A New Interpretation Approach using Tobit Analysis : Simulations based on Type I Tobit of Amemiya
- Focused on Childcare Services -

(주)파이슨 대표이사*
Sun-Young Park, Ph.D.

〈목 차〉

〈Abstract〉

The purposes of this study were first to construct statistical and econometric models based on Amemiya's Type I Tobit mainly addressing the issue of statistical efficiency; second to explore income, price, and curvilinear age effects on the explained variable in order to illustrates its statistical marginal effects related to econometric issues; finally to provide invaluable insight for graphical simulations as a new interpretation approach using Tobit analysis. Results indicated that interpretation for the mean marginal effects of three possible cases of dependent variable was more likely to be evident to understand Tobit results compared to conventional analysis only using latent variable, beta. Results also revealed that prediction value of dependent variable can be possibly and easily projected by the independent variable changed whereas only beta value can not illustrate its projection as independent variables' changes

◆ 주요어: Tobit, marginal effect, simulation

Ⅰ. INTRODUCTION

There has been a lot of issues of statistical research methodology in Family and Consumer Studies. Two big streams are mainly argued among statisticians and researchers. One may prefer quantitative analysis, others may prefer qualitative analysis or some researchers may concern both.

* (주)파이슨(www.pycn.co.kr)은 한국중소IT벤처기업연합회(PICCA) 공인 해외마케팅전문업체임.

Household's decisions regarding the demand and choice for child care services, for instance, is of great concern among consumer or family scholars. Traditionally, there have been active researches on mother's major role in the family with respect to child care, child care giver, mothers' labor force, family's role as a major provider and so on. Child care has become an important family issue as well as a public policy issue because women's labor force participation has greatly increased since 1950. Furthermore, child care service is a necessary and important consideration, not only for the working households but also for all households with young children.

In spite of the major interest in child care services for all types of households, there has been a few research focusing on research methodology regarding statistical efficiency and the way of interpretation using simulation which is more likely to be needed in statistical methodology(Mok et. al, 1995; Park, 1995; Park and Mok,1995; Park et.al, 1995) Therefore, this study will explore the issues of demand in microeconomic perspectives focusing statistical issues as well as metholodological issues.

The goals of this study therefore are described as follows: first to construct statistical and econometric models based on Amemiya's Type I Tobit mainly addressing the issue of statistical efficiency; second to explore income, price, age, and other socio-demographic effects on the explained variable in order to illustrate its statistical marginal effects with related to econometric issues; finally to provide invaluable insight for graphical simulations as a new approach for the statistical interpretation.

Ⅱ. THEORETICAL CONSIDERATION

Tobit model represents regression model in which the dependent variable is observed in only some of the ranges. The pioneering work of Tobit model was done by Tobin (1958). In his work, he estimated household expenditure on durable goods using regression model but he considered the fact that the value of the dependent variable could not be negative. He called his specific case of regression model as limited dependent variable method. This Tobit model is also known as a censored normal regression model (Maddala, 1983; 1992). In Amemiya's 1984 article titled "Tobit Models: A Survey", he classified the diverse types of Tobit. In this paper, type I Tobit model introduced by Amemiya(1984,1985) will be introduced since demand model for child care services mainly falls into such type.

1. Standard Tobit Model

: Type I Tobit : $p(y1<0) \cdot p(y1)$

According to Amemiya(1984,1985), a household is assumed to maximize the utility, subject to budget constraint, which is described as follows.

Max U(y,z)
subject to $y + z \; x$
$y \geq 0$

where y is a household's actual child care expenditures

z is all other expenditures
x is income
$y + z \leq x$ stands for budget constraints, and
$y \geq 0$ stands for boundary constraint.

Also, suppose y* is the solution of the unconstrained optimization(the utility maximization subject to income constraint only), which is denoted by

(1.1) $y_i^* = \beta' x_i + \mu_i$,

where xi are the income and other variables

μ_i indicates all the unobservable variables affecting the household's utility.

We can also define yi* as desired expenditures or potential expenditures. Thus, we can rewrite equation (1.1) as

(1.2) $y_i = y_i^*$ if $y_i^* > 0$

(1.3) $y_i = 0$ if $y_i^* \leq 0$

where yi are observed if a household's potential expenditures is greater than zero ($y_i^* > 0$), yi* are unobserved if a household's potential expendirures is less than zero($yi^* \leq 0$), xi are observed variables,

and ui~iid N (0, σ^2).

Therefore, yi stands for the actual child care expenditures, and x stands for annual earned income, price of child care, financial asset, age of mother, education of mother, mother's working hour, and other variables.

The likelihood function of the standard Tobit model is given by

$$(1.4) \quad L = \prod_0 [1 - \Phi(\beta' x_i/\sigma)] \cdot \prod_1 \sigma \phi [(y_i - \beta' x_i/\sigma)]$$

where Φ, ϕ, stand for the cumulative distribution function and density function respectively of the standard normal distribution, and $\prod_0$ stands for the product over those i for which $yi = 0 (yi^* \leq 0)$, and $\prod_1$ means the product over those i for which $yi = 1 (y_i^* > 0)$.

Maddala (1983,1992) also described the likelihood function for the type I Tobit model which is the same as the description (1.4), and it takes the following form as:

$$(1.5) \quad L = \prod_{y^*>0} \sigma f[(y_i - \beta' x_i)/\sigma] \prod_{y^* \leq 0} (-\beta' x_i/\sigma)$$

where $f(t) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$

$$F(\beta' x_i/\sigma) = \int_{-\infty}^{\beta' x_i/\sigma} f(t)dt$$

The description of the likelihood function is very important because of the maximum value of the function with respect to β and σ^2, yielding consistent and efficient estimators. Numerous studies about consumer's purchase on some durable goods using Tobit model use this type of Tobit.

2. Three Mean Marginal Effects For Type I Tobit

With limited dependent variables, Maddala (1983,1992) and Green(1990,1991) noted three mean marginal effects in Type I Tobit model :

Total mean marginal effect for overall samples

(1.6) $\partial E(y)/\partial X_i = \Phi(z)\beta_j$

Mean marginal effect for latent variables, desired expenditures

(1.7) $\partial E(y^*)/\partial X_i = \beta_j$

Mean marginal effect for the households that had positive expenditures

(1.8) $\partial E(y|y^*>0)/\partial X_i$

$$= \beta_j [1 - z\{\phi(z)/\Phi(z)\} - \{\phi(z)/\Phi(z)\}^2]$$

where $\phi(z)$ is the density function of the standard normal distribution, $\Phi(z)$ is the cumulative distribution function of the standard normal distribution, and $z = x_i'\beta$.

3. Expected Value of Tobit

In addition, the effects of independent variables which show the statistical significance on the explained variable are graphically compared through simulation results. The expected values of dependent variable for all households based on the type 1 Tobit results are given based on the following two equations by Maddala(1983). Following two equations will be used for the simulation as a expected value of prediction.

$$(1.9)\quad E(Y_i) = \Phi_i (\beta' X_i + \sigma (\phi_i / \Phi_i)) + \sigma \phi_i \cdot 0 = \Phi_i \beta' X_i + \sigma \phi_i$$

$$(1.10)\quad E(Y_i \mid Y_i>0) = \beta' X_i + E (u_i \mid u_i >- \beta' X_i) = \beta' X_i + \sigma (\phi_i / \Phi_i)$$

Ⅲ. RESEARCH METHODOLOGY

1. Data Source

The most commonly used household spending data in the U.S.A. is the Consumer Expenditure Survey (CES). The data source in this study was the 1990-1992 CES data. These data have been published by the Bureau of Labor Statistics (BLS). The BLS has been collecting data on the living expenditures and spending patterns of U.S. households for many years. The Consumer Expenditure Survey is a well designed survey where the primary focus is on gathering data related to household expenditures for goods and services purchased and used in our daily lives.

2. Sample Selection

For the purpose of preparing the data set in this study, data for households who were interviewed for any four consecutive quarters during the period from 1990 to 1992 were retrieved and their expenditures were summed to get the total expenditures of each consumer unit. The Consumer Expenditure Survey provides detailed information on expenditures and demographic characteristics of U.S. households. The data used for this study included only consumer units who participated in the interview for four consecutive quarters during 1990 and 1992.

For the purpose of this study, only households with children aged less than seventeen years were analyzxed. The sample size was 971 urban households for those whom have at least one child under age six in the 1990-1992 period.

3. Calculation of Mean Marginal Effects and Simulation

For the calculation of Z score based on the equation from (1.6) to equation (1.8), both SAS and Limdep were used. Also Excel was used for the simulation for the expected or predicted value of explained variable.

4. Definitions and Measurement of Variables

1) Explained Variable

In the estimation of the demand for child care services, the dependent variable is annual total child care expenditure including nursery school, day care, babysitting and "other" types of homecare. In the Consumer Expenditure Survey data set, expenditures on nursery and daycare included the education expenditure section, and babysitting and other home

care are in the miscellaneous expenditure section. Thus, a new aggregated variable, total child care expenditures, is the sum of total household expenses on nursery, daycare, babysitting and other types of home care. The one-stage Tobit model contains this value as a dependent variable.

2) Exogenous Explanatory Variables

The variables listed below will be used not only for basic descriptive statistics but also as independent variables in the further statistical analysis of the demand for child care. In order for illustration of the simulation, three major independent variables can only be explained since the focused issue in this paper is mainly to test a new approach in research methodological issue on a basis of tobit results rather than to show the independent variables effects on the dependent variable.

Price of child care

The Consumer Price Index for child care was incorporated as a proxy for child care prices. As a supplement, the American Chamber of Commerce Researchers Association's (ACCRA) Cost of Living Index(1992) was used in the process of determining Child Care Price.

Household income and financial assets

There were two different variables considered: household annual after-tax income and total financial assets. The household income variable was an after-tax feature and was adjusted by retirement pensions. In the empirical analysis in this study, natural logarithmic income will be used. The household's financial assets variable represented by financial assets was also created, which consists of amounts invested in their own farm or business; amounts in checking accounts, brokerage accounts and other similar accounts; amounts in savings accounts at banks, savings and loans, and credit unions; estimated market value of all stocks, bonds, mutual funds and other such securities; amount of money owned; and amounts in U.S. Saving Bonds.

Mother's age, age squared, and age cubed

Mother's age was measured as the actual reported age of the respondent. Age squared and age cubed was used for the demand and choice model to explore a possible curvilinear relationship between mother's age and household's total child care expenses.

5. Basic Descriptive Statistics

791 urban households with at least one child under age six were used for the analysis. Among the total sample (N=971), about 37.60 percent of the households (N=365) had child care expenditures compared to 62.41 percent of households (N=606) that did not have child care expenditures〈Table 1〉.

Ⅳ. ANALYSES AND RESULTS

1. Mean Marginal Effects and Simulations

As a tool of interpretation, three mean marginal effects can be suggested. Usually researchers are more likely to interpret latent variable's marginal effect which is acknowleged as a beta in Tobit output, however it is not quite correct to interpret Tobit output since latent variable is not actual value of dependent variable.

Hence, in this section, results from type I Tobit model based on three mean marginal effect are discussed. The mean marginal effects in Tobit

〈Table 1〉 Basic Descriptive Statistics of Households with Children aged under Six (N=971), All Households with and without Positive Child Care Expenditures, Pooled Sample, 1990-1992.

Variables	Mean /Freq	S.D.
Total Annual Child Care Expenditure	$755.15	$1,839.18
Expenditure per child	$611.70	$1,397.88
Annual Income[1)]	$34,986.02	$21,039.28
Total Financial asset	$9,079.67	$20,462.20
Wife's Labor Force Participation		
No Work	34.0%	
Part-time	23.2%	
Full-time	42.8%	
Time per week(if work)	22.47	18.88
Age of Mother or Wife	31.27	5.55
〈20	1.5%	
21-30	45.5%	
31-40	48.8%	
〉41	4.5%	
Education of Mother		
Less high school	15.55%	
High school grad	60.35%	
Some college	14.42%	
More college	9.68%	
Number of Children	2.1	1.0
Household Type		
H/W, own children only, oldest child 〈 6	46.9%	
H/W, own children only, oldest child 6-17	38.2%	
H/W, own children only, oldest child 〉 17	1.8%	
Female-Headed household, at least one child 〈18	13.2%	
Region		
Urban Northeast	20.39%	
Urban Midwest	26.78%	
Urban South	26.05%	
Urban West	26.78%	

models are also presented using LIMDEP and SAS. The Simulation results based on the estimated parameters are graphically illustrated using Excel. The marginal effects of Tobit models are twofold: one is based on constrained optimization and the other is based on unconstrained optimization. The mean marginal effects of a latent variable are the solution of unconstrained optimization where only the income constraint is restricted (expenditure can be negative or greater than or equal to zero). On the other hand, marginal effects for both households with and without child care expenditures are the solution of constrained optimization where both income and boundary constraints are restricted (expenditure is greater than or equal to zero).

Thus, since an expected value of the latent variables gives us information on the potential allocation of the child care budget, marginal effects

1) Family income was adjusted by retirement pension. Thus, Family Income =After tax income - Retirement pension

of the latent (unobserved) variables indicate the marginal change of the household's potential pay for child care expenditures (latent variable) with respect to the changes of other independent variables, holding preferences and relative prices constant.

The effects of independent variables which show the statistical significance (income, price and mother's age) on the demand for child care services are graphically compared through simulation results. In simulation, the expected values of child care expenditures for all households based on the Tobit results are given based on the equations (1.9) and (1.10).

〈Table 2〉 represents three mean marginal effects of Tobit model where marginal effects of potential expenditures, actual expenditures within the total sample, and actual expenditures within the sample with positive child care expenditures are given.

2. Income Effects and its simulation

In 〈Table 2〉, income is revealed to be a significant factor at the 0.0001 level. The mean marginal effects of the log of income on total child care expenditures are $304.87 for entire sample and $258.50 for households with positive expenditures, respectively, whereas potential pay for child care based on the unconstrained optimization is $936.27.

〈Figure 1〉 shows the simulation result of income effect on demand for child care, with potential and actual child care expenditures estimated for various

〈Table 2〉 Estimates on Demand for Child Care Services, Represented by Potential, Actual Child Care Expenditure and Mean Marginal Effects, Restricted Type I Tobit Model[a].

Variables(Xi)	β	$\partial E(y)/\partial X_i$	$\partial E(y\ y^*>0)/\partial X_i$	P-value
Intercept	-18,515.65	-6,029.14	-5,112.24	.0151
Lnincome	936.27	304.87	258.50	.0001
LnPrice	-1,540.62	-501.66	-425.37	.0788
LnFA*	40.90	13.32	11.29	.1490
Hours/week	56.13	18.28	15.50	.0001
Age	1,020.43	332.28	281.74	.0504
Agesq	-2,615.91	-851.81	-722.27	.0489
Agecub	2,130.91	693.88	588.35	.0441
Education				
(LessHigh school omitted)				
High	68.69	22.37	18.97	.8682
College	900.40	293.19	248.60	.0766
Morecol	1,539.93	501.43	425.17	.0061
Number of Child				
Age 〈2	441.21	143.67	121.82	.0834
Age 3-5	596.13	194.11	164.59	.0142
Age 6-11	-349.28	-113.73	-96.44	.0661
Age 12-17	-822.84	-267.94	-227.19	.0156
Sigma	3,097.10			
Log-L	-3,157.00			

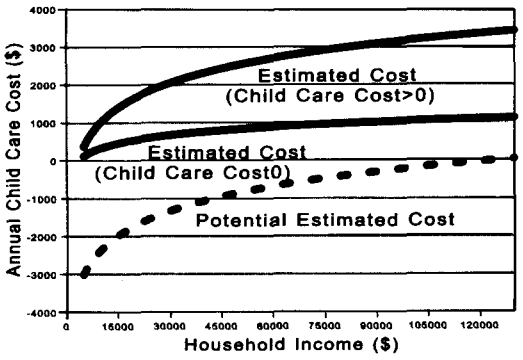
LIMDEP was used for calculating three mean marginal effects.

a: Z score=-0.45, (z)=0.33, (z)=0.36

*: FA indicates Financial Assets

income levels. As indicated by the simulation result, when a household's income is greater than $ 110,000, estimated potential expenditure on the basis of the unconstrained optimization starts to change from zero expenditures to positive expenses. The simulation results also indicate that households' estimated expenditures for households within positive expenditure groups, holding other things constant, are about $2,000 for households with $ 30,000 income; $2,700 for households with $60,000 income; and $3,500 for households with $115,000 income.

〈Figure 1〉 Simulation Results Based on Tobit Estimates in Table 2 with respect to equations (1.9) and (1.10): Effect of Household Income on the Potential, Actual Child Care Expenditure, Holding Preferences and Relative Prices constant, Measured at Sample Mean Level



The simulation analysis also shows that the average household with young children aged less than six (including households with zero child care expenditures) spends $500 if income is $30,000; $ 950 if income is $60,000; and $1,050 if income is $115,000.

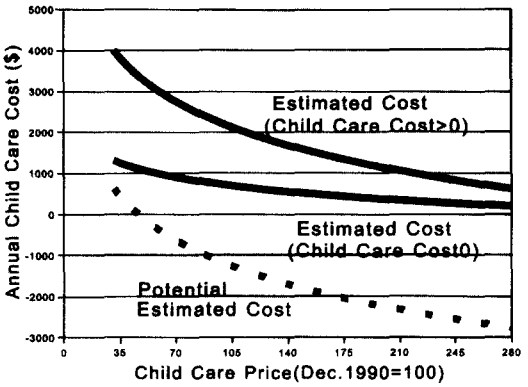
3. Price Effects and its simulation

This study attempts to incorporate price information into the model. Logarithmic price is only statistically significant at the 0.08 level. However, it is very important to see the price effect because one of the main purposes of this study was to calculate the price elasticities of the demand for child care. Average child care price during 1990-1992 in the United States was 117 when the 4th quarter of 1990 was coded as 100. The simulaiton of the price effect is illustrated in Figure 2. According to simulation result, if the child care price doubles (i. e, from a child care index of 100 to 200), the households' expected child care expenditures would decrease from $2,350 to $1,200 for households with positive child care expenditures, whereas for all households, the expenditures would decrease from $900 to $400. Thus, price effects are greater in higher expenditure groups, holding everything else constant. Also, the price effects are expected to be the greatest under the unconstrained optimization.

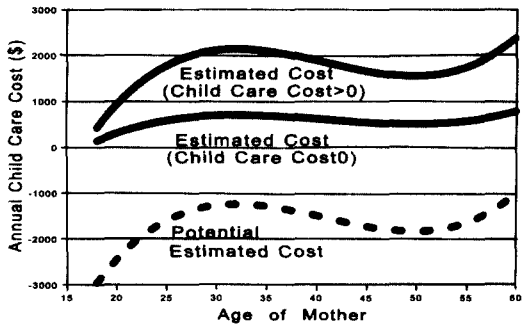
4. Mother's age, age squared, and age cubic effects and its simulation

〈Figure 3〉 shows curvilinear age effects on the dependent variable and their simulation results. Based on Tobit estimates there was a curvilinear age effect of mother's age on the potential, actual child care expenditure, holding income, prices and preferences are constant measured at sample mean level. Comparing to numerical interpretation of beta, researchers would possibly interpret trend or projection of expected value of actual and potential explained variable according as independent variables change. For example, at age of 34, estimated value of dependent variable shows the maximized value about 2000 dollar for samples of positive

〈Figure 2〉 Simulation Results Based on Tobit Estimates in Table 2 with respect to equations (1.9) and (1.10): Effect of Child Care Price on the Potential, Actual Child Care Expenditure, Holding Incomes and Preferences Constant, Measured at Sample Mean Level



〈Figure 3〉 Simulation Results Based on Tobit Estimates in Table 2 with respect to equations (1.9) and (1.10): Effect of Curvilinear Age on the Potential, Actual Child Care Expenditure, Holding Incomes and Preferences Constant, Measured at Sample Mean Level



child care cost only and about 800 dollar for samples with zero cost, respectively. Their predicted value of dependent variable for both samples decrease from mother's age 34 until age at 55 which may interpret that mother's age 55 households are likely to pay for others such as grandson or granddaughter and so forth.

Simply the beta value of Tobit shows the trend or likelihoold which is the bottom line is Figure 3 just like the other two previous figures so that researchers should not interpret the meaning of number of beta from SAS or SPSS Tobit output because its value only means the tendency or likelihood.

V. CONCLUSIONS

This study attempted to test three goals to construct statistical and econometric models based on Amemiya's type 1 Tobit, and to explore exogenous independent variable's effect on the explained variable with relations to illustrate its statistical marginal effects and finally to suggest a new approach of methodological issue using simulations in Tobit analysis. Simulation results were based on the results from the type I Tobit estimates in which the household's potential expenditures for child care were analyzed, and the estimated mean marginal effects for all households and for households with positive child care expenditures were compared, where sample means were incorporated as a meaning of holding other things being constant. Thus, if author would manipulate special value of independent variable interested or focused, the possible tobit simulation results can vary according to possible different assumptions.

In this paper, simulations were conducted for income, price of child care, and mother's age only because main goal of this study was mainly for showing a new approach to the illustration of statistical interpretation.

Results also indicated that prediction value of y can be possibly and easily traced by the independent variable changed whereas only beta value can

not illustrate its projection as independent variables' changes.

There would be possible results that child care is a normal good and a necessity for the households that already had positive expenses (N=365), while child care is a luxury good for all the households with young children (N=971) since the projection can derive such interpretations. The own price elasticity of quantity demanded for child care was -1.65 and -1.26 for both households with total samples (N=971) and positive expense groups (N=365), respectively because changed values of x and y axis can be derived from the simulation results. Once researchers know the trends and exact intercept or unit changed value, the elasticities of independent variables which are percentage changed of each variable can be easily calculated if we apply basic calculations to its simulation results.

Simulation of the statistically significant variables based on the Tobit result was the one of the purposes of this study. Thus, illustrating the estimated child care expenditures as an explained variable subject to both constrained and unconstrained optimization according to the changes of income, price, mother's age was unique in this study compared to other studies.

This study provides useful information to the consumer or family economist or family resource management educator, financial planner and counselor, and public policy maker in that a new approach to tobit's results using marginal effects and its simulation of expected value would be newly introduced in family and consumer field.

□ 접 수 일 : 2001년 9월 15일
□ 심사완료일 : 2001년 10월 29일

【REFERENCES】

Amemiya, T.(1984). Tobit models : a survey, *Journal of Econometrics*, 24, 3-61.

Amemiya, T.(1985). Advanced econometrics, Harvard University Press, Cambridge: MA

American Chamber of Commerce Researchers Association(1992). Cost of Living Index.

Green, W. H.(1991). LIMDEP, Version 6.0, Mainframe, Bellport, NY:Econometric Software Inc.

Green, W. H.(1990). Econometric analysis, New York, NY:Macmillan Publishing Co.

Maddala, G.S.(1983). Limited dependent and qualitative variables in econometrics, Cambridge: Cambridge University Press.

Maddala, G.S.(1992). Introduction to econometrics, Second Edition, NY: Macmillan Publishing Company.

Mok, C.J., Park, S., and Chern, W.S.(1995). Food consumption pattern of asian-american household: an empirical assessment of the household production theory, *Proceeding of the 1st Asian Consumer and Family Economics Association Meeting*,Vol.1, Malaysia, July, 12-24.

Park, S.Y.(1995). Child care services: two statistical / econometric approaches to household choice and demand, The Ohio State University Dissertation.

Park, S. and Mok, C.J.(1995). Childcare expenditures of households: tobit analyses for different family types, *Family Economics and Resource Management Biennial*, Vol. 1,193-198.

Park, S., Wasnich, W., and Fox, J.(1995). The impact of rising prices on out-of-pocket health care expenditures:1980-1991, *Family Economics and Resource Management Bien-*

nial, Vol. 1, 199-204.

Tobin, J.(1958). Estimation of relationships for limited dependent variables, *Econometrica*, 26, 24-36.

U.S. Bureau of the Census.(1995). Child care costs greater burden For the poor, *Census Bureau Reports, Current Population reports*, October, Washington, DC: Government Printing Office.