ORIGINAL RESEARCH ARTICLE UHD JOURNAL OF SCIENCE AND TECHNOLOGY

Using Tobit Model for Studying Factors Affecting Blood Pressure in Patients with Renal Failure

Raz Muhammad H. Karim*, Samira Muhamad

Department of Statistics, College of administration and Economic, University of Sulaimani, Sulaymaniyah, Iraq

ABSTRACT

In this study, the Tobit Model as a statistical regression model was used to study factors affecting blood pressure (BP) in patients with renal failure. The data have been collected from (300) patients in Shar Hospital in Sulaimani city. Those records contain BP rates per person in patients with renal failure as a response variable (Y) which is measured in units of millimeters of mercury (mmHg), and explanatory variables (Age [year], blood urea measured in milligram per deciliter [mg/dl], body mass index [BMI] expressed in units of kg/m^2 [kilogram meter square], and Waist circumference measured by the Centimeter [cm]). The two levels of BP; high and low were taken from the patients. The mean arterial pressure (MAP) was used to find the average of both levels (high and low BP). The average BP rate of those patients equal to or >93.33 mmHg only remained in the dataset. The 93.33 mmHg is a normal range of MAP equal to 12/8 mmHg normal range of BP. The others have been censored as zero value, i.e., left censored. Furthermore, the same data were truncated from below. Then, in the truncated samples, only those cases under risk of BP (greater than or equal to BP 93.33mmHg) are recorded. The others were omitted from the dataset. Then, the Tobit Model applied on censored and truncated data using a statistical program (R program) version 3.6.1. The data censored and truncated from the left side at a point equal to zero. The result shows that factors age and blood urea have significant effects on BP, while BMI and Waist circumference factors have not to affect the dependent variable(y). Furthermore, a multiple regression model was found through ordinary least Square (OLS) analysis from the same data using the Stratigraphy program version 11. The result of (OLS) shows that multiple regression analysis is not a suitable model when we have censored and truncated data, whereas the Tobit model is a proficient technique to indicate the relationship between an explanatory variable, and truncated, or censored dependent variable.

Index Terms: Tobit Model, Censored Regression, Truncated Regression, Renal Failure, Blood Pressure

1. INTRODUCTION

In economic and social research, many types of regression models applied. Their use is dependent on the nature of the data. The Tobit model is regarded as the most appropriate statistical model for solving those cases that the dependent variable is censored or truncated [1]. Tobit regression has been the subject of great theoretical interests in numerous practical applications. It has been developed and used in many fields, such as econometrics, finance, and medicine [1], [2]. Furthermore, it is regarded as a linear regression model where only data on the response variable incompletely observed; the response variable is censored at zero. Kidney diseases are common diseases worldwide; it is a global public health problem affecting 750 million persons globally [3]. It plays an important role in preserving normal body functions. Most people are not aware of their impaired kidney functions. In

Access this article online

DOI: 10.21928/uhdjst.v4n2y2020.pp1-9

E-ISSN: 2521-4217

P-ISSN: 2521-4209

Corresponding author's e-mail: Raz Muhammad H. Karim Department of Statistics, College of administration and Economic, University of Sulaimani, Sulaymaniyah, Iraq. E-mail: razmauhammad@gmail.com

Received: 16-05-2020 Accepted: 26-07-2020 Published: 01-07-2020

fact, kidney failure is a "silent illness" that sometimes has no obvious early symptoms. Many people with kidney diseases are not conscious that they are at high risk of kidney failure, which could require dialysis or transplantation. Often the disease such as diabetes with high blood pressure (BP) may cause kidney damage. Hypertension (high BP) is both a cause and a consequence of renal diseases, which are difficult to distinguish its types clinically [4]. Hence, the importance of this research comes as studies the factors that affect BP in patients with renal disease and knowing the real causes of it. This is crucial for medical staff and specialists (doctors) to eliminate problems and limit the spread of kidney diseases because high BP is both a cause and a consequence of kidney diseases. In this study, we find an influence on each independent variable of the dependent variable (BP). It is known that the normal BP range is 12/8 mmHg [5]. This value change due to many factors, and any change in this range make many health problems. Therefore, controlling BP and finding factor, everybody should take care of it. In this study, the data collected from patients in a dialysis center at Shar hospital in Sulaimani city. The two levels of BP; high and low BP from the patients (as dependent variables) and some independent variables (Age, blood, urea, body mass index [BMI], and Waist circumference) were taken. Each patient has their own specific BP (high, low), then we could not take high and low BP separately for our study. That is why the mean arterial pressure (MAP) was performed. It is an average arterial pressure contains high and low BP [6]. A threshold point equal to 93.33 is determined and found by MAP equation [7], equal to 12/8 mmHg, which is a normal range of BP. We assumed any value lower that range is equal to zero. Therefore, the Tobit regression model is used because some variables are equal to zero for a number of observations. This is a phenomenon that can generally be termed censored or truncated data. After that multiple regression model performed for the same data based on ordinary least square (OLS) analysis, it is found that a multiple regression model is not suitable for analysis because there are a number of observations in the dependent variable equal to zero. The use of OLS models in the case of censored sample datasets and depending on the number of zeros makes OLS estimated bias [8].

2. AIM OF THE STUDY

The aim of this study is to detect the impact of the independent variables (Age, blood, urea, BMI, and waist circumference) on dependent variables (BP) in patients with renal failure putting these results in front of specialists to eliminate a problem using a statistical model (Tobit model). Knowing which factor in the independent variable more effect on the dependent variable also comparison between (OLS) and Tobit model estimation to knowing which of them are suitable models for estimation.

3. RELATED WORK

Odah *et al.* [9] displayed the most significant factors affecting loans provided by Iraqi banks and the best methods to estimate the data using a Tobit regression model and OLS method. Liquidity and loan repayment were found to affect loans from the Iraqi Banks, while the effects of interest rate and borrowers were not statistically significant. The outcome of Tobit and OLS estimations indicate that bias will result when estimating Iraqi bank loans using OLS if bank loans are limited.

Prahutama *et al.* [10] used a Tobit regression model to study factors that affect household expenditure on education in Semarang city. The dependent variable used in this study is household expenditure for education. The independent variables used include the Education of the Head of the Household, Occupation of the Head of the household, number of household members, Number of Working Household Members, the proportion of household members who attend school in Junior High School, Senior High School and College, and food expenditure in households and regions. Based on the Tobit regression analysis proportion of household members who are taking education in college is the most significant contribution to the high cost of household expenditure.

Ahmed [11] applied a Tobit (Truncated), (censored) data regression models and multiple regression with the least square method for persons whose levels exceed 120 g/dl under the risk of diabetes injure, in the sample data (n = 500) on the assumption that blood sugar (y), depends on the explanatory (Age: X1, Cholesterol: X2 gram/deciliter, and Triglycerides: X3 gram/deciliter). The results revealed that the censored regression model was more applicable than the other regression models (truncated, and multiple regression), the two factors (Age and triglycerides have highly significant effects on the blood sugar.

Ahmad *et al.* [12] used Tobit regression analysis and data envelopment analysis (DEA) to address some of the important working capital management policies and efficiency regarding the manufacturing sector of Pakistan. To achieve that data from 37 firms have been taken for the periods 2009–2014.

Tobit regression analysis concludes that the average period has significant negative impacts on efficiency and current ratio, gross working capital turnover ratio, and financial leverage ratio that have a positive significant impact on efficiency.

Samsudin *et al.* [13] applied the Tobit model and DEA to examine the efficiency of public hospitals in Malaysia and identify the factors affecting their performance. The study analysis was based on 25 public hospitals in the northern region of Malaysia. According to the result of this study found that the daily average number of admission, the number of outpatient per doctor, and hospital classification have significant influences on hospital inefficiencies.

Odah *et al.* [14] investigated the factors affect divorce decision, and determine the most important factors causing divorce in Iraq through using the Tobit regression model and probit regression model. The data were collected through the application of the questionnaire. According to Tobit regression analysis results, marital infidelity is the main reason for the increase in divorce cases, as well as the preoccupation of the couple with social networking sites. After using the probit model, it found that age, social media sites, and income have a significant impact on the decision to divorce.

Zorlutuna *et al.* [15] applied Tobit regression analysis for the measurement of lung cancer patients. Data taken from Sivas Cumhuriyet University Faculty of Medicine Research and Application Hospital Oncology Center consists of 535 patients who have lung cancer. Tobit regression results show that when the dependent variable phase of the patient's disease, the patient's gender, patient's condition, and the pathological consequences of the disease were found to be statistically significant variables. The sex of patient has positive effect on the stage of the disease, while pathological condition has negative influences.

Anastasopoulos *et al.* [16] provided a demonstration of Tobit regression as a methodological approach to gain new insights into the factors that significantly influence accident rates. Using 5 years of vehicle accident data from Indiana, the estimation results show that many factors relating to pavement condition, roadway geometrics, and traffic characteristics significantly affect vehicle accident rates.

4. TOBIT MODEL

The regression analysis is one of the statistical methods used to explain the relationship between explanatory variables and the dependent variable. Therefore, choosing an appropriate model for the available data is a necessity of this analysis. In many statistical analyses of individual data, the dependent variable is censored. If the dependent variable is censored, the use of a conventional regression model with this type of data will lead to a bias in the estimation of the parameters there for the best model for this type of data is the Tobit model [17]. The Tobit model family of statistical regression models defines the relationship between censored or truncated continuous dependent variables and some independent variables [18]. It has been used in many areas of applications, including dental health, medical research, and economics [2]. The Tobit model refers to a regression model where the range of dependent variables is limited in some ways [16]. A model invents by Tobin in which it is supposed that the dependent variable has a number of its values clustered at limited value, usually zero [19]. This model was first introduced statistical literature in the 1950s and was called "censored normal regression model." It has been used for health studies since the 1980s. The Tobit model is an efficient method for estimating the relationship from Probit between an explanatory variable and truncated or censored dependent variable. The origin of the Tobit model is from Probit analysis and multiple regressions. The benefit of this model, using all the information that either Probit models (or logit) or OLS, would allow separately [20].

4.1. The Structural Equation Model [21], [22]

The structural equation in the Tobit model is

$$y^*_{i=}X_i{}^2 + e_i \quad (1)$$

Where $e_{i\sim}$ $N(0,\sigma^2)$. y^* is a latent variable that is observed for values greater than τ and censored otherwise. The observed y is defined by the following measurement equation

$$y_i = \begin{cases} y^* \text{ if} \left| y^* > \tau \right. \\ \tau_y \text{ if} \left| y^* \leq \tau \right. \end{cases} \quad (2)$$

In the typical Tobit model, we assume that τ=0, i.e., the data are censored at 0. T use we have

$$y_i = \begin{cases} y^* \text{ if} \left| y^* > 0 \right. \\ 0 \text{ if} \left| y^* \leq 0 \right. \end{cases} \quad (3)$$

4.2. Estimations

As we have seen from earlier, the likelihood function for the censored normal distribution is

$$L = \prod_i^N [\frac{1}{\sigma}\varnothing\left(\frac{y-\mu}{\sigma}\right)]^{di}[1-\varnothing(\frac{\mu-\tau}{\sigma})]^{1-di} \quad (4)$$

Where τ is the censoring point. In the traditional Tobit model, we set t=0 and parameterize μ as $X_i\beta$. This gives us the likelihood function for the Tobit model:

$$L = \prod_{i}^{N} [\frac{1}{\sigma} \varnothing\left(\frac{y_i - X_i\beta}{\sigma}\right)]^{di} [1 - \varnothing(\frac{X_i\beta}{\sigma})]^{1-di} \quad (5)$$

The log-likelihood function for the Tobit model is

$$\ln L = \sum_{i=1}^{N} \{d_i\left(-\ln\sigma + \ln\varnothing\left(\frac{y_i - X_i\beta}{\sigma}\right)\right) + (1-d_i)\ln\left(1-\varnothing\left(\frac{X_i\beta}{\sigma}\right)\right)\} \quad (6)$$

The overall log-likelihood is made up of two parts. The first part corresponds to the classical regression for the uncensored observations, while the second part corresponds to the relevant probabilities that observation is censored.

5. TRUNCATION, CENSORING, (TRUNCATED AND CENSORED) DISTRIBUTION, AND MARGINAL EFFECT

The leading causes of incompletely observed data are truncation and censoring.

5.1. Truncation

The effect of truncation occurs when the observed data in the sample only drawn from a subset of a larger population [23]. On the other hand, a dependent variable in a model is truncated, if observations cannot be seen when taking value with a certain range. This means, both the independent and the dependent variables are not observed when the dependent variable is in that range [24]. There are two types of Truncation: from below and from above (Truncation from left and Truncation from right). Figs. 1 and 2 explain the probability distribution of Truncated from below [11].

5.2. Censoring

The idea of "censoring" is that some data above or below the threshold is misreported at the threshold. Hence, the observed data are generated by a mixed distribution with both a continuous and a discrete component. The censoring process may be explicit in the data collection process, or it may be a by-product of economic constraints involved in constructing the data set [24]. When the dependent variable is censored, values in a certain range are all transformed to (or reported as) a single value [25]. Fig. 3 Explain the probability distributions of Censored from below [11].

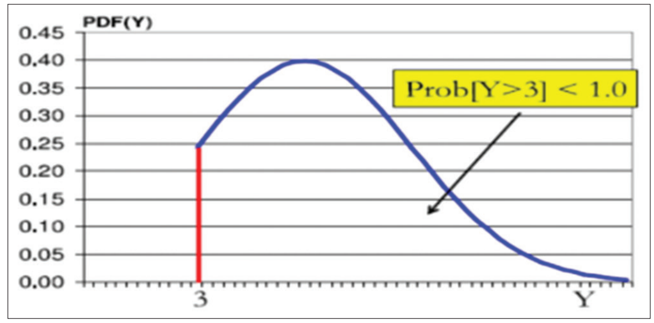

Fig. 1. Truncated from below with the probability distribution explaining (threshold = 3) [11].

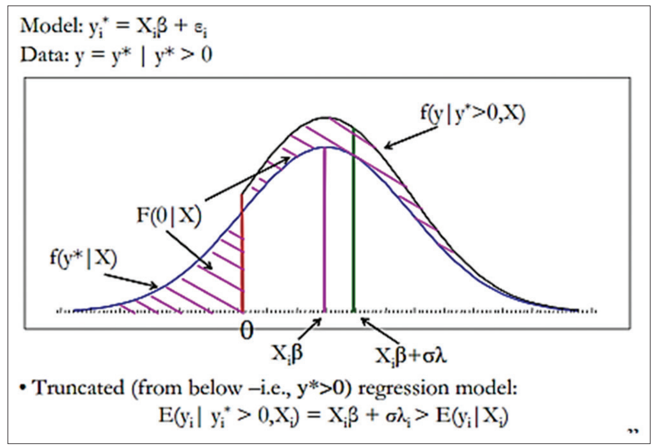

Fig. 2. Truncated normal distribution [11].

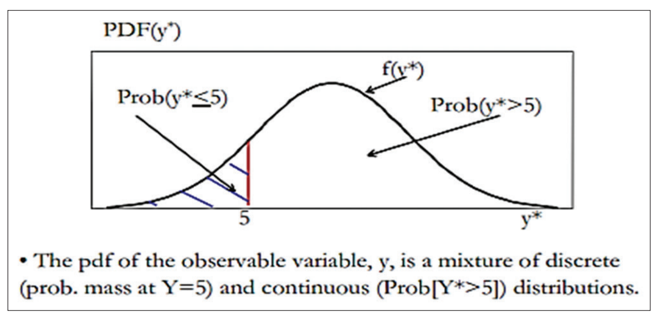

Fig. 3. Censored from below with the probability distributions explaining (threshold = 5) [11].

5.3. (Truncated and Censored) Distribution [21]

After formally considering the Tobit model, we need some results about truncated and censored normal distribution. These distributions are at the foundation of most models for truncation and censoring. The results are given for censoring and truncation on the left, which translate into censoring from below in the Tobit model. Corresponding formulas are given for censoring and truncation on the right, and both on the left and on the right.

5.3.1. Truncated normal distribution [21]

Let y denote the observed value of the dependent variable. Unlike the normal regression, y is the incompletely observed value of a latent depended variable y^*. Recall that with truncation, our sample data are drawn from a subset of a large population. In effects with truncation from below, we only observe $y=y^*$ if y^* is larger than truncation point τ. In effect, we lose the observation on y^* that are smaller or equal to τ when this is the case, we typically assume that the variable y/y >τ follows a truncated normal distribution. Thus, if a continuous random variable y has pdf f(y) and τ is constant. Then we have:

$$f\left(y / y>\tau\right)=\frac{f(y)}{P(y>\tau)} \tag{7}$$

We know that

$$P\left(y>\tau\right)=1-\Phi\left(\frac{\tau-\mu}{\sigma}\right)=1-\Phi\left(\alpha\right) \tag{8}$$

Where $\alpha=\frac{\tau-\mu}{\sigma}$ and Φ(.) is the standard normal cdf. The density of the truncated normal distribution is

$$f\left(y / y>\tau\right)=\frac{f(y)}{1-\Phi(\alpha)}=\frac{\frac{1}{\sigma}\phi(\frac{y-\mu}{\sigma})}{1-\Phi(\frac{\tau-\mu}{\sigma})}=\frac{\frac{1}{\sigma}\phi(\frac{y-\mu}{\sigma})}{1-\Phi(\alpha)} \tag{9}$$

Where Φ(.) is the standard normal pdf.

The likelihood function for the truncated normal distribution is

$$L=\prod_{i-1}^{N}\frac{f(y)}{1-\Phi(\alpha)} \tag{10}$$

Or

$$\ln L=\prod_{-1}^{N}(\ln\left[f(y)\right]-\ln\left[1-\Phi\left(\alpha\right)\right])$$

5.3.2. Censored normal distribution [21]

When a distribution is censored on the left, observations with values at or below τ are set to τ_y

$$y=\begin{cases} y^* \text{ if} \left| y^* > \tau \right. \\ \tau_y \text{ if} \left| y^* \le \tau \right. \end{cases} \tag{11}$$

The use of τ and τ_y is just a generalization of having τ and τ_y set as 0. If a continues variable y has a pdf f(y) and τ is constant, then we have

$$f\left(y\right)=\left[f(y^*)\right]^{d_i}\left[F(\tau)\right]^{1-d_i} \tag{12}$$

In other words, the density of y is the same as that for y^* for y >τ and is equal to the probability of observation of y^* < τ if y=τ. d is an indicator variable that equals 1 if y >τ. The observation is uncensored and is equal to 0 if y = τ the observation is censored.

$$\mathrm{P}\left(\text{censored}\right)=\mathrm{P}\left(y^* \le \tau\right)=\Phi\left(\frac{\tau-\mu}{\sigma}\right)=1-\Phi(\frac{\mu-\tau}{\sigma}) \tag{13}$$

And

$$\mathrm{P}\left(\text{uncensored}\right)=1-\Phi\left(\frac{\tau-\mu}{\sigma}\right)=\Phi(\frac{\mu-\tau}{\sigma}) \tag{14}$$

Thus, the likelihood function can be written as

$$L=\prod_{i}^{N}\left[\frac{1}{\sigma}\phi(\frac{y-\mu}{\sigma})\right]^{d_i}\left[1-\Phi(\frac{\mu-\tau}{\sigma})\right]^{1-d_i} \tag{15}$$

5.4. The Marginal Effect [11]

The estimated (β_k) vector shows the effect of (x_k) on (y_i). Thus, coefficients and the marginal effect of a variable is the effect of a unit change of this variable on the probability P(Y = 1|X = x), given that all other variables are constant, The slope parameter of the linear regression model measures directly the marginal effect of the variable on the other variable. There are three possible Marginal Effect

1. Marginal effect on the latent dependent variable, y^*:

$$\frac{\partial E[y^*]}{\partial x_k}=\beta_k \tag{16}$$

Thus, the reported Tobit coefficients indicate how a one-unit change in an independent variable alters the latent dependent variable.

2. Marginal effect on the expected value for y for uncensored observations:

$$\frac{\partial E[y \mid y>0]}{\partial x_k}\beta_k\left\{1-\lambda(\alpha)\left[\frac{X_i\ \beta}{\rho}+\lambda(\propto)\right]\right\} \tag{17}$$

3. Marginal effect on the expected value for y (censored and uncensored):

$$\frac{\partial E[y]}{\partial x_k}=\Phi(\frac{X_i\,\beta}{\rho})\beta_k \tag{18}$$

TABLE 1: Samples are taken from (300) patients

ID	y: Blood pressure rate above 93.33 mmHg	X1: Age	X2: Blood urea	X3: BMI	X4: Waist circumference
1	96.67	28	132	17.82	110
2	105	31	140	21.8	65
3	86.54 (0)	35	50	35.16	32
4	104.67	55	70	28.72	102
5	45.66 (0)	32	35	28.96	101
6	90.41(0)	40	38	20.81	48
7	126.67	51	260.6	42.97	120
8	123.33	20	80	22.23	90
9	45.15(0)	25	36	32.53	115
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
98	96.67	32	148	21.91	60
99	77.88 (0)	33	47	24.56	60
100	123.33	77	113	25	97
.	.	.	.	.	.
.	.	.	.	.	.
197	100	40	172.4	16.89	100
198	96.67	35	147.8	37.5	60
199	93.33	73	195.4	20.9	60
200	88.99 (0)	34	67	21.12	55
.	.				
.	.				
299	93.33	65	129.7	20.68	37
300	146.67	50	120	25.71	112

TABLE 2: Descriptive statistics of dependent and independent variables in the study

Blood pressure (Y)	Min	Max	Mean	Median
	46.15	**146.67**	**82.25**	**100.00**
Age (X1)	18.00	87.00	51.39	49.00
Blood urea (X2)	17.40	404.00	118.86	117.00
BMI (X3)	13.84	42.97	25.46	23.44
Waist circumference (X4)	30.0	150.0	68.4	60.0

TABLE 3: Results of censored regression model: censored (formula=Y~X, left=0, right=Infinity, data=my data)

Coefficients	Estimate	Std. error	t value	Pr(>t)
Intercept	-5.85378	15.93610	-0.367	0.713
Age	0.80211	0.17584	4.562	5.08e-06 ***
Blood urea	0.48305	0.04437	10.886	2e-16 ***
BMI	-0.44822	0.45381	-0.988	0.323
Waist circumference	-0.06718	0.09777	-0.687	0.492

Total (*n*=300 observations, Left-censored=66 observations, Uncensored=234 observations, left censored (Y<93.33 then Y*=0: observation)

6. RESULTS

In this part, results will be presented to the applied side of the study using statically package (R program) version 3.6.1 and Stratigraphy program version 11.

TABLE 4: Results of the truncated regression model

Coefficients	Estimate	Std. error	*t* value	Pr(>t)
Intercept	3.669531	14.627314	0.2509	0.8019
Age	0.754674	0.160378	4.7056	2.531e-06 ***
Blood urea	0.432102	0.039962	10.8128	2.2e-16 ***
BMI	-0.429849	0.417015	-1.0308	0.3026
Waist circumference	-0.061769	0.089385	-0.6910	0.4895

Table 1 shows that a sample is taken from (300) patients with kidney diseases in dialyzes center in Shar Hospitals. The two levels of BP; high and low BP from the patients (as dependent variables) and some independent variables (age, blood, urea, BMI, and Waist circumference) were taken. We found the average of BP by MAP equation that is contain each (high and low) BP, we could not take high and low BP separately because we determined threshold point equal to 93.33 founded by MAP equation, equal to 12/8 mmHg which is a normal range of BP.

6.1. Descriptive Statistics of Dependent and Independent Variables

Table 2 shows all measures of descriptive statistics. The descriptive statistics give an overview of working with the minimum, maximum, mean, and median of (Age, blood urea, BMI, Waist Circumference), and the results are 18,

TABLE 5: Fitting multiple regression model (OLS) using Stratigraphy program

Model	Unstandardized confections		Standardized confections	*t*	Sig.	Collinearly statistic	
	B	Std. error	Beta			Tolerance	VIF
(Constant)	14.236	12.512		1.138	0.256		
Age	0.647	0.140	0.217	4.630	0.000	0.923	1.083
Blood urea	0.388	0.034	0.535	11.325	0.000	0.913	1.095
BMI	-0.315	0.358	-0.040	-0.881	0.379	0.982	1.019
Waist circumference	-0.048	0.077	-0.029	-0.627	0.531	0.976	1.024

TABLE 6: Model summery

Model	R	R square	Adjusted R square	Std. error of the estimate	Change statistics				
					R Square change	F change	df1	df2	Sig. F change
1	0.632a	0.399	0.391	34.85585	0.399	49.015	4	295	0.000

a. Predictors: (Constant), Waist Circumference, BMI, Blood Urea, Age

17.40, 13.84, and 30 respectively. The max numbers of those variables are 87, 404, 42.97, and 150 respectively. The mean and median of all independent variables are 51.39, 118.86, 25.46, 68.4, 49.00, 117.00, 23.44, and 60.0 respectively.

6.2. Fitting Tobit Model (Censored and Truncated): Regression Model Using Statically Package (R Program)

Table 3 shows that *P*-value: 2.22e-16 and Log-likelihood: -1278.455on 6 Df, wald- statistics 173.8 on 4 Df, Akaike information criterion (AIC)=2566.91, AIC={-2(log-likelihood)+2K}, where K is the number of model parameter plus the intercept. Log-likelihood is a measure of model fit the higher the number the better the fit, and the minimum AIC is the score for the best model. Mean square error (MSE)=0.9305

Table 4 shows that Log-likelihood= -1476.9 on 6 Df, and the AIC=2963.8. MSE=0.993.

From the output of Tables 5-7 shows that the results of fitting a multiple linear regression model to describe the relationship between BP and 4 independent variables. Since the *P*-value in the ANOVA table is <0.05, there is a statistically significant relationship between the variables at the 95.0% confidence level. Table 6 represent the R-Squared statistic indicates that the model as fitted explains 39.9258% of the variability in BP. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 39.1112%. The standard error of the estimate shows the standard deviation of the residuals to be 34.8559. Table 7 shows the analysis Variance of dependent and independent variables.

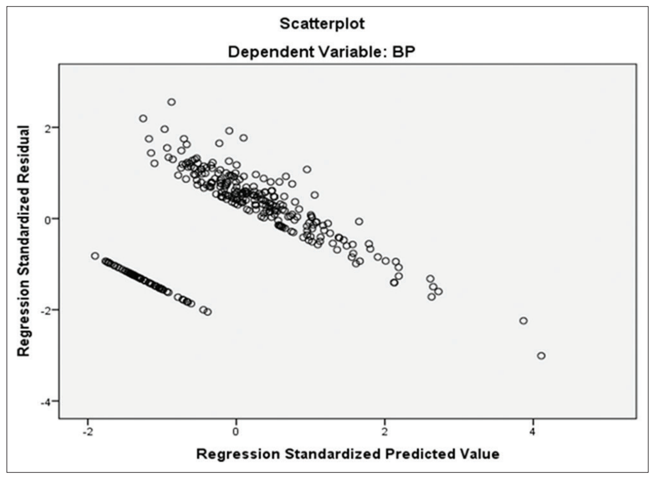

Fig. 4. Standardized residual for multiple regression model using (ordinary least square).

Fig. 4 is a standardized residual for multiple regression models using (OLS). It is clear that the (OLS) method not a suitable method when data censored.

7. DISCUSSION

Analyzing medical data with a Tobit model when it has a threshold point; it help experts (doctors and medical staffs) to identify factors affecting blood pressure in patients with kidney failure. In this study, the Tobit model (censored and truncated) regression model, and a multiple regression model with the least square method (OLS) applied to the data size (n=300) for the cases their rates are greater than or equal to (93.33). By taking the hypothesis that the (BP y) depends on the expletory variables (age, blood urea, BMI, and Waist

TABLE 7: Analysis of variance

Source	Sum of squares	Df	Mean Square	F-Ratio	P-Value
Model	238198.378	4	59549.595	49.015	0.000[b]
Residual	358404.477	295	1214.930		
Total (Corr.)	596602.856	299			

b. Dependent Variable: Blood Pressure

TABLE 8: Results of marginal effects

Coefficients	Marg. eff.	Std. error	*t* value	Pr(>t)
Age	0.772056	0.168918	4.5706	7.17e-06 ***
Blood urea	0.464948	0.042204	11.0168	2.2e-16 ***
BMI	-0.431430	0.436737	-0.9878	0.3240
Waist circumference	-0.064662	0.094095	-0.6872	0.4925

circumference) and comparing their results, the following important points are concluded below.

The result in Table 2 shows all measures of descriptive statistics. The descriptive statistics give an overview of working with the minimum, maximum, mean, and median of (age, blood urea, BMI, and Waist Circumference), and the results are 18, 17.40, 13.84, and 30, respectively. The maximum numbers of those variables are 87, 404, 42.97, and 150, respectively. The mean and median of all independent variables are 51.39, 118.86, 25.46, 68.4, 49.00, 117.00, 23.44, and 60.0, respectively.

The results of analysis censored regression model in Table 3 show the final result with all significant variables for the phenomenon study, the results of parameter estimation and t value analysis, the significant factors affecting BP. P = 2.22e-16 and Log-likelihood= -1278.455 on 6 Df, Wald statistic= 173.8 on 4 Df, AIC=2566.91. The Log-likelihood is a measure of the model fit the higher number of it is a better fit. The minimum AIC is the score for the best model. The MSE is equal to 0.9305. We know that the (β) is the relationship between the response variable and covariates, if (+β) it means a positive relationship and if (-β) means a negative relationship. Through the result in Table 3 appear the relationship between the variables (age and blood urea) is positive because the variables have a positive relationship with the dependent variables (BP) and those variables (Age and blood urea) have highly significant effects on BP. Furthermore, the relationship between (BMI) and BP is negative. If there is an increase in (BMI) by one unit the (BP) decreases by (-0.44822). The factors (BMI and Waist circumference) appeared to have no significant effects on BP.

From Tables 3-8 show that the censored regression model for the samples is a more suitable model than other regression models (Truncated, Marginal, and Multiple). This result found by comparing their AIC, log-like values, and MSE.

The censored with the marginal effects from Table 8 shows that the two variables (age and blood urea) have highly significant effects. The changes in years make BP significantly increasing by 0.77%. This means that the effect of age for any case in the sample with Std. error is by 0.16%. Furthermore, one unit of blood urea for each point increases by 0.46% with stander error (0.04).

In the result of multiple regression models, using (OLS) method, we detected that since the p-value in the ANOVA table is <0.05, there is a statistically significant relationship between the variables at the %95 confidence interval the R-square statistic indicates that the model as fitted explains 0.39 of the variability BP. And theoretically, as defined, the OLS (unconditional estimates) are bias.

8. CONCLUSION

In this study, both Tobit regression analysis and OLS analysis were used for studying factors affecting the BP. In this work, the data collected from 300 patients in a dialysis center at Shar hospital in Sulaimani city. The two levels of BP; high and low from the patients (as dependent variables) and some independent variables (age, blood, urea, BMI, and Waist circumference) were taken. Each patient has own specific BP (high and low). Then, we could not take high and low BP separately for our study. That is why the MAP was performed. It is an average arterial pressure contains high and low BP. When studying BP as a dependent variable, we find that variable data are censored at zero. In this case, the Tobit model is most suitable model to use. It was found that the two factors (age and blood urea) have highly significant effects on BP. However, the two variables (BMI and Waist circumference) appeared to have no effects on the dependent variable. The comparison of the result from Tobit and OLS estimations shows that biased can result when estimation BP using OLS if BP restricted at the threshold point

REFERENCES

[1] T. Amemiya. "Tobit models: A survey". *Search Results Journal of Economics*, vol. 24, no. 1-2, pp. 3-61, 1984.

[2] W. Wang and M. E. Griswold. "Natural interpretations in Tobit regression models using marginal estimation methods". *Statistical Methods in Medical Research*, vol. 26, no. 6, pp. 2622-2632, 2017.

[3] D. C. Crews, A. K. Bello and G. Saadi. “2019 World kidney day editorial-burden, access, and disparities in kidney disease”. *Brazilian Journal of Nephrology*, vol. 41, pp. 1-9, 2019.

[4] R. A. Preston, I. Singer, and M. Epstein. “Renal parenchymal hypertension: Current concepts of pathogenesis and management”. *Archives of Internal Medicine*, vol. 156, no. 6, pp. 602-611, 1996.

[5] J. A. Staessen, Y. Li, A. Hara, K. Asayama, E. Dolan and E. O’Brien. “Blood pressure measurement anno 2016”. *American Journal of Hypertension*, vol. 30, no. 5, pp. 453-463, 2017.

[6] R. N. Kundu, S. Biswas and M. Das. “Mean arterial pressure classification: A better tool for statistical interpretation of blood pressure related risk covariates”. *Cardiology and Angiology: An International Journal*, vol. 6, no. 1, pp. 1-7, 2017.

[7] D. Yu, Z. Zhao and D. Simmons. “Interaction between mean arterial pressure and HbA1c in prediction of cardiovascular disease hospitalisation: A population-based case-control study”. *Journal of Diabetes Research*, vol. 2016, p. 8714745, 2016.

[8] C. Wilson and C. A. Tisdell.“OLS and Tobit estimates: When is substitution defensible operationally?” In: *Economic Theory, Applications and Issues Working Papers*, University of Queensland, School of Economics, Queensland , 2002.

[9] M. H. Odah, A. S. M. Bager and B. K. Mohammed. “Tobit regression analysis applied on Iraqi bank loans”. *American Journal of Applied Mathematics and Statistics*, vol. 7, no. 4, p. 179, 2017.

[10] A. Prahutama, A. Rusgiyono, M. A. Mukid and T. Widiharih. “Analysis of Household Expenditures on Education in Semarang City, Indonesia Using Tobit Regression Model”. In: *E3S Web of Conferences*, vol. 125, p. 9016, 2019.

[11] N. M. Ahmed. “Limited Dependent Variable Modelling (Truncated and censored Regression models) with Application”. Vol. 7377. Cambridge University Press, New York, pp. 82-96, 2018.

[12] M. F. Ahmad, M. Ishtiaq, K. Hamid, M. U. Khurram and A. Nawaz. “Data envelopment analysis and Tobit analysis for firm efficiency in perspective of working capital management in manufacturing sector of Pakistan”. *International Journal of Economics and Financial Issues*, vol. 7, no. 2, pp. 706-713, 2017.

[13] S. Samsudin, A. S. Jaafar, S. D. Applanaidu, J. Ali and R. Majid. “Are public hospitals in Malaysia efficient? An application of DEA and Tobit analysis”. *Southeast Asian Journal of Economics*, vol. 4, no. 2, pp. 1-20, 2016.

[14] M. H. Odah, A. S. M. Bager and B. K. Mohammed. “Studying the determinants of divortiality in Iraq. A two-stage estimation model with tobit regression”. *International Journal of Applied Mathematics and Statistics*, vol. 7, no. 2, pp. 45-54, 2018.

[15] P. Zorlutuna, N. A. Erilli and B. Yücel. “Lung cancer study with tobit regression analysis: Sivas case”. *Eurasian Eononometrics, Statistics and Emprical Economics Journal*, vol. 3, no. 3, pp. 13-22, 2016.

[16] P. C. Anastasopoulos, A. P. Tarko and F. L. Mannering. “Tobit analysis of vehicle accident rates on interstate highways”. *Accident Analysis and Prevention*, vol. 40, no. 2, pp. 768-775, 2008.

[17] A. Henningsen. “Estimating censored regression models in R using the censReg Package”. *R Packag Vignettes*, Vol. 5. University of Copenhagen, Copenhagen, p. 12, 2010.

[18] A. C. Michalos. *Encyclopedia of Quality of Life and Well-Being*, Springer, Berlin, 2014.

[19] M. H. Odah. “Asymptotic least squares estimation of tobit regression model. An application in remittances of Iraqi immigrants in Romania”. *International Journal of Applied Mathematics and Statistics*, vol. 8, no. 2, pp. 65-71, 2018.

[20] C. Ekstrand and T. E. Carpenter. “Using a tobit regression model to analyse risk factors for foot-pad dermatitis in commercially grown broilers”. *Preventive Veterinary Medicine*, vol. 37, no. 1-4, pp. 219-228, 1998.

[21] J. S. Long. *Regression Models for Categorical and Limited Dependent Variables*. Vol. 7. Sage Publications, Thousand Oaks, 1997.

[22] A. Flaih, J. Guardiola, H. Elsalloukh and C. Akmyradov. “Statistical inference on the ESEP tobit regression model”. *J. Stat. Appl. Probab. Lett.*, vol. 6, pp. 1-9, 2019.

[23] B. R. Humphreys. “Dealing with zeros in economic data”. *University of Alberta, Department of Economics*, vol. 1, pp. 1-27, 2013.

[24] K. A. M. Gajardo. “*An Extension of the Normal Censored Regression*”. Pontificia Universidad Catolica De Chile, Santiago, Chile, 2009.

[25] W. H. Greene. *Limited Dependent Variables Truncation, Censoring and Sample Selection*. Sage, Thousand Oaks, CA, 2003.