A Double-Hurdle Model of Computer and Internet Use In American Households

Tadesse Biru Wodjao

Department of Economics
Western Michigan University

Correspondence:
Tadesse Wodajo
tadesse.b.wodajo@wmich.edu
Department of Economics
5314 Friedmann Hall
Western Michigan University
Kalamazoo, MI 49008

I would like to thank Dr. Jean Kimmel, Dr. Don Meyer and Dr. Kevin Hollenbeck for their insightful comments and suggestions. Any remaining errors are mine.

A Double-Hurdle Model of Computer and Internet Use in American Households

Abstract

This paper has two major contributions. First, it identifies the factors that influence the use of computer and Internet at home. Second, the paper shows that the double-hurdle model, compared to the standard *Tobit* and Heckman selection models, is the best econometric specification to identify the determinants of computer and Internet use at home. Computer/Internet use time is defined as the total minutes spent on three broad categories of nonmarket activities (household activities, leisure and volunteering). The study uses the 2006 American Time Use Survey (ATUS). The double-hurdle estimation reveals that the use of a home computer/Internet is governed by two independent decisions: the decision to acquire a home computer/Internet and the decision on the intensity of use. The estimation results also show that these two separate decisions are determined by different sets of factors. Most of the respondents' characteristics (age, education, disability and noncitizenship), household composition variables (number of children under different age groups), and economic variables (income and predicted wage) determine the minutes spent using computer/Internet at home. Only the location variables (urban, Northeast, Midwest and West) and two of the season variables (Spring and Winter) are the primary factors influencing the likelihood of owning a home computer/Internet.

Key words: Time allocation, household production, leisure, market work, Double-hurdle model

JEL Classification: DO1- Microeconomic behavior: underlying principle

D13 - Household production and intrahousehold allocation

C24 - Truncated and censored models

C52 - Model evaluation and selection

Background

A household performs various economically significant activities using different household technologies in which personal computers (PCs) and Internet have become a part in recent years. The adoption and use of these two technologies in American households have been steadily increasing since 1980s and 1990s, respectively. The home PC ownership grew from 8 percent in 1984 to 37 percent in 1994 and to 69 percent in 2003. The proportion of households with Internet access rose from 18 percent in 1997 to 62 percent in 2003 (Day et al (2005). Individuals use a home computer/Internet for more than just paid and unpaid work; they also use it for leisure. Examples for unpaid household activities include database management and household records keeping, whereas playing games, getting news and sports information are a few examples of leisure activities. Email or instant messaging is the other most popular use of the Internet at home.

Time allocation is best studied using time use data since these data provide detailed information on actual time use on a specific day, as opposed to the normal (or usual) time use data obtained from traditional cross-section surveys. This study employs the American Time Use Survey (ATUS) to exploit this advantage. The ATUS collects information on how people in the United States spend their time on market work and various nonmarket household activities. The ATUS also provides valuable information on respondent and household characteristics. It is a single-day time diary survey that has been conducted every year since 2003. This study draws time-use data from the 2006 survey on the actual minutes individuals spend using computer/Internet at home on three major categories of activities: (a) household activities, (b) socializing, relaxing and leisure, and (c) volunteer activities.

Time use surveys also have disadvantages as compared to other survey types. Time use data acquired on a single diary day contain too many zeros for many time use variables. For instance, in the 2006 ATUS data nearly 85 percent of the respondents report zero minutes for computer/Internet use at home, although we know from other data sources that there are computers in nearly 69 percent of American households.[1] These zero responses could arise from individuals' behavioral responses or the design of the survey. That is, the zero

[1] See Appendix A3-1.

responses could come from non-ownership of a home PC or from individuals who use PCs but did not use them on that single diary day. The traditional approach to deal with such zero-inflated data has been to use the standard *Tobit* model, originally formulated by Tobin (1958). However, this model is too restrictive as it assumes all the zeros to be the respondents' deliberate choices. Cragg's (1971) double-hurdle model overcomes this restrictive assumption. In this model, two hurdles must be crossed in order to report non-zero minutes of computer/Internet use. First, one decides whether to own a home computer/Internet, and then how many minutes to spend using it, once owned. In the literature, the two hurdles are referred to as participation and level of participation (consumption) decisions, respectively. The double-hurdle model is used in this research.

This research has two major contributions. The first contribution is that it provides information regarding the use of computer/Internet at home. In this vein, the paper identifies different sets of factors that determine the decisions to own and the intensity of use of a home computer/Internet. The second contribution of the study relates to the choice of an estimation technique. To show that the double-hurdle model is indeed superior to other most commonly used binary dependent variable models, the double-hurdle model is tested against the *Tobit* and Heckman models using likelihood ratio (LR) and Vuong tests, respectively. The tests reveal that, compared to these two models, the double-hurdle model is the best econometric specification to deal with the single-day diary data used in this study. This implies that the allocation of time for home computer/Internet use follows two independent decision paths: the decision to acquire a home computer/Internet, and the decision on the intensity of use. Therefore, the double-hurdle model is the best specification to identify the factors that influence these decisions at both stages. This finding has implications for other research using time diary data.

The rest of the paper is structured as follows. Section 1 gives a brief review of the computer and Internet use in the United States households. This is followed by a description of the data and the sources. The third section reviews the underlying theoretical background and discusses on the derivation of the time allocation model. Estimation strategies of time use data and econometric specification issues are discussed in detail in the fourth section. The last two sections present the estimation results and the conclusions of the study.

1. Computer and Internet use in the United States households

This section gives a brief description of the diffusion and use of computer and Internet in the American households using the October Supplement of the 2003 Current Population Survey (CPS) data. In addition to the labor related information gathered each month by CPS, the October Supplements routinely gather data on education-related issues and on computer and Internet access and uses from relatively large number of respondents (Day et al, 2005).[2]

The surveys show that the trend of ownership of computer and Internet at home is constantly increasing over time in the United States. The proportion of households with a computer grew from nearly 8 percent in 1984 to about 69 percent in 2003. In contrast, access to the Internet from home rose from 18 percent in 1997 to nearly 62 percent in 2003 (Figure 1).

Among those who own computers, 68, 21 and 11 percent, respectively, have at least one, two or three computers or laptops at home. The majority (78%) of these PCs are relatively new because they were bought after the year 2000. American households also use such communication and entertainment media as telephone (82.6%), cell phone (55.0%), cable TV (54.5%) and satellite TV (16.1%) (Appendix A3-1). The majority (84%) of computer owners use their home PCs for various purposes: while 53 – 91% use them for connection to the Internet, for personal emails, to complete school assignment, for playing games and word processing, the others (31 – 39%) need their PCs to work from home, for graphics and design, database management and for household records keeping (Appendix A3-2).

In 2003, slightly less than two-thirds of American households connect to the Internet via Dial-up (i.e., the slow speed Internet connection), mainly because they either do not need high speed internet (42%) or find it expensive (39%) (Appendix A3). On the other hand, only 36 percent use broadband technologies (i.e., Cable Modem and Digital Subscriber Line,

[2]The Bureau of Labor Statistics (BLS), sponsored by the National Center for Education Statistics (NCES), collects the data. The computer use data were gathered for the years CPS October 1984, 1989, 1993, 1997, 1998, 2000, 2001 and 2003, and provide detailed information on the availability of computer at school, home, and work; reasons for and frequency of computer use at school, home, and work; availability and use of Internet at school, home, and work.

DSL). In all instances, more than half of the households access the Internet at least once in a day. Those who do not access the Internet from home (38%) use their work places, schools, other people's houses and public libraries to access the Internet. For those who are traveling, airports, hotels and Internet cafés serve as Internet access sites. However, the proportions of individuals who access the Internet at these places are relatively small.

The most popular uses of Internet at home include: email or instant messaging (42%), search for information about products and services (35%), get news, weather and sports information (31%), purchasing products and services (23%) and playing games (21%) (Appendix A3-4).

2. Data and Sources

The main source of data for this study is the 2006 American Time Use Survey (ATUS). The U.S. Census Bureau, sponsored by the Bureau of Labor Statistics, has conducted this annual survey since its inception in 2003. The ATUS collects time diary information from representative households on how people living in the United States spend their time on paid work and unpaid activities. It shows the different kinds of activities people are engaged in and the time they spend doing them, disaggregated by sex, age, educational attainment, labor force status, and other characteristics, as well as by weekday and weekend days (BLS, 2007). The ATUS, for example, provides information on the amount of time people spend in market work, childcare, adult care, housework, commuting, sleeping, volunteering, religious activities, socializing, exercising, using computers/Internet at home, and relaxing. The ATUS also collects information about where and with whom each activity is conducted, and whether the activities are performed for one's job or business.

The ATUS collects information regarding computer and Internet use as a by-product of time use categories related to other activities. For example, one activity included within the broader umbrella activity of households production is computer and Internet use. In total, computer/Internet use appears three times in the ATUS' complete activity Lexicon. These various instances of computer use are grouped into three categories: (a) household activities, (b) socializing, relaxing and leisure, and (c) volunteer activities. The first one reports the total minutes spent in using computer/Internet for household and personal e-mail and messages.

The second category comprises computer/Internet uses for personal interest, excluding games[3]. For ease of presentation, the term leisure is used throughout the paper for this group of activities. The third classification refers to the minutes reported in using computer/Internet for administrative and support activities related to volunteering.[4]

The ATUS is a random subsample of respondents that have competed their final month of interviews for the Current Population Survey (CPS). Only one individual, who is at least 15 years of age, is chosen from each selected household and interviewed only one time about his or her time use on the previous day (BLS, 2007).

The 2006 ATUS consists of 12,943 households. However, the number of respondents that reported computer/Internet use at home is much less than this figure. Accordingly, those who reported the use of computer/Internet for the three time use variables of interest, namely, household, leisure and volunteer activities are only 912, 1193 and 111, respectively. In this paper, to represent all the minutes spent in using computer/Internet by a single summary variable, a fourth variable (known as computer use time) is defined as the sum of the minutes spent in the above three time uses. In terms of this aggregate measure, the total number of individuals reporting the use of computer/Internet at home becomes 1,954. This indicates that only 15 percent of the survey participants reported computer/Internet use at home on the diary day. The corresponding figure for 2003 was 12 percent. These contrast with the approximately 58 and 47 percent of U.S. households who reported in 2003 not only owning but also using their computers for home activities and to connect to the Internet , respectively.[5] The ATUS reliance on a single diary day is one reason for such small computer/Internet use response rates. Designating many computer/Internet using activities

[3] Time spent playing games on the computer and over the Internet are included in the general category of "playing games" (rather than computer/Internet use) along with other board and card games and puzzles (lexicon 120307). Thus, minutes spent playing games on the computer/Internet cannot be disentangled from other game playing. Given the wide use of computers and Internet for playing games, this would definitely understate the minutes reported for the activities categorized under "Socializing, Relaxing, and Leisure" (lexicon 120308).

[4] For specific examples on the activities given in (a), (b) and (c), refer to Appendix A1, lexicons 020904, 120308 and 150101, respectively.

[5] The data gathered on computer and Internet use by the October supplement of the 2003 CPS indicates that out of the total 69.2% households who reported owning home computers, the proportion of those who actually use their computer and connect to the Internet from home are 83.7% and 61.5%, respectively (Appendix A4-1 & 2).

under other activity categories is the other major survey design problem in ATUS contributing to the generation of zero responses.[6]

3. Underlying Theoretical Background

A household engages in a variety of economically significant market and nonmarket activities. Examples for nonmarket activities include food preparation, raising children and engagement in leisure. Technological advances make available for consumers ever-more complex gadgets to use in non-market work and play. In addition, the rapid progress in microelectronics technology has facilitated the ownership and use of less costly personal computers and Internet at home for many households in the United States in the last 20 years (NSF, 2001). Working or playing with these gadgets takes time, time that might alternatively be devoted to a different activity. For example, the time devoted to a home computer/Internet use is time not devoted to other activities. As a result, individuals face the problem of allocating their leisure time to competing nonmarket activities.

3.1 **The time allocation model**

The time allocation decision of an individual can be viewed in terms of the structure of time. The finite nature of time requires individuals to make choices among various activities based on their perceived relative utilities. These choices can then be classified into groups. Traditionally, time has two discontinuous structures consisting of paid work and leisure, where leisure is typically defined residually. Unlike paid work, leisure contains a number of activities that cannot be easily defined in operational terms useful for analysis (Feldman and Hornik, 1981).

Traditional economic theory deals with the labor/leisure choice by treating leisure as a part of a utility function, where utility is assumed to depend only on the consumption of a composite good and leisure time (Gronau, 1980). Blundell and Meghir (1986) extended Gronau's model by adding a set of taste shifter observable factors in the utility function. Kooreman and Kapteyn (1987) further disaggregated the residual leisure time into a multiple

[6]Appendix A2 identifies from the ATUS Lexicon those activities that could either be partly performed using computer/Internet or are already identified as computer/Internet using activities but reported under different categories.

of unpaid activities undertaken by individuals. Based on these modifications, a utility function can be written as:[7]

$$U = f(X,\ T_i\ ;S) \qquad \text{for } \ i = 1, 2, ..., n \qquad (1)$$

This is a one-person, one-period model where X is consumption of a composite good, T_i is the time spent on the i_{th} unpaid activity, and $S = \{H, R, F\}$ denoting a vector of individual characteristics (R), household characteristics (H) and other factors (F), such as geographic location and the diary day that are assumed to influence the individual's time preferences (Kimmel and Connelly, 2007).[8]

An individual maximizes this utility function subject to the following two interrelated constraints, known as budget[9] and time constraints, respectively:

$$\sum_{k=1}^{m} X_k = A + w(T - \sum_{i=1}^{n} T_i) \quad \text{and} \quad T = T_m + \sum_{i=1}^{n} T_i \qquad (2)$$

where X_k is the k_{th} consumption good, A is nonlabor income, w is the fixed market wage rate, T_m is market work time, and T is total time available. Maximizing (1) with respect to the constraints stated under (2) yields the optimal demand for consumption goods and for the various nonlabor time allocations included in the model:

$$X_k^* = f_k(w, A; S) \qquad \text{and} \qquad T_i^* = f_i(w, A; S) \qquad \text{for } \ i = 1, 2, ..., n \qquad (3)$$

This shows that the individual decides not only on the utility-maximizing level of consumption goods but also determines how unpaid time T_i is used. Equation (3) also indicates that the optimal allocation of consumer goods and time for various activities depend

---

[7] As opposed to Gronau (1980), Kooreman and Kapteyn (1987) use a household with both male and female partners and subdivide leisure into the time spent by each partner on a number of activities. This study employs a single-person household model because ATUS collects information from a single respondent per household only.

[8] NSF (2001) and Day et al (2005) also identify socioeconomic characteristics, demographic variables and family structure as important factors influencing the use of information technology (i.e., computer and Internet) at home.

[9] The budget constraint is derived by rearranging the expenditure equation $\sum_{k=1}^{m} P_k X_k + \sum_{i=1}^{n} wT_i \le M \equiv A + wT$, which stipulates that total expenditure (M) on goods and leisure is at most equal to the sum of the labor and nonlabor income. Price is normalized to 1 in this study. Because there is no price information in the ATUS, the survey can be treated as a cross-section and assume all respondents face the same price (Yen and Jensen, 1996).

on the price of time w and the set of taste shifter variables S. Denoting all the factors that affect these optimal allocations by Z, equation (3) can be rewritten more compactly as: $X_k^* = f_k(Z)$ and $T_i^* = f_i(Z)$. The time allocations this paper focuses on are the three earlier-defined computer/Internet activities: household, leisure and volunteer activities.

4. Estimation Strategies

4.1 Identification of estimation variables

Let the optimal leisure time in (3) T_i^* be denoted by a vector specifying the allocation of time into n number of nonmarket activities, such as food preparation, household management, childcare, personal care, maintenance and repair, computer/Internet use, socializing and relaxing, and shopping: $T_i^* = (T_1, T_2, ..., T_j, ..., T_n)$. And let the time allocated for the use of computer/Internet at home be T_j. To estimate the impact of the explanatory variables described in equation (3) on the optimal allocations of time for computer/Internet use at home, the estimation version of the time demand equation (3) can be written as:

$$T_j = \beta_0 + \beta_j^{/} Z + \varepsilon_j \qquad \text{for} \quad j = 1, 2, \ldots, m \qquad (4)$$

Estimation is carried out for computer use time by aggregating the three distinct categories of home computer/Internet use activities (household leisure and volunteer activities) over the whole sample m. Here Z denotes the vector of explanatory variables identified in equation (3) as potential determinants of the amount of time an individual allocates for the use of computer/Internet at home. These explanatory variables can be categorized into the following broad groups: economic variables, respondent characteristics, household characteristics and location variables.

4.2 Descriptive statistics of estimation variables

Descriptive statistics for variables used in estimation are presented in Table 1. As can be seen, approximately the same numbers of observations are drawn from the weekends and

weekdays samples (6,457 and 6,486, respectively). In addition, no notable difference is observed in the mean values reported under these two groups.

Respondent characteristics: On average, the respondents are 46 years old with nearly 13 years of education. Four percent of the respondents report a work-limiting disability. The majority of the survey participants are females (57%). The nonwhite population constitutes nearly 18 percent of the sample. Non-citizens account for about 8 percent of the sample.

Household characteristics: Fifty-three percent of the respondents are living with their spouses or unmarried partner. While 50 percent of the households have children under the age of 9, only 47 percent report having children aged 10 to 17. On average, there is one other adult, other than the spouse, in the household and the average family size is about 3.

Location characteristics: The majority of the respondents (82%) are living in urban areas. Respondents seem to be over-sampled from the south compared to other geographic regions due to survey design. However, here seems to be no variation in season sampling.

Economic characteristics: Of the total observations considered for analysis, 64 percent are employed and earn an average hourly wage of about $16. More than 70 percent of the respondents' spouses or unmarried partners have jobs. Nearly half of the respondents work in management, financial, professional, service and sales related occupations and in office and administrative support areas. The average annual family income for the majority (60%) of the sample households is less than $50,000.

Distribution of minutes of computer/Internet use

The total minutes spent on computer/Internet use is the dependent variable in this study. To examine the variations in the distribution of time spent on various activities, the total minutes of computer/Internet use is disaggregated based on some selected attributes believed relevant to time use decisions (such as days of the week, gender, marital, parental and employment status and level of hourly wage). Accordingly, minutes spent using the

computer/Internet for both household activities and leisure are observed to be higher on weekends than weekdays, lower for females than males and higher for the unemployed than employed individuals (Table 2b). In addition, the most common use for a computer/Internet is for leisure.

The rate of computer/Internet use shows some variations when viewed in terms of age of children, marital status, and days of the week (Table 3). For married respondents living with their spouses, the average minutes spent on computer/Internet use seem to be correlated negatively with children's age when used during the weekends, and a positive association when used in the weekdays. These opposite effects come from less use of a computer/Internet for household activities during the weekends and more for leisure activities during the weekdays. In contrast, single parent respondents with older children seem to spend on average more minutes on computer/Internet use during both weekdays and weekends.

Finally, to examine the relationship between wage and computer/Internet usage, the minutes spent in using computer/Internet are cross-tabulated against three wage levels: low, medium and high predicted wages[10] (Table 4). Assuming a normal distribution, a medium wage can be defined as the mean predicted wage plus/minus one standard deviation.[11] The association between predicted wage and computer use seems to vary with marital status. For married respondents, the average minutes of computer use tends to be related negatively to the wage during the week and positively on the weekend. However, the opposite relationship exists for unmarried respondents. On the other hand, irrespective of the day of the week, unmarried respondents on average spend more minutes on computer/Internet use compared to married respondents.

Differences between computer/Internet users and nonusers:

As a starting point, a simple comparison of means is carried out in order to examine whether the use of the computer or Internet produces any variation on individuals' allocation

[10] The mean predicted wage considered in this case is the one computed for all days of the week. For the details of the computation of the predicted wage, see section 5.1.

[11] Since the mean predicted wage (in logs) is 2.76 (or $15.84 per hour) with standard deviation of 0.405, the medium predicted wage will be in the range of 2.36 to 3.17 (or $10.59 to $23.81 per hour). Thus, the predicted wage below 2.36 becomes low and that above 3.17 is considered as high.

of time on both paid work and nonlabor activities. This comparison is based on the assumption that those who reported zero minutes for home computer/Internet use are nonusers of this technology. For this purpose, two groups of activities are selected: activities that can be performed with or without using a computer/Internet[12] and those not directly related to computer applications. Government services, financial and banking services, shopping and job search are some examples for the first group of activities. Examples for the second group include the travel time associated with the above activities, physical exercises and leisure (excluding computer games). The comparisons of the average minutes spent by each group on the selected activities are presented in Table 5.

The comparisons reveal two important findings. First, the individuals who reported zero minutes of home computer/Internet use are spending significantly more minutes on the majority of the services and activities as compared to their counterparts. Second, the number of reporting cases for all the services and activities is much higher for this same group than for home computer/Internet users. In addition, both groups are spending statistically the same amount of minutes on a few of the activities only: financial and banking services, purchase research, travel for government services, and travel for job search and interview. The significant differences observed in the average minutes spent on the selected activities may imply that the use of computer/Internet at home could probably be one potential reason for the differences in the time allocations of the two groups of respondents. Those who reported the use of home computer/Internet seem to save some minutes from the activities and services on which their counterparts are spending on average more minutes.

Similarly, to see if there is any association between home computer/Internet use and hours worked, the average minutes reported by the two groups of respondents on paid work are compared. Hence, those who use a computer/Internet at home are observed to work less by about 1 hour as compared to their counterparts. The difference is also statistically significant. In addition, noticeable differences are observed in the average minutes allocated for non-computer using leisure activities and physical exercises. Although these two

[12]Computer/Internet related activities are identified based on the various possible uses of computer/Internet reported by American households in the 2003 CPS October supplement (Appendix A3).

activities are remotely related to computer use, both groups of respondents seem to allocate different amount of time to these activities.

Finally, two tentative conclusions can be drawn from this simple comparisons of means. First, the individuals who report non-zero minutes of computer/Internet use at home seem to show different time allocation behavior as compared to those who report zero computer/Internet use minutes. Secondly, the observed time allocation differences are not limited to the activities that can be performed using a home computer/Internet. The differences also extend to the services and activities that are not directly related to computer/Internet use. Why such differences are observed between the two comparison groups is not evident at this point. This study further investigates the factors that influence the allocation of time for home computer/Internet use.

4.3 Estimation methods for time use data

Time use data have peculiar characteristics that require special consideration when using them in regression analysis. These characteristics may arise from the respondents' behavioral responses or the design of the time use survey. These unique characteristics may result in too many zeros reported for the various activities included in the time use surveys (Flood and Gråsjö, 1998; Schwierz, 2003). The same problem is also reported in labor supply and consumer expenditure surveys (Flood and Gråsjö, 1998). Specifically, because the ATUS collects just a single day's activities, many activities are likely to be reported by relatively few individuals in the survey although a far larger percent of the sample engages in the activities occasionally. In addition to computer use, the other activities likely to suffer from this single diary day problem include shopping and volunteering.[13]

For instance, in the 2006 ATUS data, the majority (84.9%) of the respondents report zero minutes for computer/Internet use at home on the diary day although, as mentioned earlier, approximately 69 percent report computer ownership (Appendix A3-1). The low usage can be explained in two ways. First, the individuals do not have a computer at home or the individuals own a computer but did not use it, for some reason, on the diary day. The zero

[13] In the 2006 ATUS, only 43.7 and 7.2% of the respondents report shopping and volunteering, respectively. If shopping disaggregated, 26.3 and 27.7% of the sample reported grocery and nongrocery shopping on the diary day.

values in the former case are related to the respondents' computer ownership decisions and are called behavioral zeros, while those in the latter case are termed as random zeros as they arise from random events.[14] Second, the design of the time use survey can also contribute to the generation of zero values due to the fact that the same time use questions are posed to all of the respondents without first asking whether they have a computer or not.

The traditional approach to deal with data that have too many zeros, yielding a censored dependent variable, has been to use the standard *Tobit* model, originally formulated by Tobin (1958). The model permits incorporation of all observations including those censored at zero, without considering the sources of the zeros. As this ignores the zero observations due to respondents' non-participation decisions, applying the *Tobit* model imposes the assumption that all the zeros arise from other factors alone (such as economic and demographic characteristics of the respondents) (Newman et al, 2003; Martínez-Espiñeira, 2006).

Heckman (1979) proposes a model that addresses the problem associated with the zero observations generated by non-participation decisions, arguing that an estimation on a selected subsample (i.e., censored estimation) results in sample selection bias. The model overcomes this problem by undertaking a two-step estimation procedure (known as *heckit*). In this estimation, a full sample *Probit* estimation is followed by a censored estimation carried out on the selected subsample. While the first estimates the probability of observing a positive outcome (known as the selection or participation equation), the second estimates the level of participation conditional on observing positive values (known as the conditional equation) (Dow and Norton, 2003). The model assumes that different sets of variables could be used in the two-step estimations. As opposed to the *Tobit* model, Heckman's (1979) model considers the zero observations to arise mainly from respondents' self-selection. In other words, this means that all the zeros come from the respondents' deliberate choices.

The *heckit* model differs from the *Tobit* model in two ways. First, the *heckit* recognizes the process to be a two-stage decision, and second it permits the use of different

[14] Carlin and Flood (1997) attribute the presence of too many zeros in the data either to censoring (behavioral or true zeros), or to faulty reporting, or other random effects (random zeros).

sets of explanatory variables in both stages of estimations. Consequently, the *heckit* can be viewed as a generalized version of the *Tobit* model.

Cragg (1971) modifies the *Tobit* model to overcome the restrictive assumption inherent in it, namely, he suggests the "double-hurdle" model to tackle the problem of too many zeros in the survey data by giving special treatment to the participation decision. The model assumes two hurdles to be overcome to observe positive values. Stated in terms of acquisition of durable goods, first, one has to desire a positive amount, and second, there have to be favorable circumstances to realize this positive expenditure. In terms of home computer/Internet use, this can be interpreted as follows. A non-zero home computer/Internet time can be observed if, first a decision whether to acquire a home computer or be connected to the Internet is made (the first-hurdle), and second random circumstances permit usage on the diary day, once it is acquired (the second-hurdle). In general, the first-hurdle refers to the participation or ownership decision and the second to the level or intensity of use.

The *heckit* and the double-hurdle models are similar in identifying the rules governing the discrete (zero or positive) outcomes. Both models recognize that these outcomes are determined by the selection and level of use decisions. They also permit the possibility of estimating the first- and second-stage equations using different sets of explanatory variables. However, the *heckit*, as opposed to the double-hurdle, assumes that there will be no zero observations in the second stage once the first-stage selection is passed. In contrast, the double-hurdle considers the possibility of zero realizations (outcomes) in the second-hurdle arising from the individuals' deliberate choices or random circumstances. This is the main difference between the two models.

The difference between the two models can be best illustrated using the following example on computer use. According to the *heckit* model, only non-computer owning respondents can report zero minutes of computer use. The model further assumes that individuals owning a home computer do not report zero values at all. On the other hand, the double-hurdle model assumes that zero values can be reported in both decision stages. The zeros reported in the first-stage arise from non-ownership and those in the second stage come from non-computer use due to the respondents' deliberate decisions or random

circumstances. In this regard, the double-hurdle model can be considered as an improvement both on the standard *Tobit* and generalized *Tobit* (*heckit*) models.[15]

Some studies use the Heckman model in the analysis of consumption decisions as a two-step process. Examples include: Cheng and Capps (1988), Heien and Wessells (1990), Heien and Durham (1991) and Saha et al (1997). On the other hand, the double-hurdle model has been widely applied in household consumption and labor supply decisions. The studies that used the double-hurdle model in consumer demand models include Deaton and Irish (1984) on household expenditures, Jones (1989, 1992) on tobacco expenditure, Burton et al (1996) on household meat expenditure in U.K., Jensen and Yen (1996) on U.S.'s food expenditure away from home, Yen and Jones (1997) on U.S.'s household cheese consumption, Newman et al (2003) on Irish household expenditure on prepared food, Fabiosa (2006) on wheat consumption in Indonesia, and Aristei et al (2007) on alcohol consumption in Italian households. Examples of uses of the double-hurdle models in studies of labor supply are Blundell and Meghir (1987), Blundell et al (1987, 1998), and Carlin and Flood (1997). The double-hurdle model is also applied in other field of studies: in models of soil conservation (Gebremedhin and Swinton, 2002), in loan default analysis (Moffatt, 2005), in the examination of charitable giving in willingness to pay studies (Verdin-Johansson, 1999), and in effects of volunteering on social capital formation (Isham et al, 2006).

This study employs the double-hurdle model with the assumption that the computer/Internet ownership and intensity of use are two distinct decisions. In addition, different sets of factors are believed to influence the decisions at each level. It appears that the double-hurdle model is not extensively used in studies that employ time diary data.[16] In contrast, the standard *Tobit* is the most favored estimation method in time use studies. The following are examples for the studies that employ *Tobit* estimations on time diary data. Kalenkoski et al (2005) investigate how parents' time spent in child care is affected by marital status and other demographic characteristics; Kimmel and Connelly (2007) examine whether mothers' time spent with their children is a production or leisure time; and Sayer et al (2004) analyze factors influencing mothers' and fathers' time investment in their children.

[15] Also known as *Tobit* type I and *Tobit* type II models, respectively (Flood and Gråsjö, 1998, 2001).
[16] Most of the studies cited above use survey data rather than time diary data.

4.4 Econometric specifications

The purpose of this paper is to investigate the factors that determine the minutes spent in using computer/Internet using the double-hurdle estimation technique. As a robustness check, the estimated parameters are compared to the corresponding standard and generalized *Tobit* estimations. Furthermore, likelihood ratio (LR) and Vuong tests, respectively, are conducted to check whether the double-hurdle estimation is indeed superior to the standard and generalized *Tobit* models. This section presents the econometric models that can be used to estimate the minutes spent in home computer/Internet using the above three estimation techniques.

4.4.1 The standard *Tobit*

The standard *Tobit* specification is defined as

$$t_i^* = X_i' \beta + \varepsilon_i \qquad with \quad \varepsilon_i \sim N(0, \sigma^2) \quad and \quad i = 1, \ldots, n \qquad (5a)$$

$$t_i = \begin{cases} t_i^* & if \quad t_i^* > 0 \\ 0 & if \quad t_i^* \leq 0 \end{cases} \qquad (5b)$$

where t_i^* is a latent endogenous variable[17] representing individual $i's$ desired level of minutes spent in using computer/Internet, and t_i is the corresponding actual observed number of minutes. X_i is a set of individual characteristics that explain both ownership and level of computer/Internet use, and β is a corresponding vector of parameters to be estimated. In this model, ε_i is assumed a homoskedastic, normally distributed error term. Equation (5b) states that the observed number of minutes become positive continuous values if only positive number of minutes are desired, but zero otherwise. Note that since there is no negative number of minutes, the censoring could be placed at zero without any loss of generality. This shows that the observed 0's on t_i can mean either a "true" 0 (i.e., due to the individual's deliberate choice) or censored 0 (i.e., those caused by data collection method).

The standard *Tobit* model is estimated using maximum likelihood methods. The log likelihood function for this estimation is:

$$LL = \sum_{0} ln\left[1 - \phi\left(\frac{X_i'\beta}{\sigma}\right)\right] + \sum_{+} ln\left[\frac{1}{\sigma}\varphi\left(\frac{t_i - X_i'\beta}{\sigma}\right)\right] \qquad (6)$$

[17]Heckman (1979) defined the latent variable as a variable that may or may not be directly observable.

where "0" indicates summation over the zero observations in the sample ($t_i = 0$) and "+" indicates summation over positive observations ($t_i > 0$). $\Phi(\cdot)$ and $\varphi(\cdot)$ are the standard normal distribution and density functions (*cdf* and *pdf),* respectively.

4.4.2 The generalized *Tobit (Heckit)*

As discussed before, to overcome the sample selection bias arising from estimations carried out using the observed variable in the sample, Heckman (1979) proposed a two-step estimation method. The first step refers to the participation (or computer ownership) decision and the second to the level of usage decision. Based on these specifications, the standard *Tobit* can be modified following Heckman (1979), and Flood and Gråsjö (1998, 2001) as:

(a) Ownership decision:

Index equation $$d_i^* = X_{1i}^{'}\beta_1 + u_i, \quad u_i \sim N(0,1) \tag{7a}$$

Threshold index equation: $$d_i = \begin{cases} 1 & if \quad d_i^* > 0 \\ 0 & if \quad d_i^* \leq 0 \end{cases} \tag{7b}$$

(b) Level of usage decision:

Computer time equation $$t_i^* = X_{2i}^{'}\beta_2 + v_i, \quad v_i \sim N(0, \sigma^2) \tag{7c}$$

Threshold computer time equation: $$t_i = \begin{cases} t_i^* & if \quad d_i = 1 \\ 0 & if \quad d_i = 0 \end{cases} \tag{7d}$$

In this specification, separate sets of factors are assumed to influence the decisions to own a home computer/Internet versus the actual minutes spent in using it, once it is owned. Hence, X_{1i} and X_{2i} are vectors of explanatory variables that affect these two-stage decisions, respectively. Both variables are also assumed to be uncorrelated with their respective error terms u_i and v_i. β_1 and β_2 are the corresponding vectors of parameters. While d_i^* is a latent index variable that denotes binary censoring, d_i is the observed value representing the individual's participation decision (i.e., if 1 it means the respondent is reporting a positive number of minutes($d_i^* > 0$), else 0). Hence, the actual observed number of minutes t_i equals the unobserved latent value t_i^* only when a positive number of minutes is reported; otherwise, it takes the value 0 (equation 7d). In this specification, the error terms are assumed

to be normally and independently distributed, implying that there is no dependence between the ownership and level of use decisions (i.e., the two decisions are made independently).

However, the Heckman (1979) model is derived based on the assumption that the error terms are correlated and the first stage decision dominates the second one. In this case, the error terms are assumed to be correlated and have a bivariate normal distribution:

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}\right] \qquad (7e)$$

where ρ is the correlation coefficient. If computer/Internet ownership dominates the level of usage, those who own a home computer/Internet will necessarily use it. This means that all the zero minutes of computer/Internet use come from non-ownership only. The implication of this is that once the first hurdle is passed, censoring is no longer relevant, as those who own computer do not report zero minutes of computer/Internet use. In such a case, the model can be estimated using *Probit* for the ownership decision and standard OLS for the non-zero computer/Internet use.

The log-likelihood function[18] for this specification is (Flood and Gråsjö[19], 1998; Aristei et al, 2007):

$$LL = \sum_0 ln[(1 - \phi(X'_{1i}\beta_1)] + \sum_+ ln\left[\Phi\left(\frac{X'_{1i}\beta_1 + \frac{\rho}{\sigma}(t_i - X'_{2i}\beta_2)}{\sqrt{1-\rho^2}}\right)\frac{1}{\sigma}\varphi\left(\frac{t_i - X'_{2i}\beta_2}{\sigma}\right)\right] \qquad (8)$$

Notice that if the error terms in (7a) and (7c) are assumed to be independent (i.e., ρ = 0), the above specification can be further simplified.

4.4.3 The double-hurdle model

In this specification, an individual has to overcome two hurdles in order to report a positive number of minutes for home computer/Internet use. The first hurdle relates to whether or not the individual owns a home computer or has access to the Internet, and the

[18] Instead of using maximum likelihood estimation, Heckman (1979) suggests a two-step method (known as *heckit*). This requires estimating first the indicator equation (7a) using a *Probit* model, and then computing the inverse Mills ratio based on the coefficient estimates ($\lambda_i(-X'_{1i}\widetilde{\beta_1}) = \varphi(X'_{1i}\widetilde{\beta_1})/\phi(X'_{1i}\widetilde{\beta_1})$). Finally, the computer time equation (7c) can be estimated using λ_i as an additional right-hand-side variable.

[19] The authors also suggest the software Limdep to estimate this log likelihood function.

second to the intensity of use by those who own a computer/Internet at home.[20] The indicator (ownership) and structural (computer time) equations of the double-hurdle model resemble those of the generalized *Tobit* model. Cragg (1971) formulates the double-hurdle model by modifying the standard *Tobit* model. For the case at hand, a slight modification of the threshold computer time equation (7d) gives the double-hurdle model:

$$t_i = \begin{cases} t_i^* & if \quad d_i = 1 \ and \ t_i^* > 0 \\ 0 & else \end{cases} \qquad (9)$$

This indicates that the observed number of minutes t_i is zero either when there is censoring at zero ($t_i^* \leq 0$) or if there is faulty reporting, or due to some random circumstance. Rewriting Equation (9) more elaborately can help show explicitly the processes involved in observing zero values (Jones, 1992):

$$\begin{aligned} t_i = t_i^* = X_{2i}^{'}\beta_2 + v_i \quad & \text{if} \quad X_{1i}^{'}\beta_1 + u_i > 0 \ \text{and} \quad X_{2i}^{'}\beta_2 + v_i > 0 \\ = 0 \qquad\qquad & \text{if} \quad X_{1i}^{'}\beta_1 + u_i > 0 \ \text{and} \quad X_{2i}^{'}\beta_2 + v_i \leq 0 \qquad (11^{/}) \\ & \text{or} \quad X_{1i}^{'}\beta_1 + u_i \leq 0 \ \text{and} \quad X_{2i}^{'}\beta_2 + v_i > 0 \\ & \text{or} \quad X_{1i}^{'}\beta_1 + u_i \leq 0 \ \text{and} \quad X_{2i}^{'}\beta_2 + v_i \leq 0 \end{aligned}$$

Hence, a positive minute of computer/Internet use is observed if only an individual owns a home computer/Internet and he/she uses it (the first condition). Non-zero values can also be observed if a non-owner of a PC reports positive minutes of usage (the third condition denoting a faulty report). On the other hand, a zero value is observed if an individual owns a PC but did not use it on the diary day (second equation), or he/she does not own a computer/Internet and hence does not report any positive minutes of usage (last equation).
Assuming the error terms in (7a) and (7c) are independent, the stochastic specification in (7e) can be written as:

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2 \end{pmatrix}\right] \qquad (7e^{/})$$

[20]Here the assumption is that individuals reporting positive minutes for a computer/Internet use at home are indirectly reporting ownership of a computer at home and access to the Internet from home.

The double-hurdle model with independent error terms can be estimated by the following log-likelihood function (Moffatt, 2005; Aristei et al, 2007):[21]

$$LL = \sum_{0} ln\left[1 - \phi(X'_{1i}\beta_1)\,\Phi\left(\frac{X'_{2i}\beta_2}{\sigma}\right)\right] + \sum_{+} ln\left[\phi(X'_{1i}\beta_1)\,\frac{1}{\sigma}\varphi\left(\frac{t_i - X'_{2i}\beta_2}{\sigma}\right)\right] \qquad (10)$$

The first term corresponds to the contribution of all the observations with an observed zero. It indicates that the zero observations are coming not only from the ownership decisions but also from the usage decisions. This contrasts with Heckman's (1979) model that assumes all the zeros are generated only by non-ownership decision. For independent error terms, equation (8) reduces to

$$LL = \sum_{0} ln[(1-\phi\,(X'_{1i}\beta_1)] + \sum_{+} ln\,\left[\Phi(\,X'_{1i}\beta_1)\frac{1}{\sigma}\varphi\left(\frac{t_i - X'_{2i}\beta_2}{\sigma}\right)\right] \qquad (8')$$

Comparing the first term of equation ($8'$) to that of equation (10) reveals that the additional term $\Phi\left(\frac{X'_{2i}\beta_2}{\sigma}\right)$ depicts the contribution of the double-hurdle model. This term captures the possibility of observing zero values in the second stage decision.

The second term in equation (10) accounts for the contribution of all the observations with non-zero minutes. The probability in the second term is the product of the conditional probability distribution and density function coming from the censoring rule and observing non-zero values, respectively (Fabiosa, 2006). For the case at hand, the former denotes the probability of passing the ownership hurdle, and the latter indicates the density of observing non-zero minutes of computer/Internet use.

Furthermore, under the assumption of independence between the two error terms, the log-likelihood function of the double-hurdle model is equivalent to the sum of the log-likelihoods of a truncated regression model and a univariate *Probit* model (McDowell, 2003; Martínez-Espiñeira, 2006; Aristei et al, 2007). Consequently, the log-likelihood function of the double-hurdle model can be maximized, without loss of information, by maximizing the two components separately: the *Probit* model (over all observations) followed by a truncated

[21]Jones (1992) was the first to derive the likelihood function.

regression on the non-zero observations (Jones, 1989; McDowell, 2003; Shrestha et al, 2006). [22] Many studies seem to choose this approach mainly because there is no statistical software to handle the estimation of the double-hurdle model.[23] However, this study estimates the log-likelihood function of the double-hurdle model using user-written programs in the Stata software.[24]

5. Model specification

This section presents the estimation results of the *Tobit*, Heckman and double-hurdle models.[25] First, the issue of equation specification and variable identification is discussed. Secondly, a log-likelihood ratio (LR) and Vuong tests are used to choose the appropriate model from these three specifications. Finally, the selected model is used to analyze the factors that influence the two-stage decisions relating to the ownership and use of a home computer/Internet.

5.1 Equation specification and identification

The selection of regressors for the *Tobit* model is straightforward. All variables that are assumed to influence the allocation of time for home computer/Internet use are included based on variables used in other time allocation studies. However, the choice of explanatory variables for the ownership and level of use equations of the Heckman and double-hurdle models is more complex. There is no clear theoretical guidance regarding equation

[22] The separability of the likelihood function, emanating from the independence assumption, permits the use of a combination of the commands incorporated in the Stata software to estimate the model: *truncreg* and *probit*. This is similar to what is demonstrated by McDowell (2003) regarding estimation of count models using *Stata*.

[23] Some researchers provide custom-built commands in *Limdep* and *Gauss* (Jones, 1992) and *Stata* (Moffatt, 2005; Fennema and Sinning, 2007).

[24] This study uses the double-hurdle model with independent, homoskedastic and normally distributed error terms. Extensions to the double-hurdle model permits introducing corrections for these error specifications (see Appendix A5).

[25] The *Tobit* model is incorporated in many computer packages so that it can be routinely estimated. Heckman's (1979) sample selection model can be estimated by maximum likelihood, but many studies use the *heckit* (the two-step estimation) instead, where a full sample *Probit* estimation is followed by least squares regression on the selected subsample (Greene, 2000). However, the double-hurdle estimation is not yet incorporated in the standard statistical software (Flood and Gråsjö, 1998; Schwierz, 2003). In this study, the three models are estimated by maximum likelihood method. For the double-hurdle model, user-written programs are used. These programs were written for Stata by Julian Fennema and are available at: http://www.sml.hw.ac.uk/somjaf/Stata/.

specifications for these two models. In most cases, the selection of explanatory variables appears to be somewhat arbitrary (Newman et al, 2003; Aristei et al, 2007). One approach is to include non-economic variables in the sample selection equation (Jones, 1992; Yen and Jensen, 1996; Newman et al, 2003; Yen, 2005; Moffatt, 2005; Aristei et al, 2007).

In this study, the choice of variables for the first and second stage equations is done through a lengthy selection procedure that involves trying out many different combinations of variables from the list of explanatory variables identified to be relevant in explaining the two-step decisions. Accordingly, a set of non-economic variables (i.e., demographic, household and location variables) are included in the selection equations of the Heckman and double-hurdle models as determinants of the decision to own a home computer/Internet. For the second-stage decision equations, economic and some of the non-economic variables are identified as determinants of the actual minutes spent in using computer/Internet at home. This is followed by an empirical method[26] of excluding those that are found to be insignificant (Jones, 1992; Yen and Jensen, 1996). Therefore, the estimation results reported in this paper are based on the final set of explanatory variables obtained after a series of estimations, starting from a specification that uses all the explanatory variables in both hurdles and gradually dropping the insignificant ones based on the LR tests.

Specifically, the first-stage decisions are identified with the following variables: age, education, disability, female dummy, race, citizenship, marital status, number of children and adults in the household, and geographic and location variables. In the second-stage equation, the following are included: age, disability, female dummy, season, income, predicted wage, and number of children and adults in the household. Since not all respondents report hourly wage, calculating the predicted wage is necessary.

Calculation of the predicted wage

Nearly 36% of the respondents in the 2006 ATUS dataset are either unemployed or not in the labor force.[27] Thus, no hourly wages are reported for this group of respondents. Excluding such observations from the regression analysis creates a sample selection problem,

[26] The exclusion of the insignificant variables is done based on LR tests.
[27] 4.08% are unemployed and 32.18% are not in the labor force.

resulting in biased parameter estimates. In addition, the non-zero reported hourly wages could also be measured with errors or considered endogenous. To correct for these problems, a sample-selection-corrected hourly wage is computed for all observations using Heckman's two-step consistent estimator (or *heckit correction*). The predicated wage is used for each individual in the sample. The variables included in the selection equation and outcome equation[28] (first and second step estimations) and the overall prediction results are reported in Appendix A4.

5.2 Estimation results

The estimation results presented in this section emphasize the two main goals of the study: learning about the various factors that influence the decisions to own and use computer/Internet at home, and selecting the best model for analyzing single-day time diary data.

The dependent variable in this paper is total minutes spent in using computer/Internet at home. This censored variable is modeled using three alternative specifications: the *Tobit*, Heckman and double-hurdle models. As discussed before, the double-hurdle model nests the others and is considered an improvement over these two specifications. Hence, the question at hand is whether the double-hurdle model is the most appropriate specification to analyze the determinants of time allocation for computer/Internet use. To this end, first, each of the three models are estimated and the results are compared. In order to account for the differences in the parameters, the maximum likelihood coefficients of the three models are summarized and reported in Appendix A6. Second, a series of tests are carried out to pick the best model. The test results are presented in Table 6.

Model selection tests

Comparison of the parameters across the three estimation techniques reveals that very few of the coefficients have conflicting (opposite) signs. There is also similarity in the significance level of the parameters across the models. In particular, the estimates in the

[28] The choice of the variables to be included in these two equations are made in such a way that the resulting mean predicted wage would be comparable to the observed mean log wage.

Tobit and Heckman specifications have almost the same sign and significance levels. In contrast, the double-hurdle model results seem to differ slightly from the other specifications in both signs and significance. What all the models share in common is that the majority of the coefficients are significant.

To identify the model that best identifies the determinants of home computer/Internet use, a series of model specification tests are carried out in the following sequence. First, the independent double-hurdle model is tested against the *Tobit* specification, and then the double-hurdle model is tested against the Heckman model. The results for these two model specification tests are presented in Table 6. The LR test of the double-hurdle model against the *Tobit* model strongly rejects the latter specification. This is an indication for the existence of two separate decision-making stages in which individuals make independent decisions regarding the ownership and computer/Internet usage at home.[29] In this case, the *Tobit* model is proved to be restrictive in a sense that it does not make any distinction between the two stages of decision making. The rejection of the *Tobit* model further shows that the observations of zero minutes for home computer/Internet use can no longer be considered as corner solutions (i.e., deliberate choices made by individuals).

When the double-hurdle model is tested against the Heckman selection model, the Vuong test rejects the latter[30]. The rejection of the Heckman selection model disproves the claim that the zero minutes observed in the data are the results of ownership decisions alone (i.e., reported by those who do not own a computer/Internet at home).

Finally, it can be concluded that both the standard *Tobit* and the Heckman specifications are inadequate to model the computer/Internet use behavior. The tests reveal that the allocation of time for home computer/Internet use follows two independent decision paths: the decision to acquire a home computer/Internet, and the decision on the extent of use. Therefore, the double-hurdle model is the best specification to identify the factors that influence these decisions at both stages.

---

[29]The conceptual framework Venkatesh et al (1985) propose for technology adoption (including computers) in the households supports this result. The authors argue that the technology adoption process has two steps: first, a household decides to acquire a technology based on its perceived needs, and second, once acquired, the household determines the amount of time to be spent and the tradeoffs to be established.

[30] The Vuong test (due to Vuong, 1989) is a test for hypothesis of model equivalence for nonnested models (Clarke, 2007).

Table 6: Specification tests

Model	Test value†	Decision
Standard Tobit vs independent double-hurdle	139.78 (23) [0.005]	Reject *Tobit*
Independent Double-hurdle vs. Heckman	8.04*	Reject Heckman

†The column reports the LR and Vuong test statistics in rows 1 and 2, respectively, the degrees of freedom of the χ^2 statistics (in round bracket), and the corresponding p-value (in squared bracket), respectively.

5.3 Discussion of results

The analyses of the estimated parameters depend on the maximum likelihood parameter estimates of the independent double-hurdle model,[31] as this is the model found to be the appropriate specification to identify the determinants of computer/Internet use at home using a single-day time diary data.

The double-hurdle model is estimated by maximizing the log-likelihood function (equation 10). The results are presented in Table 7. The coefficients in the first hurdle indicate how a given variable affects the likelihood (probability) to own a home computer/Internet. Those in the second hurdle denote how a variable influences the level of usage, given that a decision is made to own a home computer/Internet.

A general view of the results shows the following notable differences in the parameter estimates of the variables in the ownership and level of use equations. First, the majority of the variables appearing in both equations have conflicting effects in terms of both signs and level of significance. In particular, except the variables Female, Income below $50,000, Children aged 0 and 2 and 3 and 6, Midwest and West, the rest have conflicting signs in the ownership and level of use equations. Age, Disability, Urban and Spring are a few examples of coefficients that change signs. In addition, while only 6 out of 21 variables are significant in the ownership equation but not in the level of use equation, 9 out of 23 variables are significant in the level of use equation but not in the ownership equation. Secondly, while only 8 variables out of 21 are significant at 1 percent level or lower in explaining ownership, 15 out of 23 variables significantly explain the level of use. The large number of significant variables in the double-hurdle model may indicate that the variables

[31] Focus is now only on the sign and statistical significance of the coefficients of the double-hurdle model. The marginal effects of the parameter estimates will be available in future drafts of this paper. However, the signs of the marginal effects will remain the same as the signs of the coefficients.

under consideration best identify the level of use decisions than the computer/Internet ownership decisions.

Turning to the analysis of the parameter estimates, factors that significantly increase the probability of ownership and level of use are discussed below.

Respondent characteristics:

Most of the respondent characteristics determine the level of use of a home computer/Internet but not the probability of ownership. Given that a decision is made to own a computer/Internet, the factors that significantly reduce the level of use include age, being female and being a noncitizen. In contrast, education[32] and disability significantly increase the level of use. However, disability has an opposite effect when it comes to owning a computer/Internet. As compared to others, individuals with physical impairments are less likely to own a PC at home. The coefficient for 'Education squared' shows that the rate of computer use increases with the level of education, but at a decreasing rate.

On the other hand, being nonwhite significantly reduces the probability of ownership, but this variable does not affect the level of usage. This means that once a decision is made to own a home PC, no significant difference is observed in terms of usage by race. Females have the same chance of owning a computer/Internet as males, but they spend less time using it as compared to males. That is, being female negatively affects the rate of computer use at home.

Household characteristics:

All the household composition variables are found to be significant determinants of computer/Internet use but not ownership. None of these variables has any significant influence on the probability of owning a PC at home. Having children under the ages of 12 years is found to significantly reduce the amount of time to be spent on using computer/Internet at home. Children older than 13 years do not have any effect on the amount of time their parents' spend on computer/Internet use. Instead, they tend to enhance

[32]This variable cannot be used in the ownership equation since convergence could not be achieved when education is inserted in the first-hurdle. One possible explanation for the lack of convergence is using the same variables in both hurdles, which makes identification difficult.

the rate of use, though not at a statistically significant level. Moreover, neither the size of the household nor the presence of a spouse or unmarried partner in the household are significant factors in the ownership decision of a home computer/Internet.

Location and season characteristics:

The variables in this category are found to affect mainly the ownership decision rather than the usage. In particular, all the location variables significantly increase the probability of computer ownership. Living in the urban areas, Northeast, Midwest and West, as compared to living in nonmetropolitan areas and the south, raises one's probability of owning a home computer/Internet. However, none of these factors has any significant impact on the rate of use.

On the other hand, two of the season variables (Spring and Winter) significantly reduce the probability of computer/Internet ownership. But the dummy for Winter has also a positive association with computer/Internet use. This indicates that, once a decision is made to own a home PC, individuals tend to spend more time using computers in the Winter than the rest of the year. Finally, the parameter estimates show that individuals' rates of use of computer/Internet are higher during the weekdays than the weekends.

Economic characteristics:

As discussed before in connection with equation specification, economic variables are selected to identify the second hurdle only.[33] Earning an annual income below $50,000 significantly reduces the rate of computer/Internet use at home. Although the probability of owning computer/Internet tends to be negatively affected by low annual family income, this effect is not statistically significant. In addition, the predicted wage is related negatively to the allocation of time for computer/Internet use at home. Compared to low wage earners, medium and high wage earning individuals spend less time on home computer/Internet.

[33]Since there is no theoretical guidance as to which variable to include in each hurdle, an attempt was made to include a number of economic variables in both hurdles during estimation. However, the maximum likelihood estimation fails to converge with this broadened specification, particularly whenever these variables are included in the first hurdle (ownership equation), so excluded from the final estimation. The variables that survived the exclusion are 'Income below $50,000' for both hurdles and wage measures for the second hurdle.

6. Summary and Conclusions

The main aim of this paper is to identify the factors that determine both the probability of owning a home computer/Internet and those influencing the intensity of use. Hence, it is necessary to go beyond the typical binary dependent variable methods applied to cross-sectional and time diary surveys. To this end, this study uses the double-hurdle model to address the issues involved. In this regard, the research has two major contributions. First, it identifies the factors that influence the use of computer/Internet at home. Second, the paper shows that the double-hurdle model, as compared to the standard *Tobit* and Heckman selection models, is the best econometric specification to identify these factors.

The double-hurdle estimation is carried out on the total minutes spent on two broad categories of computer/Internet use (household activities and leisure) using the 2006 ATUS data. The results show that the use of a home computer/Internet follows two independent decisions: the decision to acquire it and the decision concerning intensity of use. The estimation results also reveal that the level of computer/Internet use is mainly determined by respondents' characteristics (age, education, disability and noncitizenship), household composition variables (number of children under different age groups), and economic variables (income and predicted wage). Particularly, low annual income and medium to high-predicted wage are negatively associated with the time spent at home on using computer.

On the other hand, the location variables are the primary determinants of the probability of computer/Internet ownership. A few of the season variables (spring and winter) and respondent variables (disability and being nonwhite) are also observed to influence the likelihood of owning a computer/Internet at home.

In general, the study shows that the majority of the variables under investigation affect the use level of usage rather than the probability of computer/Internet ownership. This implies that the time allocation decision is central in the use of this technology at home. Based on this observation, it may be possible to draw a preliminarily conclusion that the ownership decision is probably dominated by the level of use decision. This calls for a further study that addresses the issue of dependence (with computer usage dominating the

ownership decisions), relaxing the assumption of "independence" used in this study. In addition, the following are the possible extensions for future studies:

(1) Increasing the sample size by using more than one year of ATUS data;
(2) Computing the marginal effects of the double-hurdle model;
(3) Work on error correction specifications.

In conclusion, the double-hurdle model is perhaps the most important contribution as it carries implications for many other time use research applications, which involve activities with too many zeros in the data, most like computer/Internet use, and commonly estimated by the *Tobit* model. Examples for such activities include volunteering and shopping. This study uses the ATUS time diary data to investigate computer/Internet use at home. For a more comprehensive study of computer use both at home and outside the home (such as working place, schools, public libraries and Internet cafés), future studies might need to use other nationally representative data sources in conjunction with the ATUS data.

References

Amemiya, T. and J. Powell (1981) "A Comparison of the Box-Cox Maximum Likelihood Estimator and the Non-linear Two Stage Least Squares Estimator," *Journal of Econometrics*, 17 (354-381).

Arabmazar, A. and P. Schmidt (1981) "Further Evidence on the Robustness of the Tobit Estimator to Heterscedasticity," *Journal of Econometrics*, 17 (253-258).

Arabmazar, A. and P. Schmidt (1982) "An Investigation of the Robustness of the Tobit Estimator to Non-normality," *Econometria*, 50:4 (1055-1063).

Aristei, D., F. Perali, and L. Pieroni (2007) "Cohort, Age and Time Effects in Alcohol Consumption by Italian Households: A Double-hurdle Approach," *Empirical Economics*, DOI 10.1007/s00181-007-0142-5.

Blundell, R. and C. Meghir (1986) "Selection Criteria for a Microeconometric Model of Labor Supply", *Journal of Applied Econometrics*, 1:1 (55-80).

Blundell, R. and C. Meghir (1987) "Bivariate Alternatives to the Univariate Tobit Model," *Journal of Applied Econometrics*, 1:34 (179-200).

Blundell, R., J. Ham, and C. Meghir (1987) "Unemployment and Female Labor Supply," *The Economic Journal*, 97, Supplement: Conference Papers (44-64).

Blundell, R., J. Ham, and C. Meghir (1998) "Unemployment, Discouraged Workers and Female Labor Supply," Research in Economics, 52:2 (103-131)

Bureau of Labor Statistics (2007) American Time Use Survey User's Guide, Understanding ATUS 2003 to 2006.

Burton, M., R. Dorsett, and T. Young (1996) "Changing Preferences for Meat: Evidence from UK Household Data, 1973-1993," *European Review of Agricultural Economics*, 23 (357-370).

Carlin, P. and L. Flood (1997) "Do Children Affect the Labor Supply of Swedish Men? Time Diary vs. Survey Data," *Labour Economics*, 4 (167-183).

Cheng, H. and O. Capps, Jr. (1988) "Demand Analysis of Fresh and Frozen Finfish and Shellfish in the united States," *American Journal of Agricultural Economics*, 70:3 (533-542).

Clarke, K. (2007) "A Simple Distribution-Free Test for Nonnested Model Selection," *Political Analysis Advance Access*, DOI: 10.1093/pan/mpm004 (February).

Cragg, J. (1971) "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," *Econometrica*, 39:5 (829-844).

Day, J., A. Janus and J. Davis (2005) "Computer and Internet Use in the United States: 2003", *US Census Bureau, Special Studies*, Series P23-208 (October).

Deaton, A. and M. Irish (1984) "Statistical Models for Zero Expenditures in Household Budgets," *Journal of Public Economics*, 23:1-2 (59-80).

Dow, W. and E. Norton (2003) "Choosing Between and Interpreting the Heckit and Two-Part Models for Corner Solutions," *Health Services & Outcome Research Methodology*, 4 (5-18).

Fabiosa, J. (2006) Westernization of the Asian Diet: the Case of Rising Wheat Consumption in Indonesia, Center for Agricultural and Rural Development, Iowa State University, *Working Paper* 06-WP422.

Feldman, L. and J. Hornik (1981) "The Use of Time: An Integrated Conceptual Model", *Journal of Consumer Research*, 7.

Fennema, J. and M. Sinning (2007) "Double-Hurdle Models with Dependent Errors and Heteroscedasticity," *Stata Users Group, German Stata Users' Group Meetings 2007*, Number: 05, http://ideas.repec.org/p/boc/dsug07/05.html

Flood, L. and U. Gråsjö (1998) Regression Analysis and Time Use Data A comparison of Microeconometric Approaches with Data from the Swedish Time Use Survey (HUS), Göteborg University, School of Economics and Commercial Law, *Working Papers in Economics,* No 5.

Flood, L. and U. Gråsjö (2001) "A Monte Carlo Simulation Study of Tobit Models," *Applied Economics Letters*, 8 (581-584).

Gauthier, A., J. Gershuny and K. Fisher (2006) Multinational Time Use Study (MTUS) User's Guide and Documentation, Version 2 (February).

Gebremedhin, B. and S. Swinton (2002) "Investment in Soil conservation in Northern Ethiopia: the Role of Land Tenure Security and Public Programs," *Agricultural Economics*, 29 (69-84).

Greene, W. (2000) Econometric Analysis, 4th ed. Macmillan, New York.

Gronau, R. (1980) "Home Production – A Forgotten Industry", *The Review of Economics and Statistics*, 62:3 (August).

Heckman, J. (1979) "Sample Selection Bias as a Specification Error," *Econometrica*, 47:1 (153-161).

Heien, D. and C. Durham (1991) "A Test of the Habit Formation Hypothesis Using Household Data," *The Review of Economics and Statistics*, 73:2 (189-199).

Heien, D. and R. Wessells (1990) "Demand Systems Estimation with Microdata: A Censored Regression Approach," *Journal of Business and Economic Statistics*, 8:3 (365-371).

Isham, J., J. Kolodinsky, and G. Kimberly (2006) "The Effects of Volunteering for Nonprofit Organizations on Social Capital Formation: Evidence from a Statewide Survey," *Nonprofit and Volunteer Sector Quarterly*, 35:3 (367-383).

Jensen, H. and S. Yen (1996) "Food Expenditures Away from Home by Type of Meal," *Canadian Journal of Agricultural Economics*, 44 (67-80).

Jones, A. (1989) "A Double-Hurdle Model of Cigarette Consumption," *Journal of Applied Econometrics*, 4 (23-39).

Jones, A. (1992) "A Note on Computation of the Double-Hurdle Model with Dependence with an Application to Tobacco Expenditure," *Bulletin of Economic Research*, 44:1 (67-74).

Kalenkoski, C., D. Ribar, and L. Stratton (2005) "Parental Child Care in Single-Parent, Cohabiting, and Married-Couple Families: Time-Diary Evidence from the United Kingdom," *American Economic Review*, 95:2 (194-198) (May).

Kimmel, J. and R. Connelly (2007) "Determinants of Mothers' Time Choices in the United States: Caregiving, Leisure, Home Production, and Paid Work," *Journal of Human Resources*, 42:3 (643-681) (Summer).

Kooreman, P. and A. Kapteyn (1987) "A Disaggregated Analysis of the Allocation of Time within the Household", *Journal of Political Economy*, 95:21.

Martínez-Espiñeira, R. (2006) "A Box-Cox Double-Hurdle Model of Wildlife Valuation: The Citizen's Perspective," *Ecological Economics*, 58:1 (192-208).

McDowell, A. (2003) "From the help desk: hurdle models," *The Stata Journal*, 3:2 (178-184).

Moffatt, P. (2005) "Hurdle Models of Loan Default," *Journal of the Operational Research Society*, 56 (1063-1071).

National Science Foundation (NSF), Division of Science Resources (2001) The Application and Implications of information Technologies in the Home: Where Are the Data and What Do They Say? NSF 01-313; Author, Maria C. Papadakis (February).

Newman, C., M. Henchion and A. Matthews (2003) "A Double-Hurdle Model of Irish Household Expenditure on Prepared Meals," *Applied Economics*, 35:9 (1053-1061).

Saha, A,, O. Capps, Jr. and P. Byrne (1997) " Calculating Marginal Effects in Models for Zero Expenditures in Household Budgets Using a Heckman-Type Correction" *Applied Economics*, 29:10 (1311-116).

Sayer, L., S. Bianchi and J. Robinson (2004) "Are Parents Investing Less in Children? Trends in Mothers' and Fathers' Time with Children," *American Journal of Sociology*, 110:1 (1-43).

Schwierz, C. (2003) The Effects of Taxes and Socioeconomic Variables on Market Work and Home Production in Norway in the Years 1970 to 2000, Department of Economics, University of Oslo, Memorandum No. 33/2003.

Shrestha, R., J. Alavalapati, A. Seidl, K. Weber and T. Suselo (2006) "Estimating the Local Cost of Protecting Koshi Tappu Wildlife Reserve, Nepal: A Contingent Valuation Approach," *Environment, Development and Sustainability*, DOI 10.1007/s 10668-006-9029-4.

Tobin, J. (1958) "Estimation of Relationships for limited dependent Variables," *Econometrica*, 46:1 (24-36).

Venkatesh, A., N. Vitalari, and K. Gronhaug (1985) "Computing in the Home: Shifts in the Time Allocation Patterns of Households", *Communications in the ACM*, 28 (512-522).

Verdin-Johansson, M. (1999) "Determinants of Charitable Giving. Evidence from Hypothetical Data on Willingness to Pay," in Economics without Markets. Four Papers on the Contingent Valuation and State Preference Methods. Umeå University, Department of Economics, *Umeå Economic Studies*, *Working Papers* 517.

Yen, S. (1993) "Working Wives and Food Away from Home: The Box-Cox Double Hurdle Model," *American Journal of Agricultural Economics*, 75:4 (884-895).

Yen, S. (2005) "A Multivariate Sample-Selection Model: Estimating Cigarette and Alcohol Demands with Zero Observations," *American Journal of Agricultural Economics,* 87(2) (453-466).

Yen, S. and H. Jensen (1996) "Determinants of Household Expenditures on Alcohol," *The Journal of Consumer Affairs*, 30:1 (48-67).

Yen, S. and A. Jones (1997) "Household Consumption of Cheese: An Inverse Hypbolic Sine Double-Hurdle Model with Dependent Errors," *American Journal of Agricultural Economics*, 79 (246-251).

Appendix A1: Variables for which minutes spent in using computer/Internet van be reported in ATUS.

American Time Use Survey Activity Lexicon: 2006

Major Categories 1st - tier	2nd - tier	3rd - tier	Examples

02 Household Activities

09 Household Management

04 Household and personal e-mail and messages

Reading e-mail (personal or household)
Instant messaging (personal)
Sending e-mail (personal or household)
Reading/sending e-mail, not specified
Checking e-mail (personal or household)
Cleaning out e-mail inbox (personal or household)

12 Socializing, Relaxing, and Leisure

03 Relaxing and Leisure

08 Computer use for leisure (except games)

Computer use, unspecified
Computer use, leisure (personal interest)
Surfing the internet (personal interest)
Downloading files, music, pictures (personal interest)
Surfing the web (personal interest)
Participating in a chat room (personal interest)
Burning CDs (personal interest)
Designing/updating website (personal interest)
Browsing on the internet (personal interest)

15 Volunteer Activities

01 Administrative and Support Activities

01 Computer use

Writing/sending e-mail (volunteer)
Checking e-mail (volunteer)
Designing website for volunteer organization
Computer use, unspecified (volunteer)
Surfing the internet (volunteer)

Appendix A2: Variables for which time for computer/Internet use should be able to be reported in ATUS

American Time Use Survey Activity Lexicon: 2006

Major Categories 1st - tier	2nd - tier	3rd - tier	Examples

02 Household Activities

09 Household Management

01 Financial management

Making a budget
Balancing the checkbook
Checking/trading stocks
Buying/selling stocks
Researching investments
Paying bills
Filling out tax forms
Using computer to keep budget
Using computer to pay bills

02 Household and personal organization and planning

Planning household trip
Planning vacation

03 Caring For and Helping Household Members

02 Activities Related to household Children's Education

01 Homework (household children)

Helping household child with a school project

03 Home schooling of household children

Home schooling a household child

99 Activities related to household children education, n.e.c.

04 Caring For and Helping Nonhousehold Members

05 Helping Nonhousehold Adults

05 Financial management assistance for nonhousehold adults

Helping nonhousehold adult with computer applications
Helping nonhousehold adult with taxes/bills

05 Working and Work-Related Activities

04 Job Search and Interviewing

01 Job search activities

Contacting employer
Sending out resumes
Sending resumes to employers
Placing/answering ads
Researching details about a job
Researching an employer

Asking for references from former employers
Filling out job application
Submitting applications
Reading ads in paper/on Internet
Checking vacancies
Writing/updating resume
Picking up job application

03 Job interviewing
Preparing for interview

06 Education

01 Taking Class

01 Taking class for degree, certification, or licensure
Taking on-line course (degree, certification, or licensure)

02 Taking class for personal interest
Taking on-line course (personal interest)

03 Research/Homework

01 Research/homework for class for degree, certification, or licensure
Writing paper/essay (degree, certification, or licensure)
Organizing notes (degree, certification, or licensure)
Reading/sending e-mail (degree, certification, or licensure)

02 Research/homework for class for personal interest
Writing paper/essay (class for personal interest)
Organizing notes (class for personal interest)
Reading/sending e-mail (class for personal interest)

04 Registration/Administrative activities

01 Administrative activities: class for degree, certification, or licensure
Enrollment in class (degree, certification, or licensure)
Registration for class (degree, certification, or licensure)
Looking at course descriptions/listings (degree, certification, or licensure)
Paying tuition/fees for class (degree, certification, or licensure)
Getting parking pass for class (degree, certification, or licensure)

02 Administrative activities: class for personal interest
Enrollment in class (personal interest)
Registration for class (personal interest)
Paying tuition for class (personal interest)
Getting parking pass for class (personal interest)
Paying fees for class (personal interest)

07 Consumer Purchases

01 Shopping (Store, Telephone, Internet)

01 Grocery shopping
Ordering groceries
Paying for groceries

Appendix A3-1: Computer and communication media ownership at home (CPS October 2003)

	% Reporting
Have computer/laptop at home	69.2
Own 1	67.7
2	21.4
3 or more	10.9
Newest bought	
2003	20.2
2002	22.7
2001	19.1
2000	15.6
1999	8.4
1998	6.0
Before 1998	8.2
Communication or entertainment media used at home	
Telephone	82.6
Cell phone	55.0
Cable TV	54.5
Satellite TV	16.1
Pager	6.9
PDA	6.6

Appendix A3-2: Computer use at home (CPS October 2003)

	% Reporting
Use computer at home	83.7
Internet connection (personal)	90.6
e-mail (personal)	72.3
School assignment	74.8
Playing games (without Internet)	58.9
Word processing or desktop publishing	53.4
Work from home	38.8
Graphics (images, photo, etc.)	37.8
Database/spreadsheet	31.9
Managing household records/finance	31.2

Appendix A3-3: Internet connectivity at home (CPS October 2003)

		% Reporting
Connected from home		61.5
Device for connection:	Home PC	93.2
	Laptop	5.7
	TV-based Internet	0.76
	Mobile	0.15
	PDA	0.08
	Game Machine	0.02
	Other means	0.19
Connection type:	Dial-up	62.6
	DSL (digital subscriber line)	14.5
	Cable modem	21.6
	Fixed wireless connection	0.38
	Others	0.97
Why not high-speed Internet?	Don't need/not interested	41.5
	Too expensive	39.4
	Not available in the area	11.8
	Others	7.3
Last year's frequency of use:	At least once/day	53.8
	At least once/week, not every day	33.5
	At least once/month, not every week	7.8
	Less than once/month	4.9
Why not have Internet at home?	Too high costs	27.1
	Don't need/not interested	35.1
	No computer	23.5
	Lack of skill	3.7
	Others	10.6
Access Internet at (non-owners):	Work place	18.3
	School	11.6
	Someone else's house	6.3
	Public library	5.7
	Airport, hotel, etc	2.5
	Internet café/coffee shop	0.85

Appendix A3-4: Current use of Internet (CPS October 2003)

	% Reporting
Email or instant messaging	42.0
Search info about products and services	35.3
Get news, weather or sports info	30.8
Purchasing products and services	23.2
Playing game	21.3
Search info about health services or practices	18.0
Search info about government services or agencies	15.2
Complete school assignment	12.7
Online banking	11.7
Download federal, state or local government forms	11.3
Listening to radio or viewing TV or movie	10.2
Submit federal, state or local government forms	7.9
Search for a job	7.8
Read online job ads, or search online job listings	7.2
Search info about potential employers	5.4
Submitting resume or application to employer	4.3
Post resume on a job listing site or with a service	3.0
Trade stocks, bonds or mutual funds	2.8
Taking online course	2.7
For telephone calls	1.6

Appendix A4: Estimation of the predicted wage

```
Heckman selection model -- two-step estimates   Number of obs      =     11193
(regression model with sample selection)        Censored obs       =      4997
                                                Uncensored obs     =      6196
                                                Wald chi2(14)      =   2535.48
                                                Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
lwage        |
         age |   .0522724   .0068476     7.63   0.000     .0388513    .0656934
        age2 |  -.0005062   .0000874    -5.79   0.000    -.0006775   -.0003348
        educ |    .083443   .0038055    21.93   0.000     .0759844    .0909017
      female |  -.1899835   .0175677   -10.81   0.000    -.2244154   -.1555515
    nonwhite |  -.0996503   .0206858    -4.82   0.000    -.1401938   -.0591068
  noncitizen |  -.0948663   .0295709    -3.21   0.001    -.1528241   -.0369084
       urban |   .1780987   .0214851     8.29   0.000     .1359887    .2202087
       _cons |   .4645046   .1952172     2.38   0.017     .0818859    .8471232
-------------+----------------------------------------------------------------
select       |
         age |   .1331976   .0050439    26.41   0.000     .1233117    .1430835
        age2 |  -.0015045   .0000534   -28.19   0.000    -.0016092   -.0013999
        educ |   .1543135   .0258569     5.97   0.000     .1036349     .204992
       educ2 |  -.0010325   .0008637    -1.20   0.232    -.0027253    .0006604
    age*educ |  -.0016735   .0002893    -5.78   0.000    -.0022405   -.0011064
      female |  -.1915738   .0264974    -7.23   0.000    -.2435078   -.1396399
    nonwhite |  -.0345399   .0345417    -1.00   0.317    -.1022403    .0331606
  noncitizen |   .0801668   .0514446     1.56   0.119    -.0206628    .1809963
       urban |   .0302866   .0347139     0.87   0.383    -.0377514    .0983246
     child02 |  -.2476569    .033098    -7.48   0.000    -.3125278   -.1827859
     child35 |  -.1755854   .0318641    -5.51   0.000     -.238038   -.1131328
     child69 |  -.0748801   .0264325    -2.83   0.005     -.126687   -.0230733
   child1012 |   -.093557   .0311891    -3.00   0.003    -.1546865   -.0324275
    hhincome |   2.05e-06   3.59e-07     5.72   0.000     1.35e-06    2.76e-06
       _cons |  -3.202355   .2135439   -15.00   0.000    -3.620893   -2.783817
-------------+----------------------------------------------------------------
mills        |
      lambda |  -.2054656   .0818002    -2.51   0.012     -.365791   -.0451402
-------------+----------------------------------------------------------------
         rho |   -0.33144
       sigma |  .61991217
      lambda | -.20546556   .0818002
------------------------------------------------------------------------------
```

Appendix A5: Specification issues in the double-hurdle model

The main issue in this paper is specifying the use of computer/Internet at home in terms of the framework of the double-hurdle model. As discussed in the preceding sections, the double-hurdle model involves two distinct decisions: the participation decision (whether to own a home computer/Internet) and the level of participation (the extent of use of computer/Internet). The type of association between these decisions and the specifications of the error terms determine the likelihood function to be estimated. Hence, if an individual makes both decisions separately, the two decisions are modeled independently; or if both decisions are made simultaneously, they are modeled jointly; or if one decision is made first and affects the other one, they are modeled sequentially (Martínez-Espiñeira, 2006). The resulting models are called the *independence*, the *dependence* and the *dominance* models, respectively. In addition, in limited dependent variable models, the likelihood functions are derived based on the assumptions of normality and homoskedasticity of the error terms. When either assumption is violated, the corresponding maximum likelihood (ML) estimates become inconsistent (Amemiya and Powell, 1981; Arabmazar and Schmidt 1981, 1982).

This paper does not address each of these specification issues. Instead the study uses the double-hurdle model derived based on the assumptions of independent ownership and usage decisions (i.e., independent error terms) and homoskedastic and normally distributed error terms. Dealing with these error specification issues will be an area of extension in future studies.

Appendix A6: Maximum Likelihood Estimation of Tobit, Heckman and Double-Hurdle Models: Total Minutes of Computer Use

Variable	Tobit	Heckman		Double-Hurdle	
		1st Hurdle	2nd Hurdle	1st Hurdle	2nd Hurdle
Age	-1.344 (8.65)***	-0.009 (9.55)***	0.124 (0.70)	0.014 (1.31)	-1.634 (7.54)***
Education	25.319 (4.43)***		-5.588 (0.95)		26.372 (4.55)***
Educ2	-0.483 (2.37)**		0.116 (0.57)		-0.528 (2.55)**
Disability	18.787 (1.64)	-0.076 (0.99)	39.423 (3.57)***	-0.772 (1.75)*	46.254 (1.82)*
Female	-14.861 (3.31)***	-0.037 (1.34)	-14.186 (3.55)***	-0.066 (0.27)	-13.764 (2.30)**
Nonwhite	-17.703 (3.08)***	-0.120 (3.17)***	13.022 (2.40)**	-1.270 (4.97)***	14.412 (0.96)
Noncitizen	-26.850 (2.89)***	-0.264 (4.50)***	1.815 (0.20)	9.151 (0.00)	-43.196 (3.66)***
Married & spouse or partner present	-15.994 (2.95)***	-0.032 (0.93)		-0.148 (0.55)	
child02	-32.842 (4.90)***	-0.147 (3.40)***	-6.983 (1.23)	-0.095 (0.35)	-31.252 (4.34)***
child35	-20.633 (3.23)***	-0.091 (2.18)**	-7.560 (1.45)	-0.002 (0.01)	-20.202 (2.92)***
child69	-15.275 (2.82)***	-0.064 (1.81)*	-3.708 (0.89)	0.409 (1.20)	-18.603 (3.51)***
child1012	-10.312 (1.74)*	-0.025 (0.65)	-8.977 (2.00)**	0.377 (1.04)	-13.479 (2.23)**
child1317	7.454 (1.50)	0.051 (1.56)	-3.726 (1.11)	0.254 (0.80)	5.239 (0.95)
Family size	3.782 (1.23)	-0.018 (0.87)		0.051 (0.37)	
Urban	20.338 (3.43)***	0.164 (4.33)***	0.342 (0.06)	1.371 (3.93)***	-11.915 (1.05)
Northeast	6.724 (1.10)	0.055 (1.36)	1.570 (0.29)	1.269 (1.79)*	-11.887 (1.46)
Midwest	13.410 (2.44)**	0.065 (1.80)*	10.003 (2.03)**	0.512 (1.80)*	0.600 (0.08)
West	17.802 (3.16)***	0.101 (2.72)***	8.964 (1.78)*	0.759 (1.80)*	3.579 (0.48)
Weekdays	12.102 (2.90)***	0.117 (4.27)***	-12.491 (3.21)***	-0.265 (1.23)	16.053 (3.10)***
Spring	2.505 (0.42)	0.034 (0.86)	-4.220 (0.80)	-0.600 (1.69)*	10.574 (1.42)
Fall	1.586 (0.26)	0.032 (0.80)	-6.085 (1.11)	-0.365 (1.07)	5.695 (0.79)
winter	11.372 (1.91)*	0.083 (2.12)**	2.879 (0.54)	-0.643 (1.81)*	20.860 (2.58)***
Income below $50,000	-16.639 (3.46)***	-0.272 (8.94)***	15.623 (3.24)***	-0.284 (1.15)	-9.908 (1.65)*
High wage	-42.368 (3.25)***		-3.609 (0.30)		-51.507 (3.91)***
Medium wage	-33.128 (3.63)***		-6.664 (0.75)		-42.965 (4.52)***
Constant	-332.336 (8.33)***	-0.591 (6.92)***	144.069 (3.34)***	0.134 (0.17)	-274.392 (6.47)***
Observations	12943	12943		12943	
Absolute value of t statistics in parentheses for Tobit and Heckman models. Absolute value of z statistics in parentheses for double-hurdle model.					
* significant at 10%; ** significant at 5%; *** significant at 1%					
Some variables are omitted from either of the hurdles due to lack of convergence during estimation. This Could arise due to identification problem.					

Table 1: Descriptive Statistics

Variables		Weekends		Weekdays	
		Mean	**Std. Dev**	**Mean**	**Std. Dev**
Respondent characteristics	Age	45.49	17.77	46.03	17.74
	Female	0.58	0.49	0.57	0.50
	Education (years)	13.32	3.30	13.39	3.17
	Disability	0.04	0.20	0.04	0.20
	Nonwhite	0.17	0.38	0.18	0.39
	Black	0.12	0.33	0.13	0.34
	Hispanic	0.14	0.35	0.13	0.34
	Non-citizen	0.08	0.27	0.07	0.26
Household characteristics	Spouse or unmarried partner present*	0.53	0.50	0.53	0.50
	No. of children age 0 to 2	0.14	0.40	0.13	0.39
	No. of children age 3 to 5	0.15	0.41	0.14	0.40
	No. of children age 6 to 9	0.21	0.49	0.21	0.49
	No. of children age 10 to 12	0.17	0.42	0.16	0.42
	No. of children age 13 to 17	0.30	0.60	0.30	0.61
	No. of other adults in the household	0.45	0.83	0.44	0.83
	Family size	2.87	1.55	2.84	1.55
Location and season characteristics	Urban	0.82	0.39	0.81	0.39
	Northeast	0.17	0.38	0.17	0.38
	Midwest	0.25	0.43	0.24	0.43
	South	0.36	0.48	0.36	0.48
	West	0.22	0.41	0.22	0.41
	Summer	0.24	0.43	0.23	0.42
	Fall	0.25	0.43	0.24	0.43
	Winter	0.26	0.44	0.26	0.44
	Spring	0.26	0.44	0.28	0.45
Economic characteristics	Employed respondent	0.64	0.48	0.63	0.48
	Employed spouse or unmarried partner‡	0.38	0.48	0.38	0.48
	Log of predicted wage	2.76	0.41	2.77	0.40
	Family income below 50,000	0.59	0.49	0.60	0.49
	Management, professional, service, sales and office occupations@	0.52	0.50	0.51	0.50
	Construction, maintenance, production, transport and material moving	0.12	0.33	0.12	0.32
	Farming, fishing and forestry	0.004	0.07	0.004	0.07
Sample size		**6457**		**6486**	

*Those living with their spouses are 50.4%.

‡The proportion of respondents whose spouses or partners are employed is 71%.

@About 36% of the respondents are unemployed and have no occupations.

Table 2a: Average Minutes Spent per Day in Using Computer/Internet (including zeros)

Category	Total minutes of computer/Internet use	Household activities	Relaxing, socializing & leisure	Voluntary activities
Weekends	11.61 (42.12) 6457	2.49 (13.88) 6457	8.55 (37.91) 6457	0.56 (9.93) 6457
Weekdays	11.60 (42.50) 6486	2.88 (14.27) 6486	7.99 (37.76) 6486	0.72 1(2.16) 6486
Female	10.00 (37.88) 7427	2.72 (13.47) 7427	6.43 (32.14) 7427	0.85 (13.12) 7427
Male	13.76 (47.54) 5516	2.64 (14.86) 5516	10.74 (44.25) 5516	0.37 (7.60) 5516
Employed	10.29 (38.22) 8250	2.20 (11.54) 8250	7.61 (35.08) 8250	0.48 (8.72) 8250
Unemployed	13.92 (48.59) 4693	3.55 (17.65) 4693	9.44 (42.22) 4693	0.93 (14.36) 4693

Table 2b: Average Minutes Spent per Day in Using Computer/Internet (excluding zeros)

Category	Total minutes of computer/Internet use	Household activities	Relaxing, socializing & leisure	Voluntary activities
Weekends	83.92 (82.24) 893	42.38 (39.83) 380	95.50 (88.12) 578	79.11 (88.21) 46
Weekdays	70.90 (82.72) 1061	35.14 (36.78) 532	84.28 (92.83) 615	72.32 (98.64) 65
Female	69.70 (76.44) 1066	36.37 (34.67) 556	81.38 (83.55) 587	84.07 (100.76) 75
Male	85.44 (89.00) 888	40.95 (43.11) 356	97.79 (96.54) 606	56.53 (76.43) 36
Employed	69.21 (75.86) 1226	32.47 (31.37) 559	83.11 (84.74) 755	60.00 (77.57) 66
Unemployed	89.72 (91.80) 728	47.17 (45.71) 353	101.11 (99.26) 438	97.33 (111.36) 45

Table 3: Average Minutes of Compute Use by Age of Children, Marital Status and Day of the Week*

		Households with children between age				
		0 to 2	3 to 5	6 to 9	10 to 12	13 to 17
Weekends, Married/spouse present	Total computer use	75.62 (55.25) 73	65.40 (59.65) 81	77.52 (65.99) 98	70.63 (60.46) 75	65.19 (54.69) 108
	Household activities	38.33 (27.29) 24	28.79 (21.94) 33	26.28 (21.81) 39	24.10 (25.02) 31	26.10 (23.71) 48
	Relaxing, socializing and leisure	81.36 (58.54) 55	71.58 (59.16 55	84.88 69.01 69	77.22 60.38 54	78.96 55.55 70
Weekends, Not married/no spouse present	Total computer use	74.69 (67.14) 16	115.28 (147.95) 18	91.40 (104.45) 47	85.67 (68.71) 66	98.70 (101.22) 189
	Household activities	80.00 (101.14) 6	43.75 (91.80) 8	27.75 (28.21) 20	41.75 (43.01) 24	54.60 (60.15) 60
	Relaxing, socializing and leisure	59.58 (45.65) 12	132.69 (158.74) 13	116.37 (114.15) 30	97.83 (67.84) 47	104.37 (107.44) 145
Weekdays, Married/spouse present	Total computer use	46.85 (46.43) 95	52.89 (54.78) 94	46.21 (43.80) 135	49.28 (48.52) 113	53.72 (78.52) 130
	Household activities	30.00 (24.68) 48	30.39 (28.66) 44	28.34 (25.86) 76	28.61 (21.52) 54	27.27 (30.17) 62
	Relaxing, socializing and leisure	50.27 (43.08) 51	60.25 (54.68) 53	51.10 (48.59) 72	60.68 (57.62) 63	75.11 (100.99) 65
Weekdays, Not married/no spouse present	Total computer use	71.11 (62.63) 18	49.52 (41.44) 27	79.04 (129.83) 50	71.22 (65.60) 60	83.46 (92.34) 159
	Household activities	26.88 (26.98) 8	26.00 (20.72) 15	33.24 (32.00) 21	54.56 (70.17) 25	42.97 (51.98) 73
	Relaxing, socializing and leisure	100.50 (67.39) 10	63.36 (50.01) 14	88.80 (138.74) 30	79.97 (62.19) 35	87.18 (91.48) 107

* *Note: Average minutes of computer use for volunteer activities are excluded from the analysis due to small number of observations.*

Table 4: Average Minutes of Compute Use by Level of Wage, Marital Status and Day of the Week*

		Predicted wage		
		Low	Medium	High
Weekends, Married/spouse present	Total computer use	69.29 60.99 7	77.98 73.78 261	74.09 68.02 148
	Household activities	28.33 27.54 3	36.24 32.20 105	38.87 34.00 75
	Relaxing, socializing and leisure	85.00 64.03 4	88.72 75.73 174	75.68 64.85 88
Weekends, Not married/no spouse present	Total computer use	100.89 96.76 146	87.05 92.88 264	81.01 57.53 67
	Household activities	65.76 62.41 51	40.09 34.71 115	43.06 36.63 31
	Relaxing, socializing and leisure	106.15 103.49 105	107.13 104.11 166	93.49 57.11 41
Weekdays, Married/spouse present	Total computer use	68.00 50.70 5	59.74 68.51 404	55.18 59.59 130
	Household activities	33.33 25.17 3	28.59 22.27 210	30.62 34.28 66
	Relaxing, socializing and leisure	120.00 0.00 2	72.97 80.17 226	60.00 56.67 69
Weekdays, Not married/no spouse present	Total computer use	75.83 68.11 136	85.57 100.30 304	88.48 113.47 82
	Household activities	51.43 56.43 56	42.67 45.39 150	27.45 20.02 47
	Relaxing, socializing and leisure	80.14 68.98 92	101.34 111.40 176	115.00 131.68 50

* *Note: Average minutes of computer use for volunteer activities are excluded from the analysis due to small number of observations.*

† *The average minutes reported for the low-income group appear to be larger mainly because the number of observations is very small in the majority of the cases. As this makes comparison with the other groups difficult, this group is also excluded from the analysis.*

Table 5: Average minutes spent on various activities and services†

Minutes spent	Respondents who reported computer use			Respondents who did not report computer use			Differences‡
	mean	std	obs	mean	std	obs	
Computer or Internet use	76.85	82.74	1954	-	-	-	-
Government services	31.67	28.39	9	65.60	68.45	52	-2.53*
Travel: government services	21.56	28.45	9	29.89	25.81	46	-0.82
Civic obligation	12.50	3.54	2	85.04	111.22	24	-3.18**
Travel: civic obligation	10.00	0.00	2	29.89	38.68	19	-2.24*
Financial and banking services	11.41	13.14	85	13.52	17.85	283	-1.19
Grocery shopping	29.97	28.38	582	32.92	31.59	2823	-2.24*
Travel : grocery shopping	20.88	15.71	349	24.59	24.05	1753	-3.64**
Nongrocery shopping	60.45	65.53	604	72.05	75.71	2986	-3.86**
Travel: nongrocery shopping	32.15	39.28	734	36.94	46.07	3592	-2.92**
Purchase research	82.60	79.59	5	40.00	16.69	8	1.18
Job search and interview	85.55	97.43	31	130.83	134.13	87	-2.00*
Travel: job search and interview	64.08	104.77	13	57.58	70.52	45	0.21
Leisure without computer use	255.48	168.12	1889	299.73	200.38	10485	-10.21**
Physical activities	91.95	89.54	434	104.58	100.44	1737	-2.56*
Paid work	367.77	211.64	731	434.82	198.14	4108	-7.97**

† All the reported minutes, except those in the first row, are for non-computer/Internet uses.

‡ Difference between the mean minutes reported by computer users and nonusers.

* significant at 5%, ** significant at 1% for the difference in means.

Table 6: Maximum likelihood estimation of the Double-hurdle model: Total minutes of computer use (for all days)

Variable	Double-hurdle			
	1st Hurdle		2nd Hurdle	
Age	-0.026	(1.93)*	0.028	(.)
Age2	-0.0001	(0.70)		
Education	0.140	(1.59)		
Educ2	0.005	(1.31)		
Disability	-0.951	(5.64)***	102.897	(3.16)***
Female	0.106	(1.01)	-20.445	(3.52)***
Nonwhite	-0.441	(4.29)***		
Noncitizen	-0.631	(3.92)***		
Married & spouse or partner present	0.283	(2.97)***		
child02	-0.535	(3.22)***	-12.369	(1.59)
child35	-0.179	(1.01)	-11.331	(1.54)
child69	0.177	(0.92)	-16.162	(2.74)***
child1012	0.053	(0.33)	-6.790	(1.04)
child1317	0.207	(1.24)	-2.580	(0.44)
Other adults	-0.066	(0.77)	8.662	(1.87)*
Urban	0.322	(3.25)***		
Northeast	0.204	(1.80)*		
Midwest	0.338	(3.08)***		
West	0.221	(2.06)**		
Spring			3.832	(0.57)
Fall			1.481	(0.22)
Winter			14.317	(2.07)**
Job occup1 (if employed)			-46.363	(9.48)***
Job occup2 (if employed)			-66.438	(7.81)***
Employed spouse			-22.606	(4.38)***
Income above $50,000			15.554	(3.34)***
High wage			-13.240	(2.31)**
Medium wage			-13.249	(.)
Constant	-1.276	(1.97)**	-47.13	(.)
No. of observations	12,943			

Absolute value of z statistics in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%
Job occup1: Management, professional, service, sales and office occupations
Job occup2: Construction, maintenance, production, transport and material moving