Journal of Economics
and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq

Comparison Between Nelson-Olson Method and Two-Stage Limited Dependent Variables (2SLDV) Method for the Estimation of a Simultaneous Equations System (Tobit Model)

Fulla Makee Ahmed
University of Baghdad/ Department of Missions and Cultural Relations
fullaamen@gmail.com

Ass.Prof.Dr. Eman Mohammed Abdullah
College of Administration and Economics/University of Baghdad
dr.eman20001999@ gmail.com

Received:27/7/2021 Accepted: 22/8/2021 Published: December / 2021

Abstract:

This study relates to the estimation of a simultaneous equations system for the Tobit model where the dependent variables (y_s') are limited, and this will affect the method to choose the good estimator. So, we will use new estimations methods different from the classical methods, which if used in such a case, will produce biased and inconsistent estimators which is (Nelson-Olson) method and Two- Stage limited dependent variables(2SLDV) method to get of estimators that hold characteristics the good estimator .

That is , parameters will be estimated for the limited variables and find the variance-covariance matrix for extracted estimators by the aforementioned two methods and then compare between the results of the two methods and find any better method by estimation and then finding the estimation efficiency, and this is what the study aims to .

A simultaneous equations system will be imposed for the limited model defined by two equations containing two endogenous variables one of complete observations and the other censored at zero.

The two methods were used to analyze the relationship between income and family expenditure on durable consumer goods , where the results showed that the performance of (Nelson-Olson) method is better than performing the Two-Stage limited dependent variables (2SLDV) method in obtain the lower values and all comparison measures as well as the results showed that income and expenditure one affects the other and the certificate obtained by the head of the family and the price affects the income and expenditure .

Keywords: Simultaneous Equations System , Tobit Regression Model , Nelson-Olson Method , Two-Stage Limited Dependent Variables Method(2SLDV)

1-Introduction:

The economy models lately received great attention especially limited regression models. This refers to the models that contain limited dependent variables and have several repeated observations within a certain borderline range.

And the Tobit model is among such models; it is one of the models that are used in statistical analyses for all fields whether economic medical or social that's why it's important to study this model it is the appropriate model in estimation for a simultaneous equations system involving limited dependent variables (y_s') and a set of independent variables (x_s') as well as in the case of truncated data or censored data .

Here will be mentioned some of research and studies on the Tobit model where in (1958) (James Tobin) [16] proposed the first model of the limited dependent variable now define the Tobit model and used repetitive methods to estimation the parameters of his model by using the initial values of the parameters and where in (1978) (Nelson F. , Olson L.) [15] interested in studying the simultaneous equations models with the limited dependent variables by using the Two-Stage (2SLDV) method in estimation and the resulting estimators were natural and consistent and the proposed model had new specifications and was more appropriate than Amemiya estimators and used simultaneous equations system for the limited model consisting of two equations in which one completely observations and the other is truncated. In (2018) [3](Ahmed N.) presented a study for Tobit models especially the censored and truncated regression models and the model was limited variables and used the maximum likelihood (ML) method in estimation and the resulting estimators were consistent and unbiased at the same time for the truncated and censored data and were compared with the multiple regression model by using the least squares method of estimation the itself data .

2- Simultaneous equations system

It is defined that it is a set interconnected processes of economic data that mathematically demonstrates contain on (dependent and independent) variables that affect one other simultaneously and that most economic phenomena can be represented by the form of a simultaneous equations system include overlapping relationships any relationships that affect and affect each other at the same time . (11) (2)

3-Limited Data and Truncated and Censored

1-3 Limited Data

The data are limited when there is a certain range within which the observations are located ,i.e. the observations are limited by smallest and largest value and vary depending on the nature of the phenomenon under study .

And as an example of the limited dependent variables is a study of consumer purchases on durable consumer goods and may vary the amount expend for the purchase and in sometimes no any process purchase is not made at all so then the expenditure record is zero (9) .

2-3 Truncated Data and Censored Data

Truncation is the exclusion of some of the vocabulary of the sample under study (i.e. is a case of lack in data). For example, when studying income are included people with low incomes, and we exclude people with high income from the sample. (3) (8)

and Truncated is the values of the dependent variable missing in the data at a particular point in the sample under study .

To clarify the truncation we will assumed that the y_i dependent variable observed is related to y_i^* the dependent variable unobserved as follows (12) **:**

$$y_i = \begin{cases} y_i^* & if \; y_i^* > 0 \end{cases} \quad , \quad i = 1, 2, \dots, n$$

As for the censored data, it is described as the values of the dependent variable (y) for a limited part of the observations. For example gathered the values of the dependent variable y around zero (meaning one value for y values is recorded with in a certain range) while the values of the independent variables (x_s')are viewed for all observations(i.e. full sample observations) (3)

and in the censored sample data are as follows (12) **:**

$$y_i = \begin{cases} (y_i^*, x_i) & if \; y_i^* > 0 \\ (0, x_i) & if \; y_i^* \le 0 \end{cases} \quad , \quad i = 1, 2, \dots, n$$

4- Tobit model for a Simultaneous equations

The Tobit model is one of the most important models used in statistical analysis of limited models whether economic or medical or in many other fields .

The first person to suggest this model is James Tobin (1958), where he clarified the relationship between income of the family and expenditure on durable consumer goods assuming there are several limited values in which the values of the approved variable vary (usually zero) .

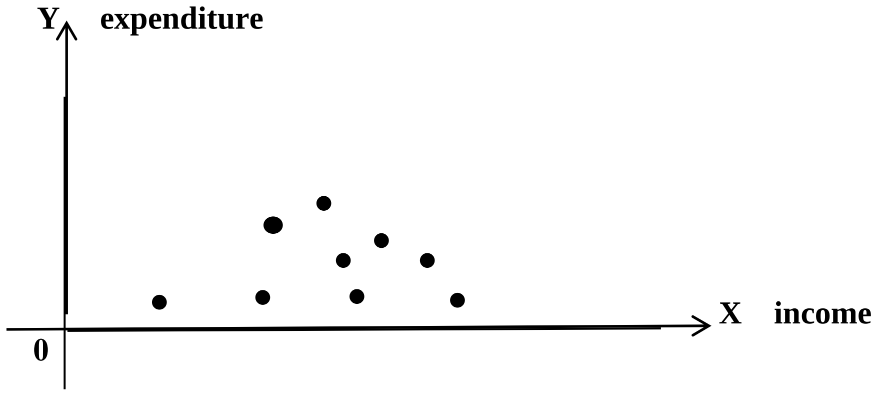
Tobin called his model the limited dependent regression model because the values of the dependent variables in the sample under test are not accessible to her fully but generalized by the name of Tobit model and this word was coined by Goldberger (1964) for the similarity between it and the Probit model, which is one of the censored regression models (6) (13) (14) (16)**.**

The figure below shows the relationship between income and expenditure on durable consumer goods for family and represents

X **: family income .**

Y **: family expenditure on durable consumer goods .**

It shows that the expenditure represented by the dependent variable data for the regression model used by Tobin does not take negative values (6) **.**



will be a simple explanation for the simultaneous equation for the Tobit model as follow :

$$y_i^* = \beta' x_i + \varepsilon_i \quad , \quad i = 1,2, \ldots, n \quad \ldots (1)$$

where

$$y_i = \begin{cases} y_i^* & if \ y_i^* > 0 \\ 0 & if \ y_i^* \le 0 \end{cases} \qquad i = 1,2, \ldots, n \qquad \ldots (2)$$

and

y_i^*: **un observed dependent variable dependent variable , :**y_i

x_i **: vector for independent variables**

β: **parameter vector**

ε_i**: error limit which is normal distributed with mean zero and variance** σ^2 **,** $\varepsilon_i \sim N(0, \sigma^2)$ **.**

5- Comparison Measures

We used the comparison measures clarified below because of their importance on the basis of which the methods of appreciation will be compared and the method with the least value for these measures is the best method in estimation .

1-5 Mean Square Error (MSE)

It is represents the difference between the real and estimation value of the dependent variable (y) divided by the number of those values .

and its formula as follow :

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}$$

and $y_i, \hat{y}_i$ **: they represent the real value and the estimation value of the dependent variable , respectively .**

n : sample size .

And when taking the square root of the mean square error will produce the measure (RMSE) it is more accurate in calculation and confirmation of the results (MSE) .

And its formula as follow :

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

2-5 Mean Absolute Error (MAE)

It is represents the sum of the absolute term of the difference between the real and estimation value of the dependent variable (y) divided by the number of those values .

And its formula as follow :

$$MAE = \frac{\sum_{i=1}^{n} | y_i - \hat{y}_i |}{n}$$

6- Methods of estimating for simultaneous equations system

That use of the estimation method for the parameters of a simultaneous equations system by the ordinary least squares method (OLS) is not feasible because the resulting estimators are biased and inconsistent. This contradicts the assumption of linearity which is based on the freedom of relationship between the values of the independent variables and the values of random boundaries; therefore, we resort to other methods and techniques of estimation [1] .

Before starting any methods, one must first identification each equation in the simultaneous equation system passing through the order condition then the rank condition so that we can get the appropriate method for estimating the parameters of the system variables .

1-6 Identification

It is the process of testing each equation separately from the equations that form the simultaneous equations system, [1] and we mean by that is it possible to obtain estimates of the structural model parameters estimated from the parameters of the reduced model and these estimates are unique and bear qualities the good estimator the represented by (unbiased , consistent) .

By the identification process the appropriate method is chosen to estimate the parameters of the variables of simultaneous equation system, and we can determine each equation from equations system if he exact identify or unidentify [11] and the identification includes two condition (order and rank) .

1- Order condition

This condition is considered important and not sufficient to complete the identification process for each equation in a simultaneous equations system; therefore, consideration should be given to the condition that follows it to complete the identification process . This condition is based on the assumption [2]:

W : represents the number of variables in the equation under test .

N : represents the number of total variables (exogenous , endogenous and time regress variables) of the simultaneous equations system under test .

L : represents number of total equations in the simultaneous equations system under test .

- if N-W=L-1 the equation is exact identified .

- if N-W> L-1 the equation is over identified .

- Otherwise the equation is unidentified , that is, the parameters cannot be estimated .

2- Rank Condition

This condition is complementary and confirmed to the previous condition to identification the equation, and to complete this condition all structural parameters must be arranged in relation to all variables in the system, and we take the parameters corresponding to the missing parameters in the equation under test and represent it as matrix and it divides into several sub-matrices from Rank (L-1) if the matrix resulting from the structural parameters is not square, then we find the determinant if the determinant is equal to zero then the equation at that time is unidentified either if the determinant for it is not equal to zero the equation is identified [(1)] .

After the two previous conditions, the identification process will be over and on light of this will be chosen appropriate methods to estimate the simultaneous equations system.

2-6 Estimation by Nelson-Olson method for Tobit model

Nelson - Olson (1977) presented this method of estimation through which they generalized the Tobit model presented by Tobin (1958) for the simultaneous equation and from that obtained the consistent estimators .

This method will be detailed by a model consisting of two equations that make up the system and as explained below, which were expressed by Nelson - Olson (1977) for the Tobit model [(5)].

$$y_{1i} = \alpha_1 \; y_{2i}^{*} + \beta_1' x_{1i} + u_{1i} = \alpha_1 \; y_{2i}^{*} + x_{1i}' \beta_1 + u_{1i} \quad \ldots.(3)$$

$$y_{2i}^{*} = \alpha_2 y_{1i} + \beta_2' x_{2i} + u_{2i} = \alpha_2 \; y_{1i} + x_{2i}' \beta_2 + u_{2i} \quad \ldots.(4)$$

$$y_{2i} = \begin{cases} y_{2i}^{*} & if \;\; y_{2i}^{*} > 0 \\ 0 & if \qquad {}^{o}/_{w} \end{cases} \qquad i = 1, 2, \ldots, n \qquad \ldots.(5)$$

Where

y_{1i}, y_{2i} : vector for endogenous observed variables

y_{2i}^{*} : vector for endogenous un observed variables

α_2 , α_1 : model constants (unknown)

x_{1i}', x_{2i}' : Row vectors from rank $(1\times k_1)$, $(1 \times k_2)$ for the independent variables for equations (3) and (4) .

β_1 , β_2 : a column vectors from rank $(k_1 \times 1)$, $(k_2 \times 1)$ of the unknown parameters for equations (3) and (4) .

And $k = k_1 + k_2$

u_{1i}, u_{2i} : The random error with mean zero and variance σ^2 and which is distributed normally

The equations system can be expressed by the following equations :

$$y_1 = \alpha_1 y_2^{*} + X_1 \beta_1 + u_1 \ldots..(6)$$

$$y_2^{*} = \alpha_2 y_1 + X_2 \beta_2 + u_2 \ldots..(7)$$

$$y_2 = \begin{cases} y_2^{*} & if \; y_2^{*} > 0 \\ 0 & {}^{o}/_{w} \end{cases}$$

The reduced form of two equations (6) and (7) can be obtained by compensation the equation (7) by equation (6) we get :

$$y_1 = X\pi_1 + V_1 \ldots..(8)$$

It is through compensation the equation (6) by equation (7) we get :

$$y_2^* = X\pi_2 + V_2 \ldots..(9)$$

Where
X : Matrix from degree ($n \times k$) which contains different columns for [X_1 , X_2]
π_1 , π_2 : vectors from rank ($k \times 1$) for the parameters of equations (8) and (9)
V_1 , V_2 : Vectors from rank ($n \times 1$) of the error limit for the equations (8) and (9) .

And

$$V_{1i} = \frac{\sigma_{12}}{\sigma_2^2} V_{2i} + e_i \quad , \quad i = 1, 2, \ldots\ldots, n \qquad \ldots.(10)$$

Where e_i independent of V_{2i}

And to find the Nelson-Olson estimators for the two parameters vectors γ_1 , γ_2 and to find the asymptotic variance and covariance matrix, the following two points must be followed [5]:
We will define γ_1 and γ_2 as follows:

$$\gamma_1' = (\alpha_1, \beta_{11}, \beta_{12}, \ldots., \beta_{1k_1}) = (\alpha_1, \beta_1')$$
$$\gamma_2' = (\alpha_2, \beta_{21}, \beta_{22}, \ldots., \beta_{2k_2}) = (\alpha_2, \beta_2')$$

First: The parameters are estimated for the reduced form of the Tobit π_1 *and* π_2 model defined by equations (8) and (9) .
Since it is estimated π_1 in equation (8) using the least squares method as follows:

$$\hat{\pi}_1 = (X'X)^{-1}X'y_1 = \pi_1 + (X'X)^{-1}X'V_1 \quad \ldots\ldots(11)$$

As for π_2 it is estimated by maximizing the likelihood function of the Tobit model to obtain π_2 and σ_2^2 which are as follows : censored

$$L(\pi_2, \sigma_2^2) = \prod_{i=1}^{N_\circ} [1 - F(\pi_2' x_i, \sigma_2^2)] \cdot \prod_{i=1}^{N_1} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(y_{2i} - \pi_2' x_i)^2}{2\sigma_2^2}} \ldots..(12)$$

Where
$N_\circ$: a set of observations in which y_{2i} is equal zero or less than zero .
N_1: the set of observations in which y_{2i} has positive values .
And

$$Fi = F(\pi_2' x_i, \sigma_2^2) = \int_{-\infty}^{\pi_2' x_i} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}(\frac{\lambda}{\sigma_2})^2} d\lambda \quad \ldots..(13)$$

$$f_i = \frac{1}{\sqrt{2\pi\sigma_2^2}} \, e^{-\frac{1}{2}\left(\frac{\pi_2' X_i}{\sigma_2}\right)^2} \quad \ldots..(14)$$

<u>Second</u>: by depending on equations (3) and (9), we obtain the following:

$$y_1 = \alpha_1(X\pi_2 + V_2) + X_1\beta_1 + u_1$$

$$y_1 = X\hat{S}\gamma_1 + H_1 \quad \ldots..(15)$$

Where

J_1: a matrix of degree ($k \times k_1$) with elements of one and zero numbers

$$X_1 = X . J_1 \quad \text{And}$$

$$H_1 = V_1 - X(\hat{\pi}_2 - \pi_2)\alpha_1 \quad \text{و} \quad \hat{S} = [\hat{\pi}_2 : J_1]$$

And To find the Nelson-Olson estimators for the parameter vector γ_1 in equation (15) we use the least squares method as follows [(5)]:

$$\hat{\gamma}_1 = \left(\hat{S}'X'X\hat{S}\right)^{-1}\hat{S}'X'y_1 \ldots..(16)$$

The Nelson-Olson estimators of the parameters vector γ_1 defined by equation (16) can be written as follows :

$$\hat{\gamma}_1 = \gamma_1 + \left(\hat{S}'X'X\hat{S}\right)^{-1}\hat{S}X'H_1 \quad \ldots\ldots(17)$$

From equation (17) one can obtain the asymptotic variance and covariance matrix for the $\hat{\gamma}_1$ vector as follows:

$$\text{V-COV}(\hat{\gamma}_1) = \left(\hat{S}'X'X\hat{S}\right)^{-1} \hat{S}'X' \ \text{V-COV}(H_1) \ X\hat{S}\left(\hat{S}'X'X\hat{S}\right)^{-1} \ldots\ldots(18)$$

$$\text{V-COV}(H_1) = \text{V-COV}(V_1) + \alpha_1 X \, \text{V-COV}(\hat{\pi}_2) + 0 - 2\alpha_1 \text{COV}(V_1, X\hat{\pi}_2) \quad \ldots\ldots(19)$$

It is noted that the formula (19) depends on the variance and covariance matrix of the vector ($\hat{\pi}_2$) which can be found after finding the estimators ($\hat{\pi}_2$) as follows:

by depending on equations (4) and (8), we obtain the following :

$$y_2^* = \alpha_2(X\pi_1 + V_1) + X_2\beta_2 + u_2$$

$$y_2^* = X\hat{\pi}_1\alpha_2 + X_2\beta_2 + V_2 - \alpha_2 X(\hat{\pi}_1 - \pi_1)$$

$$y_2^* = XD\gamma_2 + H_2 \quad \ldots\ldots(20)$$

J_2: a matrix of degree ($k \times k_2$) with elements of one and zero numbers

Where $X_2 = X . J_2$

and

$$H_2 = V_2 - \alpha_2 X(\hat{\pi}_1 - \pi_1) \quad , \quad D = [\hat{\pi}_1 : J_2]$$

It is noted from equation (20) that a detailed formula for equation (9) can be written considering that :

$\pi_2 = D\,\gamma_2$ ….. (21)

The defined likelihood function can be written as formula (12) after compensating the equation (21) in it as follows:

$L\left(\pi_2, \sigma_2^2\right) = L\left(\pi_1, \gamma_2, \sigma_2^2\right)$ …..(22)

And when you compensating $\hat{\pi}_1$ knowledge of equation (11) with the likelihood function (22), we get the following :

$L\left(\hat{\pi}_1, \gamma_2, \sigma_2^2\right)$ …..(23)

And depending on the likelihood function (23) the parameters vector γ_2 and the parameter σ_2^2 can be estimated as the two researchers Nelson-Olson were able to find their estimators by maximizing the likelihood function (23).

And to find the asymptotic variance and covariance matrix for the vector of estimators $\hat{\gamma}_1$ defined by equation (18) it is necessary to obtain the asymptotic distribution of the estimators vector $\hat{\pi}_2$ for the purpose of obtaining the asymptotic variance and covariance matrix as shown in formula (19).

Amemiya (1979) explain that it is:

$$\hat{\pi}_2 - \pi_2 \triangleq (I\ ,0\ )[-E\frac{\partial^2 LogL}{\partial\theta^*{}_1\partial\theta_1^{*\prime}}]^{-1}\frac{\partial LogL}{\partial\theta_1^*} \quad \text{….. (24)}$$

Where

$\theta_1^{*\prime}$: Parameters vector ($\pi'_2\ , \sigma_2^2$).

$(I\ ,0\ )$: a matrix that includes a identity matrix and a column vector whose elements of zeroes .

A: means that both sides of formula (24) have asymmetric asymptotic distribution .

Therefore, the variance and covariance matrix of the vector estimators $\hat{\pi}_2$ is given by the following equation :

$$V - COV(\hat{\pi}_2) = [I\ \ \vdots\ \ 0]\left[-E\frac{\partial^2 LogL}{\partial\theta_1^*\partial\theta_1^{*\prime}}\right]^{-1}\begin{bmatrix} I \\ \dots \\ 0 \end{bmatrix} = V_\circ \quad \text{…...(25)}$$

Where

$V_\circ$: asymptotic variance and covariance matrix for estimators vector $(\hat{\pi}_2)$

And compensating equation (25) into formula (19) we get the following:

$$V - COV(H_1) = \sigma_1^2 I - \alpha_1^2 X[I\ \vdots\ 0]\left[E\frac{\partial^2 LogL}{\partial\theta_1^*\partial\theta_1'}\right]^{-1}\begin{bmatrix} I \\ \dots \\ 0 \end{bmatrix}X' +$$

$$\alpha_1 X[\,I\ \ \vdots\ \ 0]\ \left[\ E\frac{\partial^2 LogL}{\partial\theta_1^*\partial\theta_1^{*\prime}}\ \right]^{-1} E\left[\frac{\partial LogL}{\partial\theta_1^*}.V_1'\right] +$$

$$\alpha_1 E\left[V_1\frac{\partial LogL}{\partial\theta_1^{*\prime}}\right]\left[E\frac{\partial^2 LogL}{\partial\theta_1^*\partial\theta_1^{*\prime}}\right]^{-1}\begin{bmatrix} I \\ \dots \\ 0 \end{bmatrix}X' \quad \text{…..(26)}$$

Depending on equation (26) can find the asymptotic variance and covariance matrix for the estimators vector $\hat{\gamma}_1$ by compensating it with equation (18).

And by taking the log of the likelihood function defined by formula (12) as follows:

$$Log\ L = \sum_{i=1}^{N\circ} Log(1 - F_i) - \frac{N_1}{2} Log(2\pi) - \frac{N_1}{2} LogL\ \sigma_2^2 - \frac{1}{2\sigma_2^2}\sum_{i=1}^{N_1}(y_{2i} - \pi_2'\ X_i)^2 \quad \ldots\ldots(27)$$

To maximize the function (27), we take the partial derivatives with respect to the vector of the parameters π_2 and the parameter σ_2^2 to obtain the estimates for π_2 and σ_2^2 as follows [4] :

$$\frac{\partial LogL}{\partial \pi_2} = -\sum_{i=1}^{N\circ} \frac{f_i X_i}{1 - F_i} + \frac{1}{\sigma_2^2}\sum_{i=1}^{N_1}(y_{2i} - \pi_2' X_i)\ X_i \ldots..(28)$$

$$\frac{\partial LogL}{\partial \sigma_2^2} = \frac{1}{2\sigma_2^2}\sum_{i=1}^{N\circ} \frac{\pi_2' X_i f_i}{1 - F_i} - \frac{N_1}{2\sigma_2^2} + \frac{1}{2\sigma_2^4}\sum_{i=1}^{N_1}(y_{2i} - \pi_2' X_i)^2 \ldots\ldots(29)$$

we find second derivative for equations (28) we obtain : And

$$\frac{\partial^2 LogL}{\partial \pi_2 \partial \pi_2'} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} a_{11}^*(i) X_i\ X_i' \ldots..(30)$$

[5] : **Where**

$$a_{11}^*\ (\ i) = \frac{1}{\sigma_2^2}\left(\pi_2' X_i f_i - \frac{\sigma_2^2 f_i^2}{1 - F_i} - F_i\right) \ldots..(31)$$

And we find second derivative for equations (29) we obtain :

$$\frac{\partial^2 LogL}{\partial\ \pi_2 \partial \sigma_2^2} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} a_{12}^*(i)\ X_i \ \ldots..(32)$$

Where [5]**:**

$$a_{12}^*(i) = -\frac{1}{2\sigma_2^4}\left[\ (\pi_2' X_i)^2 f_i + \sigma_2^2 f_i - \frac{\sigma_2^2 \pi_2' X_i f_i^2}{1 - F_i}\right] \ldots..(33)$$

And

$$\frac{\partial LogL}{\partial (\sigma_2^2)^2} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} a_{22}^*\ (i) \ldots..(34)$$

Where [5] **:**

$$a_{22}^*\ (i) = \frac{1}{4\sigma_2^4}\left[\frac{(\pi_2' X_i)^3 f_i}{\sigma_2^2} + \pi_2' X_i f_i - \frac{(\pi_2' X_i f_i)^2}{1 - F_i} - 2F_i\right] \ldots..(35)$$

$$E\,\frac{\partial^2 LogL}{\partial\theta_1^*\partial\theta_1^{*\prime}} = \begin{bmatrix} X' & O' \\ o' & 1' \end{bmatrix} A^* \begin{bmatrix} X & o \\ O & 1 \end{bmatrix} \quad \ldots\ldots(36)$$
$$= \underline{X}' A^* \ \underline{X}$$

: Where

A^*: a diagonal matrix whose diagonal elements are equal to $a_{ij}^*(i)$

O : matrix of degree $(n \times k)$ its elements of zeroes.

o : column vector of zeroes and 1: column vector of ones .

Can be obtained the vector of the estimators

$\theta_2^{*\prime} = \left(\gamma'_2, \sigma_2^2\right)$ depend on the likelihood function defined by the formula (23) to be $L\ (\hat{\pi}_1, \theta_2^*\ )$which is maximized to obtain the estimators of the parameters vector θ_2^* and then the asymptotic variance and covariance matrix can be obtained with the same steps that went through previously or it can be written as follows (5) :

$$\hat{\theta}_2^* - \theta_2^* = -\left[E\frac{\partial^2 LogL}{\partial\theta_2^*\partial\theta_2^{*\prime}}\right]^{-1}\left[\frac{\partial LogL}{\partial\theta_2^*} + E\frac{\partial^2 LogL}{\partial\theta_2^*\partial\pi_1'}.(\hat{\pi}_1 - \pi_1)\right] \ldots\ldots(37)$$

Thus,

$$\begin{aligned} V - COV(\hat{\theta}_2^*) &= \left[E\ \frac{\partial^2 LogL}{\partial\theta_2^*\partial\theta_2^{*\prime}}\right]^{-1}\Big(-E\ \frac{\partial^2 LogL}{\partial\theta_2^*\partial\theta_2^{*\prime}} \\ &+ \sigma_1^2 E\frac{\partial^2 LogL}{\partial\theta_2^*\partial\pi_1'}(X'\mathrm{X})^{-1}E\frac{\partial^2 LogL}{\partial\pi_1\partial\theta_2^{*\prime}} \\ &+ E\frac{\partial^2 LogL}{\partial\theta_2^*\partial\pi_1'}(X'\mathrm{X})^{-1}X'E\left[V_1\frac{\partial LogL}{\partial\theta_2^{*\prime}}\right] \\ &+ E\left[\frac{\partial LogL}{\partial\theta_2^*}V_1'\right]X(X'\mathrm{X})^{-1}E\frac{\partial^2 LogL}{\partial\pi_1\partial\theta_2^{*\prime}}\Big)\left[E\frac{\partial^2 LogL}{\partial\theta_2^*\partial\theta_2^{*\prime}}\right]^{-1} \ldots.(38) \end{aligned}$$

From equation (21) can infer of relationship between θ_1^* and θ_2^* as follows :

$$\begin{bmatrix} \pi_2 \\ \sigma_2^2 \end{bmatrix} = \begin{bmatrix} D\,\gamma_2 \\ \sigma_2^2 \end{bmatrix}$$
$$\begin{bmatrix} \pi_2 \\ \sigma_2^2 \end{bmatrix} = \begin{bmatrix} D & o \\ o' & 1 \end{bmatrix}\begin{bmatrix} \gamma_2 \\ \sigma_2^2 \end{bmatrix}$$

$$\theta_1^* = \underline{D}\theta_2^* \quad \ldots\ldots(39)$$

We have :

$$\frac{\partial LogL}{\partial\theta_2^*} = \underline{D'}\frac{\partial LogL}{\partial\theta_1^*} \quad \ldots\ldots(40)$$

And also

$$= \underline{D'}\,E\frac{\partial^2 LogL}{\partial\theta_1^*\partial\theta_1^{*\prime}}\underline{D} \quad \ldots\ldots(41)\ \ E\frac{\partial^2 LogL}{\partial\theta_2^*\partial\theta_2^{*\prime}}$$

And depended on equation (36) we get the following :

$$E\,\frac{\partial^2 LogL}{\partial\theta_2^*\partial\theta_2^{*\prime}} = \underline{D'}\ \underline{X'}\ A^*\underline{X}\ \underline{D}\ \ldots\ldots(42)$$

As for the quantity $E\left[\frac{\partial LogL}{\partial\theta_2^*}V_1'\right]$ It shall be as follows :

$$E\left[\frac{\partial LogL}{\partial\theta_2^*}V_1'\right] = \underline{D'}\ E\left[\frac{\partial LogL}{\partial\,\theta_1^*}V_1'\right]\ \ldots\ldots(43)$$

Where

$$E\left[\frac{\partial LogL}{\partial\,\theta_1^*}V_1'\right] = -\sigma_{12}\,\underline{X'}\,A^*\begin{bmatrix} I \\ 0 \end{bmatrix}$$

And upon it , the equation (43) after compensation becomes as follows:

$$E\left[\frac{\partial LogL}{\partial\theta_2^*}V_1'\right] = -\sigma_{12}\,\underline{D'}\ \underline{X'}\ A^*\begin{bmatrix} I \\ 0 \end{bmatrix}\ \ldots\ldots(44)$$

Finally, when deriving the two formulas $\frac{\partial LogL}{\partial\pi_2}$ and $\frac{\partial LogL}{\partial\sigma_2^2}$ defined in equation (27) with respect to the parameters vector π_1 we obtain the following :

$$E\,\frac{\partial^2 LogL}{\partial\theta_2^*\partial\pi_1'} = \underline{D'}\,\frac{\partial LogL}{\partial\theta_1^*\partial\pi_1'}\ \ldots\ldots(45)$$

Where

$$E\,\frac{\partial LogL}{\partial\theta_1^*\,\partial\pi_1'} = \alpha_2\underline{X'}\ A^*\begin{bmatrix} I \\ 0 \end{bmatrix}\ X\ \ldots\ldots(46)$$

And by compensation equation (46) into equation (45) :

$$E\,\frac{\partial^2 LogL}{\partial\theta_2^*\partial\pi_1'} = \alpha_2\underline{D'}\,\underline{X'}\,A^*\begin{bmatrix} I \\ 0 \end{bmatrix}\ X\ \ldots\ldots(47)$$

Thus, we obtain the following asymptotic variance and covariance matrix [5]:

$$\begin{aligned} &V - COV\left(\hat{\theta}_2^*\right) \\ &= \left(\underline{D'}\,\underline{X'}\,A^*\,\underline{X}\,D\right)^{-1}\Big[-\underline{D'}\ \ \underline{X'}\ A^*\,\underline{X}\,\underline{D} \\ &+ \left(\alpha_2^2\sigma_1^2\right. \\ &\left. - 2\alpha_2\sigma_{12}\right)\underline{D'}\ \ \underline{X'}\ A^*\begin{bmatrix} I \\ 0 \end{bmatrix}\ X\,(X'X)^{-1}X'(I\ \ 0)\,A^*\,\underline{X}\,\underline{D}\Big]\left(\underline{D'}\,\underline{X'}\,A^*\ \underline{X}\,\underline{D}\right)^{-1}\ \ldots\ldots(48) \end{aligned}$$

Amemiya (1979) proved the final form of the asymptotic variance and covariance matrix for the estimators vector $\hat{\gamma}_2$ as follow :

$$V - COV(\hat{\gamma}_2) = \begin{bmatrix} I \\ 0 \end{bmatrix} V - COV\left(\hat{\theta}_2^*\right)(I\quad 0)\ \ldots\ldots(49)$$

3-6 Estimation by Two-stage limited dependent variables(2SLDV) method

This method used for the simultaneous equations system is exact identified, and this method is considered one of the important and widely used methods in the economic field especially for regression models that include limited dependent variables (y' s) and for censored data and for easily of its use and simplicity of calculating where the estimators resulting from this method are unbiased and consistent and thus have become as popular no less than to use of Two- stage least square method.

This method was developed before Nelson-Olson (1978) [5] [10].

The simultaneous equations system can be expressed by (3) , (4) , (5) equations the aforementioned in the first method .

The steps of this method will be on two stages; therefore, it is called by Two-Stage [7] [15].

1- The first stage will include determining the endogenous variable for the equation whose parameters will be estimated which was the result of it exact identified .

2- And then the reduced formula for the endogenous variable is found mentioned in the previous point .

3- Using the ordinary least squares method to estimate π_1 and estimate π_2 by the maximum likelihood, and we can find the estimated values of the endogenous variables of the system and then move on to the second stage which includes :

4- The use of the maximum likelihood method again to estimate the parameters of the structural equations of the system. This is done after replacing the estimated values ($\hat{y}$) the mentioned in the first stage which replaced the real values of endogenous variables y_s' in the structural equations.

The reduced form of the model is represented by the two equations (8) and (9) following:

$$y_1 = X\pi_1 + V_1$$
$$y_2^* = X\pi_2 + V_2$$

And π_1 estimated by ordinary least square method as follows :

$$\hat{\pi}_1 = (X'X)^{-1}X'y_1 \;\ldots\ldots (50)$$

The estimates for the reduced form will be as follows [15]:

$$\hat{y}_{1i} = \hat{\pi}'_1 x_i \quad \ldots\ldots(51)$$
$$\hat{y}^*_{2i} = \hat{\pi}'_2 x_i \quad \ldots\ldots(52)$$

Therefore, the asymptotic variance and covariance matrix for $\hat{\gamma}_1$ will be as follows [5] :

$$V - COV(\hat{\gamma}_1) = [S'X'\,V - COV(H_1)^{-1}XS]^{-1} \quad \ldots\ldots(53)$$

And the asymptotic variance and covariance matrix for $\hat{\gamma}_2$ will be as follows [5]:

$$V - COV(\hat{\gamma}_2) = \left(D'V_\circ^{-1}D\right)^{-1}D'\,V_\circ^{-1}V_2\,V_\circ^{-1}D\left(D'V_\circ^{-1}D\right)^{-1} \ldots\ldots(54)$$

And

$$V_2 = \mathrm{C}\,(X'\mathrm{X})^{-1} + V_\circ \quad \ldots\ldots(55)$$

$$C = \alpha_2^2\sigma_1^2 - 2\alpha_2\sigma_{12}$$

7- Application

For the purpose of studying estimation methods as a practical application, the data was taken from the central agency for statistics from the economic and social survey of the family of the year (2012) and the size of sample (n=250) family it included wages and salaries income to study income and expenditure on durable consumer goods for the Iraqi family and all the governorates of Iraq (18) governorates and the most important factors affecting them .

And it was applied the Nelson-Olson method and the Two-stage limited dependent variables method (2SLDV) to estimate the parameters of the Tobit system of a simultaneous equations, and thus estimate the entire system .

The model is determinant variables (dependent variables) because both income and expenditure have a minimum and a higher limit where the equations system was built according to Keynesian theory to analyze the relationship between income and expenditure on consumer durables good and the independent variables (x_1 , x_2) were chosen on an economic basis as they are the most variables affecting income and expenditure .

Was expressed for the simultaneous equations system previously with equations (3) ,(4) ,(5) .

8- Results

We estimate parameters of a simultaneous equations system for Tobit model and the finding of the variance-covariance matrix of the estimators extracted by Nelson-Olson method and Two-stage limited dependent variables method (2SLDV) and the two methods will be compared according to the comparison measures (mean square error (MSE). Root mean square error (RMSE) , mean Absolut error (MAE) .

The results were as follows :

Parameters Estimation :

Nelson - Olson Method	α_1	1.625086	β_1	5.804347
	α_2	0.612346	β_2	0.028352
Two- Stage Method	α_1	0.026136	β_1	1093.567
	α_2	3.044963	β_2	332.1376

$$\hat{y}_1 = 1.62\, y^*_2 + 5.80\, x_1$$
$$\hat{y}^*_2 = 0.61\, y_1 + 0.02\, x_2$$

From the first equation, it turns out that the impact of expenditure on income is 1.62 i.e. increasing family income by one unit leads to an increase in expenditure by 1.62 and the amount of impact of the head of family certificate on income is 5.80 i.e. the certificate has an impact on income .

From the second equation, it turns out that the effect of income on expenditure is 0.61 and that effect of price on expenditure is 0.02 .

**v_cov (γ_1) = [2.50665374926432e-06 0.0624310019881790
0.0624310019881790 1920.81557297558]**

**v_ cov (γ_2) = [0.285682285299109 473.514293187380
473.514293187380 814030.100658778]**

$$\hat{y}_1 = 0.02\ y_2^* + 1093.56\, x_1$$

$$\hat{y}_2^* = 3.04\, y_1 + 332.13\, x_2$$

From the first equation, it turns out that the amount of 0.02 of expenditure affects the income and the amount of 1093.56 the certificate of the head of the family affects the income .

From the second equation, it turns out that the amount of 3.04 of income affects the expenditure and the amount 332.13 from the price affects the expenditure .

**v_ cov (γ_1) = [113785.190030593 613993.922320955
613993.922320955 3437168.45207971]**

**v_cov (γ_2) = [16235.0259684979 142826.806753870
142826.806753870 1303238.85959866]**

For Comparison Measures:

methods	MSE	RMSE	MAE
Nelson-Olson	0.000232	0.015231	0.001983
Two-Stage	0.00547	0.073959	0.00605

The results proved that Nelson-Olson method is better than Two-stage limited dependent variables 2SLDV method by obtaining the lowest values and for all comparison measures .

As for the estimation efficiency that proved Nelson-Olson method efficient relative to the Two-stage limited dependent variables 2SLDV method and the results were as follows :

Eff (γ_1) = 2.202970131e-11

Eff (γ_2) = 1.759666328e-05

9- Conclusions

Was used Nelson-Olson method and Two-stage limited dependent variables (2SLDV) method to analyze the relationship between income and expenditure on durable consumer goods where the results showed that (Nelson-Olson) method is better than Two-Stage limited dependent variables (2SLDV) method in obtaining the lower values and for all comparison measures as well as the results showed that income and expenditure one affects the other i.e. the increase in income by one unit leads to an increase in expenditure, which corresponds to economic theory and that the independent variable x_1 represented by the certificate of the head of family affects the dependent variable y_1 represented family income of 5.80 i.e. the certificate affects income and the independent variable x_2 represented by the price affects the dependent variable y_2^* of represented by expenditure of 0.02 according to the study data.

References:

1 - Kazem A. Dr. Amory Hadi, Basem Shlebeh Muslim, (2002), “Advanced Economic Measurement - Theory and Practice,” Baghdad, Al-Taif Press.

2- Yahya Muzahim Muhammad, Mahmoud Hamdoun Abdullah, (2009), “Using the two methods 2SLS , 3SLS in estimation of the simultaneous equations system for International grain prices for the period (1961-2001)”, research from Al-Rafidain Development Journal,
Issue 93, Volume 31, College of Administration and Economics, University of Mosul .

3- Ahmed , Dr. Nawzad M. , (2018),"Limited Dependent Variable Modeling (Truncated and Censored Regression models) with Application",PP.82-96 .

4- Amemiya , T.(1973), " Regression analysis when the dependent variable is truncated normal ",Vol.14, PP.997-1016.

5-Amemiya , T.(1979), "The estimation of a simultaneous equation tobit model ", International Economic Review,Vol.20,PP.169-181.

6- Amemiya T., (1984) ," Tobit model : A survey",Econometrics Journal, Vol 24,PP. 3-61.

7- Amemiya T.,(1985),"Advanced Econometrics ", Book, PP.397-398 .

8- CAMERON,A. Colin , TRIVEDI, K. PRAVIN,(2005), "Microeconometrics Methods and Applications", Book , chapter 16 ,PP.529.

9- Cragg , John G., (1971) , " some statistical models for Limited dependent variables with application to the demand for durable goods", Vol.39,pp829.

10-Fei Lee L, Maddala G. and TrostR,(1980),"Asymptotic Covariance matrices of Two-Stage Probit and Two-Stage Tobit methods for Simultaneous equations models with Selectivity",Vol.48,PP.491-503.

11- -Fomby ,Thomas B., Hill ,R. Carter and Johnson Stanley R. ,(1988),"Advanced Econometric Methods", Book , PP.439 , Chapter 25, PP.555-561.

12- https://www.cambridge.org /core , Chapter 8, Truncated and Censored Data , (2010),University of Southern California .

13- KOC Seyma. , SAHIN. Mustafa,(2018) ,"Tobit Model and an Application",PP.73-80.

14- McDonald , john F., Moffitt, Robert A. ,(1980) ,"The uses of tobit analysis " the Review of economics and statistics ,Vol . 26, pp. 318-321.

15- Nelson F. , Olson L.,(1978), "Specification and Estimation of A Simultaneous Equation Model With Limited Dependent Variables",Vol.19,PP.695-709.

16 - Tobin , J. , (1958) , "Estimation of Relationships for Limited Dependent Variables", Econometrica,26 ,PP. 24-36.

Journal of Economics and Administrative Sciences Vol.27 (NO. 130) 2021, pp. 210-226

مقارنة بين طريقة Nelson-Olson وطريقة ذات المرحلتين للمتغيرات المعتمدة المحددة (2SLDV) لتقدير منظومة المعادلات الانية لنموذج (Tobit)

الباحث/ فلة مكي أحمد
جامعة بغداد / قسم البعثات والعلاقات الثقافيه
fullaamen@gmail.com

ا. م. د. ايمان محمد عبد الله
كلية الادارة والاقتصاد/جامعة بغداد
Dr.eman20001999@gmail.com

Received:27/7/2021 Accepted: 22/8/2021 Published: December / 2021

مستخلص البحث:

تتعلق هذه الدراسة بتقدير منظومة المعادلات الانية لانموذج Tobit حيث المتغيرات المعتمدة ($Y's$) محددة وهذا سيؤثر على طريقة اختيار المقدر الجيد لذلك سنستعمل طرق بديلة في التقدير تختلف عن الطرق التقليدية التي اذا استخدمت في مثل هذه الحالة ستنتج مقدرات متحيزة (biased) وغير متسقة (inconsistent) وهما طريقة (Nelson-Olson) وطريقة ذات المرحلتين للمتغيرات المعتمدة المحددة (2SLDV) لنحصل على مقدرات تحمل صفات المقدر الجيد .

اي سيتم تقدير المعلمات للمتغيرات المحددة وايجاد مصفوفة التباين والتباين المشترك للمقدرات المستخرجة بالطريقتين السابقة الذكر ثم المقارنة بين نتائج الطريقتين وايجاد اي طريقة أفضل في التقدير ثم ايجاد كفاءة التقدير وهذا ماتهدف اليه الدراسة .

وقد تم فرض منظومة معادلات انية للانموذج المحدد بمعادلتين تتضمن متغيرين داخليين احدهما كامل المشاهدات والاخر خاضع للرقابة (Censored) عند الصفر .

حيث وظفت الطريقتين لتحليل العلاقة بين الدخل والانفاق للاسرة على السلع الاستهلاكية المعمرة حيث اظهرت النتائج ان اداء طريقة (Nelson-Olson) افضل من اداء طريقة ذات المرحلتين للمتغيرات المعتمدة المحددة (2SLDV) وذلك في الحصول على اقل القيم ولمقاييس المفاضلة كافة وكذلك اظهرت النتائج ان الدخل والانفاق يؤثر ويتأثر احدهما بالاخر وإن الشهادة الحاصل عليها رب الاسرة والسعر يؤثران على الدخل والانفاق .

نوع البحث : ورقة بحثية

المصطلحات الرئيسة للبحث: منظومة المعادلات الانية , نموذج انحدار Tobit , طريقة Nelson-Olson, طريقة ذات المرحلتين للمتغيرات المعتمدة المحددة (2SLDV) .

*البحث مستل من رسالة ماجستير