Journal of Economics and Administrative Sciences
2024; 30(141), pp. 510-528
P-ISSN 2518-5764
E-ISSN 2227-703X

Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq
DOI: https://doi.org/10.33095/s8vcxh81

The use of the (Tobit) Model in Studying the Variables affecting the Increase in the Number of People with Systolic Hypertension, in the Presence of the problem of Heterogeneity of Random Error Variance

Mustafa Hasan Dari*
Department of Statistics
college of Administration and Economics
University of Baghdad
Baghdad, Iraq

Baydaa Ismael AbdulWahhab
Department of Statistics
college of Administration and Economics
University of Baghdad
Baghdad, Iraq

mostafa.hasan1201a@coadec.uob aghdad.edu.iq
baidaa_29@coadec.uobaghdad.edu.iq

***Corresponding author**

Received:27/8/2023 Accepted:25/10/2023 Published Online First: 30 /6/ 2024

Abstract:

The process of regression analysis is based on a number of basic assumptions, and if one of these assumptions is not available, this will lead to obtaining inaccurate results, and the most of these assumptions is homoscedasticity, violating this assumption causes a problem of heteroscedasticity of the random error variance the, which may come from different variances This cause to misleading and inaccurate decisions.

Recently Econometric models received attention, especially limited regression models, which contain specific response variables and several repeated observations within a certain range, including Tobit models and also considered imitations of the Censored Regression Model. Problem arise when it appears in the data heteroscedasticity, which makes methods of estimating the parameters of the linear regression model giving incorrect and inaccurate results, hence we get Biased parameters that do not have minimalist variance, as well as getting an unreliable p-value.

The objective of this research is to study the estimation of the multiple standard (TOBIT) model which features that the variable (Y) specific variant. In heteroscedasticity, which both likelihood function method weighted and Genetic Algorithm was used in simulation.

To sum up, the results showed, that Genetic algorithm better outcomes than weighted likelihood function method therefore, it was used in the practical application.

Paper type: Extracted research.

Keywords: Model (Tobit), The problem Heteroscedasticity, Addressing the heteroscedasticity in the Standard (Tobit) Model, Weighted likelihood function method, Genetic algorithm (G.A)

P-ISSN 2518-5764
E-ISSN 2227-703X

1.Introduction :

Studying regression models, structure of their data, and estimating the parameters of those models by performing statistical inference are extremely important issues of many researchers. Regression analysis is one of the effective statistical methods that describes the form of the relationship between dependent variables and independent variables (Hadi and Dakhel, 2010). And it is also one of the important topics used in many scientific studies (Albayati and Majid, 2018). Also used to analyse experimental problems in the social, economic, and life sciences (Ahmed and Abdullah, 2021). One of the most prominent assumptions is the non-fulfilment of the condition of homogeneity of random error variance (heteroscedasticity), this condition causes one of the most prominent economic measurement problems facing researchers when they build a linear regression model for a different condition, and this leads to misleading and inaccurate decisions (Abdullah, 2018). Recently, econometric models getting a lot of attention, especially limited regression models, which contain specific response variables and have several repeated observations within a certain range, which are also called Tobit models and are considered an imitation of censored regression models. regression models) (Khauoka, 2010). The Tobit model is widely used in the economic, medical and social fields, because most economic studies require its applied in reality for best represent the studied phenomenon (Alhason, 2021). The Tobit model is characterized by dependent variable is determined. Depending on the normal state, it differs from truncated regression models. The model is called a truncated regression model when observations outside a certain range are missing (for the dependent variable and the independent variables), while it is called a specific regression model when the independent variables are observed over an open range while the response variable is observed within Specific range.

1.1 literature review:

Several studies have addressed the problem of heteroscedasticity in Tobit model, including:

Lee et al (1980) developed a two-stage method for simultaneous equations in the estimation of the (Tobit and (Probit) models in which the dependent variables are specific and the comparison between the results of the two models was through the derivation of the covariance matrix and the covariance of the estimators.

Greene (1981) studied the accurate description of the bias of the least squares (OLS) estimators in the (Tobit) model, and the truncated regression model, where he used the method of moments to correct the bias of the (OLS) estimators.

Brännäs and Laitila (1988) dealt with the estimation of parameters and the testing of individual parameters in (Tobit) models in the presence of the problem of non-conformity of variance. Where the statistical properties of the quasi-parametric greatest potential estimators were evaluated.

Carson (1988) studied the effect of the threshold point of determination on the estimators of the greatest possibility, and he concluded that assuming the threshold point of determination incorrectly leads to inconsistent estimates of the estimators of the greatest possibility.

Blundell and Smith (1989) studied the estimation of the regression models for the dependent variables specified for the system of simultaneous equations of the Tobit model as a special case in the estimation, where they developed an effective test approximate to the Tobit model of the simultaneous

Showers and Shotick (1994) used the analysis of the (Tobit) model to analyze the effect of family characteristics on the demand for total insurance, as this approach examines the marginal change in the demand for insurance, as well as the change in the probability of purchasing insurance.

Zuehlke (2003) discussed model (TOBIT) the case that the threshold point of determination is not zero, and through a simulation study, he proved that the approximate distribution of the greatest possible estimator is greatly affected by the estimation of the threshold point of determination.

Richard and Sun (2007) developed the results of Richard and presented an estimator for the threshold point of determination when it is unknown, and they noticed during their study that the estimators of the maximum possibility of and depending on the estimator have a degree of efficiency for the maximum possibility estimator when the true value of is known.

Henningsen (2010) explained how censored regression models including the standard Tobit model can be estimated in R programming for the CensReg add-on package

Chang (2011) studied computationally simulated two-level dynamic board estimator of the Tobit Model developed by (Cragg, 1971) and the probability function was simulated through procedures based on an iterative algorithm formulated by Geweke H. Then the simulation estimates are applied to study the labor supply for married women.

Zhang et al (2017) developed the system of simultaneous equations of the (Tobit) model on the impact of the role of off-farm labor on forest land transfers in China, where the two-stage method of specific dependent variables (2SLDV) was used.

Odah et al (2017) developed Tobit model to identify the determinants of profit distribution among some competing companies in the Iraqi Stock Exchange, and they used a data set for these companies between 2005 and 2015 and reviewed their results in analytical tables.

Muhammad and Abdul Karim (2021) discussed the issue of using some functions in a controlled regression model, such as the Tobit model and the Log-BurrXIIEE model, and estimating the parameters of the two models. Real data on patients with kidney failure were adopted.

Abdulkareem and Mohammed (2022) studied paper that included a comparison between two new censored regression models extended from some continuous distributions with the Burr-XII system. The two models are the Log-BXII Weibull model (LBXIIW) and the Log-BXII exponential model (LBXIIEE). The comparison results showed that the LBXIIW model is better than the LBXIIEE model according to the values of the model selection criteria when creatine is a dependent variable in the model.

The research problem lies when the data suffers from a heteroscedasticity problem, which makes standard Tobit model parameter estimation methods useless in giving correct and accurate results, and thus obtaining biased parameters that do not have the property of minimal variance, in addition to obtaining an unreliable (p) value.

This research aims to use the standard multiple regression model (Tobit) in the presence of the problem of heterogeneity to study the variables that affect the increase in the number of people with high blood pressure (systolic), using the weighted maximum likelihood function method, and the genetic algorithm. A comparison was made between the methods using the comparison criterion (mean square error) (MSE).

2.Material and Methods :

2.1 truncated data:

Truncation of data occurs when some data are excluded for each of the dependent variable and independent variables (regressors variable), for example when studying income (Ahmed, 2018). Given that income is a response variable, low-income people are included and people with income are excluded height of the sample. In terms of the effect, the amputation is achieved when the sample data is withdrawn from a partial group of the large community of the phenomenon (ie the sample has a lack of data) (Cameron and Trivedi, 2005).

2.2 Censored Data:

This type occurs when the dependent variable is missing some data or is limited but the independent variable is complete. For example, if the income levels of people are all within the sample, but the reason is that there are some incomes at a high level, we put a code for them. For example, if the income level is specified by a certain amount, then the income that reaches it or greater than it is controlled. Censoring and vice versa. Controlled data can be defined as data or values that were not seen, or the impossibility of measuring their observations, or not recording complete information about viewing during the study period (Lawless, 2011). It can also be defined as the process of determining the number of failed units for a specific period of time for the experiment or setting a specific time for the experiment and then knowing the number of failed units (Arellano-Valle et al, 2012).
The types of controlled data include:

2.2.1 Type-I Controlled data :

It is also called (censorship from the right), and it is when we have a sample of size (n) and (o) a predetermined period of time, (c) is the number of observations that were observed during the specified period, and the observed value (z) is a random result, and the time is The experiment holds, the number of failures seen is a random variable because the failure times on the are right missing and (c) is a random variable that can only be determined after (o) the time is up (Balakrishnan and Kundu, 2013).

2.2.2. Type-II Controlled data:

It is also called (control from the left), and it is when the sample size (n) and the number of observations that can be observed (c) are fixed and predetermined. Therefore, the time of the experiment will be time (t) which is the unspecified random variable. This type of censorship is used in time-to-event (or event-time) studies (Balakrishnan and Kundu, 2013). where the main difference between type one and type two is the random outcome under controlled data (Rinne, 2008).

2.3 Model (Tobit):

In regression analysis sometimes it is not possible to get data of the response variable (Y) exactly or it may be required to take only part of the data, in such case such types of models cannot be estimated by least squares method, because the data It cannot represent the community completely, i.e. it has been truncated or censored (Mohammed, 2016). To analyse this type of data, finite dependent variable or latent variable models have been developed. One of the models developed for controlled data is the Tobit model, or controlled natural regression model (Alam et al, 2020). The (Tobit) model is a result of a combination between the multiple regression model and (Probit) analysis, where the values of the dependent variable (Y*) are not observed for some observations, although the values of the independent variables (X's) can be observed for all observations. It was proposed by (James Tobin) in (1958), to describe the relationship between the dependent variable and the explanatory variables, through his study of household expenditures on durable goods using a regression model, taking into account the fact that expenditures (as a dependent variable for his model) are a positive value that cannot be negative. The (Tobit) model deals with the data of the dependent variable on the basis that it consists of two parts, and each part takes a specific distribution function (Mohammed, 2014). Observations with zero and negative values take the cumulative distribution function (C.D.F), and observations that take quantities greater than zero will take The probability density function (P.D.F), and by multiplying the two functions (C.D.F) and (P.D.F), we get the mixed function that expresses the (Tobit) model (Tobin, 1958).

The general form of (Tobit) model (Aljanabi and Alhamzawi, 2020) is:

$$Y_i^* = \beta' X_i + U_i \qquad , U_i \sim N(0, \sigma^2) \qquad (1)$$

$$Y_i = \begin{cases} Y_i^* = \beta' X_i + U_i & \text{if } Y_i^* > \lambda \\ 0 & \text{if } Y_i^* \leq \lambda \end{cases} \qquad (2)$$

Where the symbol (Yi) represents the dependent variable, defined for each i=1,2,...,n, and ($Yi*$) represents the response variable (latent variable) generated by a conventional linear regression model that is not observed when Yi* < 0. The symbol (Xi) represents the explanatory variables defined for each i = 1,2,...,n, the symbol (β) represents the parameters of the model, (Ui) represents the random error and represents the set of unobserved variables affecting (Yi*), The symbol (λ) represents the restriction point (Simonof et al, 2007). It is also called the Tobit type I model or standard, and it is mathematically defined by the following formula (Amemiya, 1984):

$$Y_i^* = \beta' X_i + U_i \qquad , i = 1, 2, \ldots, n$$

$$Y_i = \begin{cases} Y_i^* & \text{if } Y_i^* > 0 \\ 0 & \text{if } Y_i^* \leq 0 \end{cases} \qquad (3)$$

2.4 The problem of heteroscedasticity is the random error variance

The term homoscedasticity assumes that prediction errors are similar or have the same prevalence, as it consists of two parts (homo meaning the same and scedastic meaning dispersion (derived from the Greek word skédasi, meaning dispersion) (Gujarati, 2002). Thus, it is an assumption required to justify the use of t and F tests when inferring the coefficients of the linear regression model (Altemimi et al, 2014). This assumption is also required to establish confidence intervals for coefficients or prediction periods associated with the response variable (Shaikhi, 2011), when it is not true Assuming that the error variance is constant, the normal least squares estimates tend to have regression coefficients with very high variances (Abd, 2016), which leads to misleading results and unreliable p-values.

An important assumption of the linear regression model is that the error variance (u_i), conditioned by the chosen values of the explanatory variables, is a constant number equal to σ^2. This is the assumption of homogeneity of variance, i.e. (Hoffmann, 2022).

$$E(u_i)^2 = \sigma^2 \qquad , i = 1, 2, \ldots, n \qquad (4)$$

Where we notice that the conditional variance of the variable (Yi) conditional on the variable (Xi), remains the same regardless of the values that the variable (Xi) takes. On the other hand, we notice from Figure (4), that the conditional variance of the variable (Yi) increases with the increase of the variable (Xi). Here, the differences in (Yi) are not the same. Hence, there is heterogeneity, i.e. that:

$$E(u_i)^2 = \sigma_i^2 \qquad (5)$$

There are many reasons why error variances (ui) are heteroscedasticity as follows:

• The problem of heteroscedasticity can arise as a result of the presence of outliers or values, which are called outlier observations, which are very different observations (whether very small or very large) from the rest of the observations in the sample. Inclusion or exclusion of such an observation, especially if the sample size is small, can significantly alter the results of the regression analysis.

• Another source of heterogeneity of variance is skewness in the distribution of one or more explanatory variables included in the model. An example of this is economic variables such as income, wealth, and education, as it is known that the distribution of income and wealth in most societies is unequal, with few possessing the largest part of income and wealth (Gajardo, 2009).

• Not following error-learning models, as people learn, whereby their errors in behavior become smaller over time or the number of errors becomes more consistent. In this case, the error variance is expected to decrease.

• Other sources of variance inhomogeneity: As researcher David Hendry notes, variance inequality can arise either because of incorrect transformation of the data (for example, ratio or first difference transformations) or because of incorrect shape or form of a function (For example, linear versus logarithmic linear models) (Hendry, 1995).

2.5 Addressing the problem of heterogeneity in the standard (tobit) model:

One of the problems that occur in regression models of different types is the problem of heteroscedasticity of variance for random errors, and therefore each of them has a variance σi^2 = Var (u_i) and for each (i=1,2,3,…….,n1,n1+1,n1+2,……,n) The presence of this problem affects the Tobit regression model by affecting the response variable and the explanatory variables in it, and this can be explained as follows (Kadum and Muslm, 2002)

Rewrite the Tobit regression model as follows:

$$Y_i = \begin{cases} Y_i^* = X_i'B + U_i & \text{if } y^* > 0 \\ 0 & \text{if } y^* \le 0 \end{cases} \quad (6)$$

$$U_i \sim N(0, \sigma^2 w) \qquad w = \frac{1}{w_i}$$

W_{n*n}

So y* will be:

$$\sqrt{w_i} y^* = \left[\sqrt{w_i} \quad \sqrt{w_i}\, X_{1i} \quad \sqrt{w_i}\, X_{2i} \quad \dots\dots\dots.\sqrt{w_i}\, X_{ki}\right] \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_K \end{bmatrix} + \sqrt{w_i}\, U_i \quad (7)$$

$$\tilde{y}_i = \tilde{X}_i' B + \tilde{U}_i$$

For n_1 observations we get the following:

$$\begin{bmatrix} \sqrt{w_1} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{w_2} & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \sqrt{w_{n1}} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_{n1} \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{w_1} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{w_2} & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \sqrt{w_{n1}} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n1} & \cdots & x_{kn1} \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

$$+ \begin{bmatrix} \sqrt{w_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{w_{n1}} \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_{n1} \end{bmatrix}$$

$$P^{-1} y_{n1} = P^{-1} \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_{n1}' \end{bmatrix} \beta + P^{-1} \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_{n1} \end{bmatrix}$$

$$P^{-1} Y_1^* = P^{-1} X_1' \underline{B} + P^{-1} U \quad (8)$$

$$\tilde{Y}_1 = \tilde{X}_1'B + \tilde{U}_1$$

Since:

$$P_1^{-1} = \begin{bmatrix} \sqrt{w_1} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{w_2} & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \sqrt{w_{n1}} \end{bmatrix} \qquad P_1 = \begin{bmatrix} \frac{1}{\sqrt{w_1}} & 0 & 0 & \cdots\cdots & 0 \\ 0 & \frac{1}{\sqrt{W_2}} & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{w_{n1}}} \end{bmatrix}$$

$$P_1P_1' = \begin{bmatrix} \frac{1}{w_1} & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & \frac{1}{w_{n1}} \end{bmatrix} = w_1 \Rightarrow (PP')^{-1} = w_1^{-1} = \begin{bmatrix} w_1 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & w_{n1} \end{bmatrix}$$

Therefore, for y*> 0

$$E\left(U_1^*U_1^{*'}\right) = E[((P_1^{-1}U_1)(P_1^{-1}U_1)']$$

$$= E\left[P_1^{-1}U_1U_1'P_1^{-1'}\right]$$

$$= P_1^{-1}EU_1U_1'P_1^{-1}$$

$$= P^{-1}\sigma^2 w_1 P_1^{-1'}$$

$$= \sigma^2 P_1^{-1}P_1P_1'P_1^{-1'} = \sigma^2 In \qquad (9)$$

As for y*≤ 0, the steps are the same

$$E\left(U_0^*U_0^{*'}\right) = E[((P_0^{-1}U_0)(P_0^{-1}U_0)']$$

$$= E\left[P_0^{-1}U_0U_0'P_0^{-1'}\right]$$

$$= P_0^{-1}EU_0U_0'P_0^{-1}$$

$$= P_0^{-1}\sigma^2 w_0 P_0^{-1'}$$

$$= \sigma^2 P_0^{-1}P_0P_0'P_0^{-1'} = \sigma^2 In \qquad (10)$$

$$E\left(U_0^*U_0^{*'}\right) = \sigma^2 In$$

2.6 Detecting the problem of variance inhomogeneity:

Most often, in economic studies, there is only one sample value (Yi) corresponding to a given value of the variable (Xi), and there is no way to know σi^2 from just observing Yi. Therefore, in most cases involving econometric investigations, heterogeneity may be a matter of hunch, guesswork, or prior empirical experience. There are some tests for detecting heterogeneity (Alquraishi, 2007). and most of these tests are based on the residuals of the OLS method (Hoffmann, 2022).

2.6.1 Breusch–Pagan–Godfrey Test:

It is one of the important tests used to detect the existence of the problem of heterogeneity of the error variance, and this test requires the normal distribution of the data (Bekhet and Fathalla, 2012). as it was proposed in 1979 by the researchers (Breusch and Pagan), and to clarify this test, let us assume we have a regression model The following polylinearity (Gujarati and Porter, 2009):

$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i \qquad , \forall\, i = 1, \ldots, n$

Let us suppose that the error variance function σi^2 is given by the following formula:

$$\sigma_i^2 = f(\theta_1 + \theta_2 z_{2i} + \cdots + \theta_m z_{mi}) \qquad (11)$$

That is, $\sigma i2$ is a function of (non-random) explanatory variables, and (z) represents some or all of the explanatory variables, and suppose that:

$$\sigma_i^2 = \theta_1 + \theta_2 z_{2i} + \cdots + \theta_m z_{mi} \qquad (12)$$

So if $\theta_2 = \theta_3 =. . .= \theta_m = 0$, then $\sigma_i^2 = \theta_1$ means that the variance is constant. Therefore, it can be tested whether σi^2 is homogeneous, through the hypothesis that ($\theta_2 = \theta_3 =. . .= \theta_m = 0$), and this is the basic idea behind the (BPG) test. The steps to perform the test are as follows:

1.Estimation of linear regression model parameters and calculation of errors $\hat{u}_i$.

2.Estimation of variance using the Maximum Likelihood method, $\hat{\sigma}^2 = \sum \hat{u}_i^2 / n$.

3.Construct the defined p_i variables as follows:

$p_i = \hat{u}_i^2 / \hat{\sigma}^2$

It is simply the square of the remainder divided by $\hat{\sigma}^2$.

4.The regression equation (p_i over z) is estimated as:

$$p_i = \theta_1 + \theta_2 z_{2i} + \cdots + \theta_m z_{mi} + v_i \qquad (13)$$

Where v_i is the residual term of the regression equation (13).

5.The sum of squares shown ess is calculated from equation (13), where:

$$\lambda = \frac{1}{2}\, ESS \qquad (14)$$

Assuming that u_i follows a normal distribution, then with increasing sample size (n), λ follows an asymptotic chi-square distribution with a degree of freedom (m-1), that is:

$$\lambda \sim x_{m-1}^2 \qquad (15)$$

If the calculated value (λ) is greater than the tabular value (x^2) at the chosen level of significance, then we reject the homogeneity-variance hypothesis (Breusch and Pagan, 1979).

2.7 Weighted Likelihood function method

The possibility function of the model can be written as follows (Gajardo, 2009).

$$L(\tilde{Y}_i) = \prod_{n2}\left[1 - \Phi_i\left(\frac{\tilde{X}_i'\beta}{\sigma}\right)\right] * \prod_{n1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(\tilde{Y}_i - \tilde{X}_i'\beta)}{2\sigma^2}} \qquad (16)$$

By taking the logarithm of the function:

$$\log L = \textstyle\sum_{n2} \log[1 - \Phi_i] - \frac{n1}{2}\log \sigma^2 - \frac{n1}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum_{n1}(\tilde{Y}_i - \tilde{X}_i'\beta)^2 \qquad (17)$$

By partially deriving the vector parameters B and σ^2, we get the following (Fair, 1977):

$$\frac{\partial \log L}{\partial \beta} = \sum_{n2} \frac{1}{1 - \Phi_i}\left[\frac{-\partial \Phi_i}{\partial \beta}\right] - 0 - 0 - \frac{2\sum_{n1}(\tilde{Y}_i - \tilde{X}_i'\beta)\tilde{X}_i'}{2\sigma^2}$$

Based on the observation made in the source by (Amemiya, 1973). the following:

$$\frac{\partial \Phi_i}{\partial \beta} = f_i \tilde{X}_i' \quad , \; f_i = f(\tilde{Y}_i) \sim N(\tilde{X}'\beta, \sigma^2)$$

Equation is as follows:

$$\frac{\partial \log L}{\partial \beta} = -\sum_{n2} \frac{f_i \tilde{X}_i'}{1 - \Phi_i} + \frac{\sum_{n1}(\tilde{Y}_i - \tilde{X}_i' B)\tilde{X}_i'}{\sigma^2} = 0 \qquad (18)$$

$$\frac{\partial \log L}{\partial \sigma^2} = -\sum_{n2} \frac{1}{1 - \Phi_i}\left[\frac{-\partial \Phi_i}{\partial \sigma^2}\right] - \frac{n1}{2\sigma^2} - 0 - 0 + \frac{\sum_{n1}(\tilde{Y}_i - \tilde{X}_i'\beta)^2}{2\sigma^4} \qquad (19)$$

Based on the following note in the source (Amemiya, 1973):

$$\frac{\partial \Phi_i}{\partial \sigma^2} = -\frac{1}{2\sigma^2}\tilde{X}_i'\beta f_i$$

The equation is as follows:

$$\frac{\partial \log L}{\partial \sigma^2} = +\frac{1}{2\sigma^2}\sum_{n2}\frac{\tilde{X}_i'\beta f_i}{1-\Phi_i} - \frac{n1}{2\sigma^2} + \frac{\sum_{n1}(\tilde{Y}_i - \tilde{X}_i'\beta)^2}{2\sigma^4} = 0 \qquad (20)$$

To obtain the parameter σ^2, equation (18) is multiplied by) $(\beta/2\sigma^2)$, and the result is added to equation (20), as follows:
Substituting $\tilde{X}$ *and* $\tilde{Y}$ for what they are equal to, we get the following:

$$\sigma^2_{WML} = \frac{Y'W^{-1}Y - Y'W^{-1}X_{(1)}\beta}{n1} \qquad (21)$$

Then the parameter vector β can be found:

$$\beta = (\tilde{X}_{(1)}'\tilde{X}_{(1)})^{-1}\tilde{X}_{(1)}'\tilde{Y}_{(1)} - \sigma(\tilde{X}_{(1)}'\tilde{X}_{(1)})^{-1}\tilde{X}_{(0)}'\varphi$$

$$\beta = (X_{(1)}'W_{(1)}^{-1}X_{(1)})^{-1}X_{(1)}W^{-1}Y_{(1)} - \sigma(X_{(1)}'W^{-1}X_{(1)})^{-1}X_0'P_0^{-1}\varphi \qquad (22)$$

$$\hat{\beta} = \hat{\beta}_{(1)GLS} - \sigma(X_{(1)}'W^{-1}X_{(1)})^{-1}X_0'P^{-1}\varphi \qquad (23)$$

$$0 \le \varphi_i \le 1$$

It is noted from formula (22) that $\hat{\beta}_{(1)GLS}$ represents the estimators of the weighted last squares method for the parameter vector β, which depends on all observations related to the first group Y>0. Therefore, formula (22) represents the relationship between the estimators of the least squares The weighted non-zero observations and the estimators of the Tobit regression model parameters when the problem of heteroscedasticity is formed. Therefore, formula (22) represents the estimators of the Tobit model. To find these estimators, the following steps must be followed (Fair, 1977):

1. We calculate $\hat{\beta}_{(1)GLS}$ based on the non-zero data (first set).
2. We calculate $(X_{(1)}'W_{(1)}^{-1}X_{(1)})^{-1}X_{(0)}'P^{-1}$.
3. Choosing initial estimators for the parameter vector β, let it be βo, and then calculating the value of σ^2 defined by the formula (21). Then get $\sigma^{(0)}$ which is the root of $\sigma^{2(0)}$.
4. Finding a vector $\varphi^{(0)}$ using β^0 and $\sigma^{(0)}$ based on the extraction of Φ_i.
5. Calculating the estimators of the parameter vector β defined by the formula (23) using $\sigma^{(0)}$, $\varphi^{(0)}$, which is denoted by $\tilde{\beta}^{(0)}$ and the following equation is calculated:

$$\beta^{(1)} = \beta^{(0)} + \alpha\left(\tilde{\beta}^{(0)} - \beta^{(0)}\right) \quad \text{where } 0 < \alpha \le 1$$

6. The directive $\beta^{(1)}$ will act as a new initial directive for the purpose of repeating all steps except for the first step, as it takes place only once. Through the fourth step, we conclude that the Tobit estimator is an iterative estimator that depends on the parameter α, which is suggested to be taken (0.4), (Gajardo, 2009).

2.8 Genetic algorithm (G.A):

It is a set of operations to find the best solutions, and the genetic algorithm is considered one of the methods of representation and research, where the algorithm is represented as a computer simulation using chromosomes. The operation of the genetic algorithm (on Darwin's theory) depends on the survival of the fittest, which relies on simulating the work of nature, meaning that (organisms that adapt to their external environment in their environment do not apply to the principle of extinction, as organisms with strong characteristics survive while organisms with weak characteristics die (Saleh, 2015). Genetic mutation that occurs in small percentages is also one of the causes that help develop genetic traits transmitted through genes). Genetic algorithms contain a chromosome and a group of communities (Alkhafaji, 2021). A genetic algorithm starts with a set of solutions represented by chromosomes, called a community. Solutions are taken from one community group and used to form a new community group, which is driven by the possibility that the new community will be better than the old one. Moreover, solutions are selected according to their suitability to form new solutions, i.e. offspring.

The above process is repeated until some conditions are met. Computationally speaking, the basic genetic algorithm has been defined as follows (Malhotra, Singh & Singh, 2011):

Step 1: (begin) to generate random sets of chromosomes, i.e. appropriate solutions to the problem

Step 2: (fitness) Assess the fitness of each chromosome in the population.

Step 3: (new community) Create a new community by repeating the following steps until the new community is complete.

a- (Optional) Select two parent chromosomes from a community group according to their fitness. Better fit, more chance to choose father or mother.

b- (crossing over) the possibility of parents crossing over to form new offspring, i.e. children. If no cross is made, the offspring is an exact copy of the parents.

c- (mutation) With the potential for mutation, transform a new offspring at each locus.

d- (acceptance) of placing new offspring in the new community.

Step 4: (Replace) Use the newly created community to run more of the algorithm.

Step 5: (Test) If the final condition is met, stop and return the best solution in the current set.

Step 6: (loop) Go to step two.

2.8.1 Steps of the genetic algorithm according to the model used:

To estimate the parameters of the (Tobit) model, the following steps can be followed (Wadi, 2017):

1- Determine the vectors that represent the lower and upper limits of the parameters to be estimated according to the genetic algorithm, namely.

$$L_{\hat{\theta}} = \hat{\theta}_{LM} - 1.96\sqrt{var(\hat{\theta}_{LM})}$$

$$U_{\hat{\theta}} = \hat{\theta}_{LM} + 1.96\sqrt{var(\hat{\theta}_{LM})}$$

2- Starting the initial set of the genetic algorithm from chromosomes within the minimum and maximum parameters randomly, where each chromosome consists of three genes, because the number of parameters required to be estimated using this algorithm at this stage is three, and the number of chromosomes is equal to the size of the population or the number of members of the community.

3- Calculate the fitness function for each chromosome in the population with formula (3).

4- Choosing a percentage of the chromosomes that will be expressed as crossover, which is chosen according to the fitness function of these chromosomes.

5- Crossing over the chromosomes that were selected in step 4

6- Modification of new chromosomes.

7- Using an elitist strategy to fill in between generations.

Repeating steps 1-7 to achieve the closeness criterion, and when the stopping criterion is satisfied, we obtain the genetic estimates of the parameters that make the log-likelihood function in equation (16) the maximum possible.

2.9 Simulation :

The simulation experiment relied on studying the performance of the estimation methods for the normal state of the data (errors are distributed normally with average and variance, assuming a case of heterogeneity of variance), and with different combinations of model parameters, with different sample sizes, and by repeating the experiment once.

While real data were taken for a sample of 127 cases of patients with systolic blood pressure with a set of explanatory variables, which are believed to affect to some extent the disease, represented by (age and weight) in the applied side.

2.9.1 Generate independent variables:

Explanatory variables $(\boldsymbol{x_1}, \boldsymbol{x_2})$ were generated, which are the same number of real variables, according to the standard normal distribution $\boldsymbol{X \sim N(0, 1)}$.

2.9.2 Generate random errors:

Random errors (U) were generated following a normal distribution with an average of 0 and a variance of σ that is determined according to the initial σ value of the original data, then assumed less, equal and greater values were taken to identify the most important weaknesses and strengths in the estimate.

2.9.3 Generate the dependent variable:

Assuming certain values for the feature vector $(\beta_0, \beta_1, \beta_2)$, the dependent variable is generated using the generated explanatory variables plus the generated random errors.

2.9.4 Description simulation experiments

The following linear regression model was assumed:

$$\boldsymbol{Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + U_i}$$

$\boldsymbol{\beta_0}$	8.41	8.41	8.41	8.41
$\boldsymbol{\beta_1}$	0.049	0.07	0.01	0.029
$\boldsymbol{\beta_2}$	0.029	0.05	0.08	0.01

2.9.5 Generate data for dependent variable Y*

1- The sample size (n) was determined: four sample sizes were chosen: (100,200,300,400) and the experiment was repeated 500 times.

2- In order to create a controlled dependent variable that simulates the practical reality, the set point will be λ = 12. Accordingly, the new dependent variable will be the following:

$$Y = Y^* \quad , \quad if\ Y^* > 12$$
$$= 0 \quad , \quad if\ Y^* \leq 12$$

3- Finding the parameter values $\theta_i \quad i = 0,1,2$ according to the mentioned estimation methods.

4- Calculating the comparison scale between the estimation methods.

2.9.6 Mean Square Error (MSE)

Where it represents the sum of squares of the difference between the (real and estimated) value of the dependent variable (y) divided by the number of those values. Its form is as follows:

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

Since:

R: represents the number of repetitions of the experiment (500).

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

n: represents the sample size.

2.9.7 Simulation model:

A multiple linear regression model was used in the simulation experiments that includes two independent variables (x_1, x_2), and it was applied to all the estimation methods in order to make a clear comparison about the performance of the estimation methods with changing the sample size amounting to 4 sizes and different combinations of the model parameters amounted to 12, assuming a case of heterogeneity Errors varied, and the results were presented based on the number of observations used within the used estimation method.

2.9.8 Estimation methods:

Data with a state of heterogeneity of random error variance were generated in order to study the performance of the estimation methods presented in the research by the following methods As in Tables 1-12 :

1. Genetic Algorithm (GA).
2. Weighted Maximum Likelihood method (MLW).

Table 1: Mean sum square error values of MSE observations for estimation methods for different values of parameters $(\beta_0, \beta_1, \beta_2)$ and for sample size n=100, σ=1

Parameters	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$
Method	8.41 0.049 .029	8.41 0.07 .05	8.41 0.1 0.08	8.41 0.029 0.01
GA	0.7709965	0.7854435	0.8060957	0.7133931
MLW	0.8343372	0.9002265	0.9028559	0.9184383

In general, the best method for different values of the marginal slope parameters is the (GA) method and then the MLW method (MLW) according to the different values of the marginal slope parameters of the study model.

Table 2 : Mean sum square error values of MSE observations for estimation methods for different values of parameters $(\beta_0, \beta_1, \beta_2)$ and for sample size n=200, σ=1

Parameters	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$
Method	8.41 0.049 .029	8.41 0.07 .05	8.41 0.1 0.08	8.41 0.029 0.01
GA	0.7711733	0.7858911	0.8059777	0.7131654
MLW	0.9435326	0.8822548	0.8525613	0.8636939

general, the best method for different values of the marginal slope parameters is the (GA) method and then the MLW method (MLW) according to the different values of the marginal slope parameters of the study model.

Table3: Mean sum square error values of MSE observations for estimation methods for different values of parameters $(\beta_0, \beta_1, \beta_2)$ and for sample size n=300, σ=1

Parameters	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$
Method	8.41 0.049 .029	8.41 0.07 .05	8.41 0.1 0.08	8.41 0.029 0.01
GA	0.7702415	0.7862999	0.8056569	0.7132757
MLW	0.8904639	0.8900153	0.8937221	0.9008079

In general, the best method for different values of the marginal slope parameters is the (GA) method and then the MLW method (MLW) according to the different values of the marginal slope parameters of the study model.

Table 4: Mean sum square error values of MSE observations for estimation methods for different values of parameters $(\beta_0, \beta_1, \beta_2)$ and for sample size n=400, σ=1

Parameters	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$
Method	8.41 0.049 .029	8.41 0.07 .05	8.41 0.1 0.08	8.41 0.029 0.01
GA	0.7713612	0.7865843	0.8053913	0.7132899
MLW	0.8985715	0.8816064	0.8690185	0.8762214

In general, the best method for different values of the marginal slope parameters is the (GA) method and then the MLW method (MLW) according to the different values of the marginal slope parameters of the study model.

Table 5 : Mean sum square error values of MSE observations for estimation methods for different values of parameters $(\beta_0, \beta_1, \beta_2)$ and for sample size n=100, σ=1.26

Parameters	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$
Method	8.41 0.049 .029	8.41 0.07 .05	8.41 0.1 0.08	8.41 0.029 0.01
GA	1.108547	1.147520	1.086610	1.015182
MLW	1.185450	1.280840	1.230584	1.274319

In general, the best method for different values of the marginal slope parameters is the (GA) method and then the MLW method (MLW) according to the different values of the marginal slope parameters of the study model.

Table 6: Mean sum square error values of MSE observations for estimation methods for different values of parameters $(\beta_0, \beta_1, \beta_2)$ and for sample size n=200, σ=1.26

Parameters	$\beta 0$ $\beta 1$ $\beta 2$	$\beta 0$ $\beta 1$ $\beta 2$	$\beta 0$ $\beta 1$ $\beta 2$	$\beta 0$ $\beta 1$ $\beta 2$
Method	8.41 0.049 .029	8.41 0.07 .05	8.41 0.1 0.08	8.41 0.029 0.01
GA	1.108517	1.148263	1.087407	1.016099
MLW	1.314437	1.262735	1.273505	1.327398

In general, the best method for different values of the marginal slope parameters is the (GA) method and then the MLW method (MLW) according to the different values of the marginal slope parameters of the study model.

Table 7: Mean sum square error values of MSE observations for estimation methods for different values of parameters $(\beta_0, \beta_1, \beta_2)$ and for sample size n=300, σ=1.26

Parameters	$\beta 0$ $\beta 1$ $\beta 2$	$\beta 0$ $\beta 1$ $\beta 2$	$\beta 0$ $\beta 1$ $\beta 2$	$\beta 0$ $\beta 1$ $\beta 2$
Method	8.41 0.049 .029	8.41 0.07 .05	8.41 0.1 0.08	8.41 0.029 0.01
GA	1.107231	1.148034	1.087766	1.016514
MLW	1.319675	1.300971	1.326608	1.326853

In general, the best method for different values of the marginal slope parameters is the (GA) method and then the MLW method (MLW) according to the different values of the marginal slope parameters of the study model.

Table 8: Mean sum square error values of MSE observations for estimation methods for different values of parameters $(\beta_0, \beta_1, \beta_2)$ and for sample size n=400, σ=1.26

Parameters	$\beta 0$ $\beta 1$ $\beta 2$	$\beta 0$ $\beta 1$ $\beta 2$	$\beta 0$ $\beta 1$ $\beta 2$	$\beta 0$ $\beta 1$ $\beta 2$
Method	8.41 0.049 .029	8.41 0.07 .05	8.41 0.1 0.08	8.41 0.029 0.01
GA	1.108441	1.149037	1.086824	1.016529
MLW	1.292124	1.242854	1.336467	1.285404

In general, the best method for different values of the marginal slope parameters is the (GA) method and then the MLW method (MLW) according to the different values of the marginal slope parameters of the study model.

Table 9: Mean sum square error values of MSE observations for estimation methods for different values of parameters $(\beta_0, \beta_1, \beta_2)$ and for sample size n=100, σ=2

Parameters	$\beta 0$ $\beta 1$ $\beta 2$	$\beta 0$ $\beta 1$ $\beta 2$	$\beta 0$ $\beta 1$ $\beta 2$	$\beta 0$ $\beta 1$ $\beta 2$
Method	8.41 0.049 .029	8.41 0.07 .05	8.41 0.1 0.08	8.41 0.029 0.01
GA	2.607714	2.675563	2.730039	2.827383
MLW	2.716885	2.714612	2.790580	2.830932

In general, the best method for different values of the marginal slope parameters is the (GA) method and then the MLW method (MLW) according to the different values of the marginal slope parameters of the study model.

Table 10 : Mean sum square error values of MSE observations for estimation methods for different values of parameters $(\boldsymbol{\beta_0}, \boldsymbol{\beta_1}, \boldsymbol{\beta_2})$ and for sample size n=200, σ=2

Parameters	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$
Method	8.41 0.049 .029	8.41 0.07 .05	8.41 0.1 0.08	8.41 0.029 0.01
GA	2.607170	2.673164	2.729460	2.826959
MLW	2.727223	2.962954	2.923588	2.816083

In general, the best method for different values of the marginal slope parameters is the (GA) method and then the MLW method (MLW) according to the different values of the marginal slope parameters of the study model.

Table11: Mean sum square error values of MSE observations for estimation methods for different values of parameters $(\boldsymbol{\beta_0}, \boldsymbol{\beta_1}, \boldsymbol{\beta_2})$ and for sample size n=300, σ=2

Parameters	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$
Method	8.41 0.049 .029	8.41 0.07 .05	8.41 0.1 0.08	8.41 0.029 0.01
GA	2.606457	2.675698	2.729746	2.826858
MLW	2.683951	2.877319	2.811051	2.731710

In general, the best method for different values of the marginal slope parameters is the (GA) method and then the MLW method (MLW) according to the different values of the marginal slope parameters of the study model.

Table 12: Mean sum square error values of MSE observations for estimation methods for different values of parameters $(\boldsymbol{\beta_0}, \boldsymbol{\beta_1}, \boldsymbol{\beta_2})$ and for sample size n=400, σ=2

Parameters	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$	$\beta0$ $\beta1$ $\beta2$
Method	8.41 0.049 .029	8.41 0.07 .05	8.41 0.1 0.08	8.41 0.029 0.01
GA	2.607257	2.674953	2.729253	2.827832
MLW	2.863231	2.783693	2.911912	2.873036

In general, the best method for different values of the marginal slope parameters is the (GA) method and then the MLW method (MLW) according to the different values of the marginal slope parameters of the study model.

3. Discussion of Results

3.1 Description of Data:

In order to study the performance of the estimation methods of the (TOBIT) model in practice, real data were created that suffer from the problem of heterogeneity of random error variance in the health aspect represented by studying the effect of some factors on the incidence of high blood pressure (systolic), and the sample included 127. According to studies Medical, it is believed that there are many factors affecting to certain degrees in determining the amount of high blood pressure (systolic).

3.2 Analysis of Data:

With regard to the dependent variable represented by the amount of blood pressure (Y^*), medical sources indicate that the normal rate of blood pressure is (120) or (12), so the number (12) was adopted, and on this basis the dependent variable under study is considered a limited variable, since:

$$\mathbf{Y = Y^*} \qquad \mathbf{Y^* > 12}$$
$$= 0 \qquad Y^* \leq 12$$

Accordingly, the threshold point of determination is λ = 12, and in line with the presented theoretical presentation, the threshold point of determination is considered to be λ = 0, and therefore the model will devolve into the following model:

$$Y = Y^* \qquad Y^* > 12$$

$$= 0 \qquad Y^* \leq 12$$

The in independent variables were coded as follows:
X_1: The weight of the response unit is measured in kilograms.
X_2: the lifetime of the response unit measured in years.
In order to analyze the model under study, the parameters of the model were estimated and the relevant indicators calculated.

And as a preliminary measure to test the existence of the problem of heterogeneity of variance, the independent variables X_1 (weight) and X_2 (age) were studied, and to detect the presence of heterogeneity of variance, the **Breusch–Pagan–Godfrey Test** was used, which is explained in the theoretical aspect (1-11-2). Test results according to the hypotheses below:
Null hypothesis: that the variance of the error limit for all sample observations is constant, meaning that there is no problem of heterogeneity of the random error variance.

$$\boldsymbol{H_0}: \boldsymbol{\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \cdots = \sigma_i^2}$$

Alternative hypothesis: at least two different variances i.e. a heterogeneity problem exists

$$\boldsymbol{H_1}: \textit{The problem of heterogeneity exists}$$

Since the calculated (chi-square) value of (11.086) is greater than the tabular chi-square value, in addition to the value of (p-value = 0.003914), which is less than 0.05, so the alternative hypothesis will be accepted, which indicates that there is a statistical significance on The existence of the problem of heterogeneity of variance for both explanatory variables and In the light of the results of the preliminary analysis of the model, the data can be applied according to the computer program prepared by the researcher for the (DE) and (GA) methods, in order to study the performance of those methods in estimating the parameters of the standard (TOBIT) model.

Parameters	Initial value	GA
β_0	8.41	
β_1	0.049	0.8003630
β_2	0.029	

Through the table above, the parameters of the (Tobit) model were estimated in the case of heterogeneity of random error variance, using the best estimation methods extracted from the experimental side, which were as follows:

Table 13: The estimated values of the model parameters according to the best estimation methods

Parameters	GA
β_0	8.49
β_1	0.1097
β_2	0.126

P-ISSN 2518-5764
E-ISSN 2227-703X

4 Conclusion:

In terms of efficiency, the estimations of the algorithm (GA) showed that any of the variables by one unit leads to an increase in the probability of disease occurring by the values of the parameters. Likewise, we note that the value of the selection parameters according to the (tobit) model under study explains 36% of the change in the dependent variable (Y). 64% are explained by other variables that were not included in the model.

Authors Declaration:

Conflicts of Interest: None

-We Hereby Confirm That All The Figures and Tables In The Manuscript Are Mine and Ours. Besides, The Figures and Images, Which are Not Mine, Have Been Permitted Republication and Attached to The Manuscript.

- Ethical Clearance: The Research Was Approved By The Local Ethical Committee in The University.

References :

1.Abd, H.O., (2016). Econometrics, first edition, Dar Al-Kutub Press, Iraq, Karbala.

2.Abdulkareem, A.D. and Mohammed, S.R., (2022). Comparison between two new censored regression models extended from Burr-XII system with application. International Journal of Nonlinear Analysis and Applications, Vol. 13, No. 1, pp.3395-3403.

3.Abdullah, R.A., (2018). Choosing the best test method for the problem of heterogeneity of variance in the multiple regression model (with practical application), Master's, University of Karbala, Karbala.

4.Ahmed, D.N.M., (2018). Limited Dependent Variable Modeling (Truncated and Censored Regression models) with Application. The Scientific Journal of Cihan University–Sulaimanyia PP, 82-96.

5.Ahmed, F.M. and Abdullah, E.M., (2021). Comparison Between Nelson-Olson Method and Two-Stage Limited Dependent Variables (2SLDV) Method for the Estimation of a Simultaneous Equations System (Tobit Model). Journal of Economics and Administrative Sciences, Vol. 27, No. 130, pp. 210-226.

6.Alam, K. M., Xuemei, L., Baig, S., Yadong, L. and Shah, A. A. (2020). Analysis of technical, pure technical and scale efficiencies of Pakistan railways using data envelopment analysis and Tobit regression model. Networks and Spatial Economics, Vol. 20, pp. 989-1014.

7.Albayati, M.M. and Majid, H.H., 2018, Tobit Quantile Regression Bayesian model using the fourth level of prime distributions. Journal of Economics and Administrative Sciences, Vol. 105, No. 24, pp. 487-487.

8.Alhason, F.M.A., (2021), Estimation of the System of Simultaneous Equations (Tobit Model) with Practical Application, Master, University of Baghdad, Baghdad.

9.Aljanabi, S. and Alhamzawi, R., (2020). A New Bayesian Group Bridge to Solve the Tobit Model. Iraqi Journal of Science, pp.215-222.

10. Alkhafaji, M.A.A., (2021), Comparison of some estimators of skewed normal distribution parameters by adopting genetic algorithms, Master, University of Baghdad, Baghdad.

11. Alquraishi, I.K.S., (2007). Parametric methods and non-parametric methods in statistical tests, first edition, College of Administration and Economics, Al-Mustansiriya University.

12. Altemimi, Z.H.A., and others. (2014). Regression analysis", first edition, Directorate of Dar Al-Kutub for Printing and Publishing, Basra University, Iraq, Basra.

13. Amemiya, T., (1973). Regression analysis when the dependent variable is truncated normal. Econometrica: Journal of the Econometric Society, pp.997-1016.

14. Amemiya, T., (1984). Tobit models: A survey. Journal of econometrics, Vol. 24, No. 1-2, pp.361.

15. Arellano-Valle, R.B., Castro, L.M., González-Farías, G. and Muñoz-Gajardo, K.A., (2012). Student-t censored regression model: properties and inference. Statistical Methods & Applications, Vol. 21, pp.453-473.
16. Balakrishnan, N. and Kundu, D., (2013). Hybrid censoring: Models, inferential results and applications. Computational Statistics & Data Analysis, Vol. 57, No. 1, pp.166-209.
17. Bekhet, H.A., Fathalla, S., (2012). Econometrics", Dar Al-Kutub Press, Baghdad.
18. Blundell, R.W. and Smith, R.J., (1989). Estimation in a class of simultaneous equation limited dependent variable models. The Review of Economic Studies, Vol. 56, No. 1, pp.37-57.
19. Brännäs, K. and Laitila, T., (1989). Heteroskedasticity in the Tobit model. Statistical papers, Vol. 30, No. 1, pp.185-196.
20. Breusch, T.S. and Pagan, A.R., (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica: Journal of the econometric society, pp.1287-1294.
21. Calzolari, G. and Fiorentini, G., (1998). A tobit model with garch errors. Econometric Reviews, Vol. 17, No. 1, pp.85-104.
22. Cameron, A.C. and Trivedi, P.K., (2005). Microeconometrics: methods and applications. Cambridge university press.
23. Carson, R.T. and Sun, Y., (2007). The Tobit model with a non- zero threshold. The Econometrics Journal, Vol. 10, No. 3, pp.488-502.
24. Carson, R.T., (1988). The tobit model with a non-zero censoring threshold. Working Paper88-37, Department of Economics, University of California, San Diego.
25. Chang, S.K., (2011). Simulation estimation of two-tiered dynamic panel Tobit models with an application to the labor supply of married women. Journal of Applied Econometrics, Vol. 26, No. 5, pp.854-871.
26. Fair, R. C. (1977). A Note on the Computation of the Tobit Estimator. Econometrica: Journal of the Econometric Society, pp. 1723-1727.
27. Gajardo, K.A.M., (2009). An Extension of the Normal Censored Regression. Pontificia Universidad Catolica De Chile, Santiago, Chile.
28. Greene, W.H., (1981). On the asymptotic bias of the ordinary least squares estimator of the Tobit model. Econometrica: Journal of the Econometric Society, pp.505-513.
29. Gujarati, D.N., (2002). Basic Econometrics 4th ed.
30. Gujarati, N.D. and Porter, D.C., (2009). Basic econometrics. International edition McGraw-Hill/Irwin, a business unit of the McGraw-Hill companies. Inc., New York.
31. Hadi, F.H. and Dakhel,T.R., (2010). Studying the effect of increasing the registration of organic partners using the tobit model. Journal of Economics and Administrative Sciences, Vol. 59, No. 16, pp. 179-179.
32. Hendry, D.F., (1995). Dynamic econometrics. Oxford university press.
33. Henningsen, A., (2010). Estimating censored regression models in R using the censReg Package. R package vignettes, Vol. 5, pp. 1-12.
34. Hoffmann, J.P., (2022). Linear regression models: applications in R. Crc Press.
35. Kadum, A.H., Muslm, B.S., (2002). Advanced Economic Measurement - Theory and Practice, Al-Taif Press, Baghdad.
36. Khauoka, A.F.T., (2010). Comparison of some methods of estimating the Tobit model using simulation with its application in the field of health", PhD, University of Baghdad, Baghdad.
37. Lawless, J.F., (2011). Statistical models and methods for lifetime data. John Wiley & Sons.
38. Lee, L.F., Maddala, G.S. and Trost, R.P., (1980). Asymptotic covariance matrices of two-stage probit and two-stage tobit methods for simultaneous equations models with selectivity. Econometrica: Journal of the Econometric Society, pp.491-503.
39. Li, L., Simonoff, J.S. and Tsai, C.L., (2007). Tobit model estimation and sliced inverse regression. Statistical Modelling, Vol. 7, No. 2, pp.107-123.
40. Malhotra, R., Singh, N. and Singh, Y., (2011). Genetic algorithms: Concepts, design for optimization of process controllers. Computer and information science, Vol. 4, No. 2, p.39.

41. Mohammed, H.F.A.I., (2014), Using the Tobit Model in Analyzing Blood Components Data in Children with Cancer, Master Thesis, College of Graduate Studies, Sudan University of Science and Technology.
42. Mohammed, S.R. and Abdulkareem, A.D., (2021). Comparison some censored regression models with application to renal failure data. International Journal of Nonlinear Analysis and Applications, Vol. 12, pp.799-807.
43. Mohammed, S.R., (2016). Comparison Study to Some Censored Regression Model with Application. B.Sc. in Statistics, Kirkuk University.
44. Odah, M.H., Mohammed, B.K. and Bager, A.S.M., (2018). Tobit regression model to determine the dividend yield in Iraq. LUMEN Proceedings, Vol. 3, pp.347-354.
45. Rinne, H., (2008). The Weibull distribution: a handbook. CRC press.
46. Saleh, Y.I., (2015). Solving the problem of multi-objective allocation using the genetic algorithm with a practical application, Master, University of Baghdad, Baghdad.
47.
48. Shaikhi, M., (2011). Econometric Methods, first edition, Dar Al-Hamid, Amman, Jordan.
49. Showers, V.E. and Shotick, J.A., 1994. The effects of household characteristics on demand for insurance: A tobit analysis. Journal of Risk and Insurance, pp.492-502.
50. Tobin, J., (1958). Estimation of relationships for limited dependent variables. Econometrica: journal of the Econometric Society, pp.24-36.
51. Wadi, A.S., (2017). Improved estimations of the parameters of the two-dimensional sine model with a practical application", PhD, University of Baghdad, Baghdad.
52. Zhang, H., Kuuluvainen, J., Yang, H., Xie, Y. and Liu, C., (2017). The effect of off-farm employment on forestland transfers in China: A simultaneous-equation Tobit model estimation. Sustainability, Vol. 9, No. 9, p.1645.
53. Zuehlke, T.W., (2003). Estimation of a Tobit model with unknown censoring threshold. Applied Economics, Vol. 35, No. 10, pp.1163-116

Journal of Economics and Administrative Sciences
2024; 30(141), pp. 510-528
P-ISSN 2518-5764
E-ISSN 2227-703X

استعمال انموذج (Tobit) في دراسة المتغيرات التي تؤثر على زيادة اعداد المصابين بمرض ضغط الدم الانقباضي بوجود مشكلة عدم تجانس التباين الخطأ العشوائي.

مصطفى حسن ضاري (1)
جامعة بغداد/ كلية الإدارة والاقتصاد
بغداد، العراق
mostafa.hasan1201a@coadec.uobaghdad.edu.iq

بيداء إسماعيل عبد الوهاب (2)
جامعة بغداد/ كلية الإدارة والاقتصاد
بغداد، العراق
baidaa_29@coadec.uobaghdad.edu.iq

Received:27/8/2023 Accepted:25/10/2023 Published Online First: 30 /6/ 2024

مستخلص البحث:

تعتمد عملية تحليل الانحدار على عدد من الافتراضات الأساسية، وفي حالة عدم توفر أحد هذه الافتراضات سيؤدي ذلك إلى الحصول على نتائج غير دقيقة، وأكثر هذه الافتراضات هي تجانس تباين الخطأ العشوائي، حيث ان خرق هذا الفرض يسبب مشكلة عم تجانس تباين الخطأ العشوائي، والتي قد تأتي من تباينات مختلفة في المجتمع حيث ان الفروق تؤدي إلى اتخاذ قرارات مضللة وغير دقيقة. حظيت نماذج الاقتصاد القياسي مؤخرا بالاهتمام، وخاصة نماذج الانحدار المحدود، والتي تحتوي على متغيرات استجابة محددة وعدة مشاهدات متكررة ضمن نطاق معين، بما في ذلك نماذج (Tobit) وتعتبر أيضا تقليدا لنموذج الانحدار الخاضع للرقابة. تنشأ المشكلة عندما تظهر في البيانات مشكلة عم تجانس تباين الخطأ العشوائي، مما يجعل طرق تقدير معلمات نموذج الانحدار الخطي تعطي نتائج غير صحيحة وغير دقيقة، وبالتالي نحصل على معلمات متحيزة لا تحتوي على اقل تباين، وكذلك الحصول على قيمة (p) غير موثوقة. يهدف هذا البحث إلى دراسة تقدير انموذج (TOBIT)القياسي المتعدد الذي يتميز بكون المتغير (Y) محدد في ظل مشكلة عدم التجانس، تم استخدام كل من طريقة الاحتمالية الموزونة والخوارزمية الجينية في المحاكاة. خلاصة القول، أظهرت النتائج أن الخوارزمية الجينية أفضل نتائج من طريقة دالة الاحتمال الموزونة، لذلك تم استخدامها في التطبيق العملي.

نوع البحث: بحث مستل .

المصطلحات الرئيسة للبحث: انموذج (TOBIT)، مشكلة عدم تجانس تباين الخطأ العشوائي، معالجة مشكلة عدم التجانس في النموذج (Tobit)) القياسي، طريقة دالة الإمكان الأعظم الموزونة، الخوارزمية الجينية (G.A) .

*البحث مستل من رسالة ماجستير