds.), *The Economics of News Goods*, (The University of Chicago Press)
- J. Hausman (1999) ‘Cellular Telephone, New Products and the CPI’ *Journal of Business and Economic Statistics*, 17(2), 188-94.

J. Hausman (2002) 'Sources of Bias and Solutions to Bias in the CPI' *NBER working Paper*, 9298.

J. Hausman and G. Leonard (2002) 'The Competitive Effects of a New Product Introduction: A Case Study' *Journal of Industrial Economics*, 50.

J. Hausman and W. K. Newey (1995) 'Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss' *Econometrica*, 63(6), 1445-1476.

J. J. Heckman (1976) 'The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models' *Annals of Economic and Social Measurement*, 5, 475-492.

J. J. Heckman (1979) 'Sample Selection bias as a specification Error' *Econometrica*, 47, 153-161.

M.P. Hoffmann (1998) '*Developing sustainable management tactics for cucumber beetles in Cucurbits*' SARE (Sustainable Agriculture Research and Education Program) 1998 Annual Report.

IDATE (INSTITUT de l'AUDIOVISUEL et des TELECOMMUNICATIONS en EUROPE) (2005) '*Use-IT: qui consomme quoi en 2015?*'

N.L. Johnson and S. Kotz (1972) '*Distribution in statistics: continuous multivariate distributions*' Wiley, New York.

A. Nevo (2001) 'New Products, Quality Changes, and Welfare Measures Computed from Estimated Demand Systems' *NBER Working Paper* n° W8425.

A. Petrin (2002) 'Quantifying the Benefits of New Products: The Case of the Minivan' *Journal of Political Economy*, 110(4): 705-729.