Determine important variables which affecting on abortion by using method of Tobit quintile regression

Fadel hamid hadi alhusseini

Department of Statistics and Informatics Economic

, university of Craiova,Romania

Email:fadhelfadhel222@yahoo.com

assistant lecturer Saif Hosam Raheem

Department of Statistic

Al-Qadiseya University , Iraq

saifhosamrr@yahoo.com

Abortion is considering a natural method for get rid the womb from fetal at happening health problem. Such as fetal growth by abnormal picture, stop fetal growth inside the womb etc.

Global statistics is indicating that abortion happening by rate (15%-20%), per pregnancy case. Majority, abortion cases are happening in first three months of pregnancy. Phenomenon under study is contain one response variable represents a number of abortions per woman through her lifetime. Some women arrive abortion cases through her lifetime to four or five case. and others from women arrive the abortion cases through her lifetime to zero. Therefore, the response variable in this study is censored at zero. So Tobit regression model is fit for this phenomena. But study these phenomena at various locations from spread data of response variable through using Tobit quantile regression. Which, it is specialized by study quantile regression at left censored response variable. In this paper we will use Tobit quantile regression is proposed by Powell (1986) through use programming [R] package, the function ("cqr"). This function existing within (Package ‘quantreg’).

Introduction

Abortion is one of a very common thing, that the possibility of abortion rate is almost one case in every five cases pregnancies, this percentage decreases in the final stages of pregnancy, where one case of abortion approximately happens in every 100 pregnancies. There are some factors effect on abortion directly where the others of factors have influenced indirectly.

One of these factors is the mother's age during pregnancy. whenever the mother getting older, the possibility of abnormalities in chromosomes occurs to the embryo that effects on the abortion negatively. So that in the 30s the possibility of abortion reaches to one (1) of every five pregnancies. And at 42 age and older it becomes one case of every two cases.

The mother's weight during the pregnancy, any increasing or decreasing in the mother's weight during the period of pregnancy influences on increasing the possibilities of abortion, also the increasing or decreasing the sugar blood or Low blood pressure adversely effects on health of the fetus and thus the occurrence of abortion. Moreover the economical factors have obvious role in abortion cases, for instance the household income influence on abortion because the lack of income will directly reflect on the reduces the health care for the mother during pregnancy period through the reduction of the mother visiting to her doctor, also the lack of household income will increase the mothers hours work during pregnancy, or the low income effects on the kind of feeding for the mother in pregnancy period, all these factors effect on the heaths mother as a result will increase from the abortion cases .

As well as the number of births has been an extrusive impact on abortion, the wrong use of drugs increases the risk of occurring the abortion, as well as the Toxoplasmosis disease has a direct impact on the incidence of abortion or hormone (progesterone) that has a negative effect on the occurring of abortion.

Most of the factor mentioned above have been functioned as (independent variable) in this paper, while the (response variable) represents the number of abortions cases during the woman life, we notice that this variable (uncensored) during the abortion cases are more than Zero (0) and this may be only one or two or even seven cases. (as in research sample under study) through describing this variable, so the (Tobit regression model) Is to be optimized.

But when we are studying the Tobit model with a set of lines according to different quantile rations, the model of Tobit quantile regression is to be the suitable model. In this research we will depend on the model of Tobit quantile regression to study the influence of the independent variable on the response variable during various quantile ratios, in order to Estimate the parameters of this model we will rely the function (crq) which it has been found within the ((Package

'quantreg')). While the other parts of the research is organized as a follow, the second chapter (Methodology of Tobit Quantile Regression), while the third chapter will introduce the sample research and the variables study. In the fourth chapter we will explain the analysis of the results of study, the conclusions will be expressed in the fifth chapter.

2-Methodology of Tobit Quantile Regression

It is known that the main objective of analyzing the regression models is to focus on the relationship between the response variable and a set of independent variables but the nature of used data, imposes on the researcher the suitable model for these data and particularly the model of response variable. So If the data of response was from the continuous variable type and have a linear relationship with the independent variables, so that the liner regression model is optimized for these data. If the variable data binary response models the probit regression model is considered the optimized for these type of data. but if the data of variable response censored in a certain point therefore the censored regression model is considered the Appropriate that takes the following mathematical formula

$$Y = \begin{cases} c & if\ g^* \leq c \\ g^* & if\ g^* > c \end{cases} \qquad (1)$$

$$where:\ g^* = \alpha + x_i^T \beta + e_i, \qquad i = 1, \dots\dots, n$$

Where C is the censored point and g^* is latent variable and (β, β_0) are coefficients of latent variable, x_i^T is a set of independent variable , e_i is a term random error that distributes according to normal distribution in zero (0) mean and σ^2 variance.

Y is a response variable where in this model the data of the approved variable are two groups, on censored at point C and free group of various quantities

but if the quantity of censored point equals to zero (0) (C=0) in this case we will get with Tobit regression model , that considers the optimal model for this data which the response variable in it is censored at (0) point that takes the following in mathematical formula

$$Y = \begin{cases} 0 & if\ g^* \le 0 \\ g^* & if\ g^* > 0 \end{cases} \qquad (2)$$

$$where:\ g^* = \alpha + x_i^T\beta + e_i, \qquad i = 1, \dots\dots, n$$

Also its possible to express the Tobit regression model by another mathematical formulae

$$Y = \max(0,\ g^*) \qquad (3)$$

According to the equation above we find that the Tobit regression model is a special state of censored regression model see (jemes tobin 1958,wiliim 2007).

If there is a desire for the researcher to study the Tobit regression model at different quantile regression lines the Tobit quantile regression model considers the appropriate regression model that was suggested by Poll in (1986).

In recent times this model has become important in various applications such as scientific, medical, economic, engineering, psychological, and others. The mathematical formula for this model is

$$\min_{\alpha_t,\beta_t} = \sum_{i=1}^{n} \rho_\tau \left(y_i - max\{C, \alpha_\tau + x_i^T\beta_\tau\}\right) \qquad (4),$$

Since τ value including within open group of (0,1) through mathematical logic there is an unlimited Tobit quantile regression lines. Consequently, for every data of particular phenomena, there is a line of quantile informative in analyzing this data, but identifying this line is a difficult task. The equation is called (3) check function that proposed by Koenker and Bassett (1978) Where we can reformulate the equation (3) by the following formula

$$\rho_\tau(e_i) = \begin{cases} (\tau)\,|\, y - max\left(0, \alpha_\tau + x_i^T\beta_\tau\right)| & y \ge max\left(0, \alpha_\tau + x_i^T\beta_\tau\right) \\ -(1-\tau)\,|\, y - max\left(0, \alpha_\tau + x_i^T\beta_\tau\right)| & y < max\left(0, \alpha_\tau + x_i^T\beta_\tau\right) \end{cases}$$

From equation (5) The weighted deviations between the real values and the (predictive values), the positive deviations take τ weight and negative deviations take (1-τ) weight,

Therefore, the predictive equation for Tobit quantile regression above can be reformulate by another way

$$\sum d_\tau\,(y_i, \widehat{y_i}) = \tau \sum_{y_i \geq \alpha_\tau + x_i^T \beta_\tau} \left| y_i - \alpha_\tau + x_i^T \beta_\tau \right| + (1-\tau) \sum_{y_i < \alpha_\tau + x_i^T \beta_\tau} \left| \alpha_\tau + x_i^T \beta_\tau \right| \quad (6)$$

Where d_τ represents the distance between (values observation) and Tobit quantile regression line.

Where the first part of the equation (6) represents a vertical distance above the Tobit quantile regression line and the second part of the equation (6) represents the vertical distance under the Tobit quantile regression line See (Koenker 2005).

Without forgetting that Tobit quantile regression has a set of properties made it at the front of regression models where this model works without achieving the required assumptions in the regression models and one of these requires that this model does not need to achieve the hypothesis have said that the random error must be distributed normally and even the error limit in Tobit quantile regression model is free distributed , also the Tobit quantile regression model is robust against outlier, in another word that this model does not affected by the presence of outlier values.

One of its important properties it can estimate the range of the regression lines in order to provide complete coverage for all data that appears in the study, where can estimate a large group of Tobit quantile regression line at different quantile ratios of set within a group of $\tau \in (0,1)$

unfortunately, the equation (4) is (not differentiable) at the zero point (0) where you cannot find a solution to this equation. But (Koenker and D'Orey, 1987) mentioned that it can be minimized for this function through a linear programming algorithm, you can also estimate the parameter of Tobit quantile regression model by Bootstrapping method which was introduced by)Mooney, 1993).

There was appeared many algorithms for estimating the parameter of Tobit quantile regression model but most of these algorithms were inefficient when the phenomenon study data contribute on a large group of censored value which represented 0 in Tobit model.

So we are going to adopt in estimating the parameter of this model on algorithm was made by the researcher Fitzenberger (1996) which considers more efficient than previous algorithms through t using (cqr) founding within Package 'quantreg' see (Koenker 2016).

3- The research sample and study variables

Data were collected from educational of the Women's Hospital and children in Diwaniya, where the size of the sample under study (300) Show, as the study sample containing one response variable represents the number of abortion times. The time of abortions differs from one woman to another. Sometimes abortion cases for some women reach to more than one case in all her life and this may reach to 8 abortion cases but other women never have any abortion case during her life , through describing the variable above then the response variable is been censored in zero (0)

Also, the number of independent variables involved in this study are:

X_1 maternal age

X_2 mothers weight

X_3 mothers blood pressure during pregnancy

X_4 mothers blood sugar during pregnancy

X_5 the number of mothers births

X_6 monthly income of the family

X_7 the number of mother's hours working during the pregnancy

X_8 hormone progesterone

X_9 using the wrong medication

X_{10} Toxoplasmosis

In order to analyze the phenomenon study data, it will have used the function (crq) contained within (Package 'quantreg').). Also in this paper we will use seven quantile lines (as desired by the researcher) at the following rates (0.13,0.25,0.38,0.50,0.63,0.75,0.88) where these ratios determined by the following mathematical function

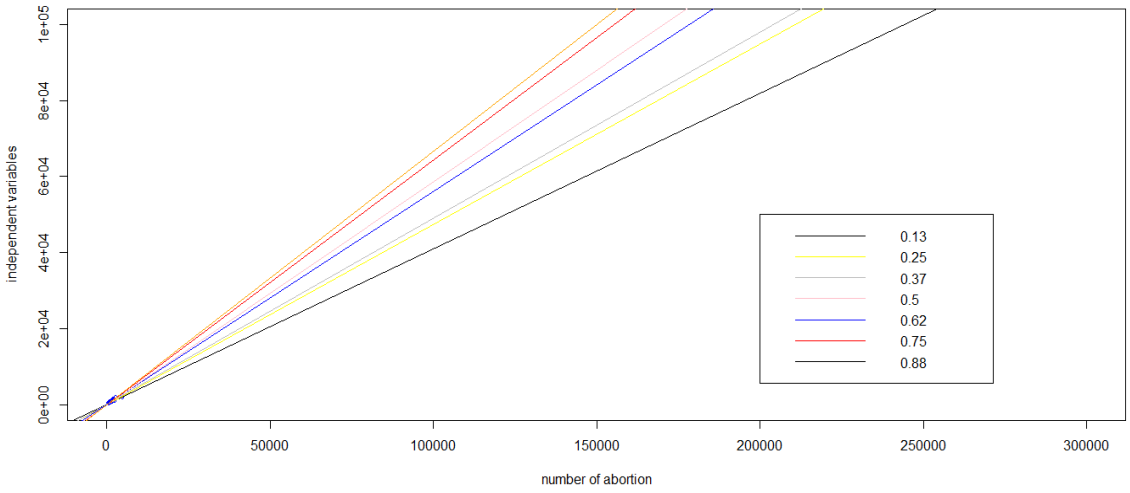
$$\tau_q = \frac{q}{Q+1} \text{ for } q = 1, \dots, Q. \qquad (7)$$

So the Tobit quantile lines estimate as a follow

$$\tau_1 = \frac{1}{Q+1} = \frac{1}{8} = 0.13,\ \tau_2 = \frac{2}{8} = 0.25,\ \tau_3 = \frac{3}{8} = 0.38,\ \tau_4 = \frac{4}{8} = 0.50$$

$$,\tau_5 = \frac{5}{8} = 0.63,\ \tau_6 = \frac{6}{8} = 0.75,\ \tau_7 = \frac{7}{8} = 0.88$$

Where it can represent the Tobit regression quantile at rations above through the following descriptive chart:



Form (1) represents Tobit quantile regression lines at supposed ratios

As previously we mentioned that the number of Tobit quantile regression lines were used in this research seven Tobit quantile that previously mention as follow:

- Table (1) shows the results of the analysis Tobit quantile regression model at quantile ratio (0.13)

The Tobit quantile ratio =0.13				
Variables	**Coefficients**	**Std. Error**	**t value**	**Pr(>|t|)**
Intercept	0.06221	0.01222	5.09152	0.00000
X_1	0.06179	0.01245	4.96449	0.00000
X_2	0.01708	0.00619	2.76042	0.00614
X_3	0.42676	0.15906	2.68301	0.00772
X_4	0.39209	0.15017	2.61106	0.00950
X_5	0.12464	0.04921	2.53290	0.01184
X_6	0.00025	0.00018	1.35836	0.17540
X_7	0.08081	0.03718	2.17349	0.03055
X_8	0.00445-	0.01442	0.30889-	0.75763

X_9	0.18246	0.20698	0.88152	0.37876
X_{10}	0.18232	0.20919	0.87152	0.38419
The pseudo-R square = 0.746388				

From the results of Table (1), we find that the ten independent variables mentioned earlier can interpret (74.63%) of variation in the response variable (the number of abortion cases) shows that this indicator ten independent variables have a clear force in the interpretation of contrast variable response. We find that all the independent variables have a direct correlation with the response variable (the number of abortions) except one variable is X_8 progesterone hormone It has an inverse relationship, according to this measure, all the results are logical in the medical side.

We find that the number of non-proper independent variables are (4) independent variables, and the remaining variables were of a proper effect on the response variable (number of abortion cases) and these proper numbers are

X_1 Mothers age here we find that the relationship of this variable with response variable (the number of abortions) is a direct proper correlation Thus, the increasing age of the mother is incremented by one unit raises the case of an abortion about (0.06179).

X_2 mothers weight the relationship of this variable with response variable (the number of abortion cases) is proper positive correlation, thus the increasing of mother's blood one unit rises abortion cases to about (0.42676).

X_4 mothers blood sugar during pregnancy, the relationship of this variable with response variable (the number of abortion cases) it's a proper **proportional** relation as a result increasing the mother's births numbers have a **proportional** proper relation thus the increasing of mother's sugar blood during pregnancy one unit rises abortion cases about (0.12464).

X_5 the number of mothers births: the relationship of this variable with response variable (the number of abortion cases) is proportional relation therefore the increasing in numbers of mothers births one unit rises the abortion cases into about (0.08081).

X_7 the number of mother's hours working during the pregnancy this variable has a proper proportional relation with response variable (number of abortion cases) thus the increasing the works hours during pregnancy period one unit rises the occurs of abortion to about (0.08081).

- Analyzing Tobit quantile regression model at quantile ratio (0.25)

- Table (2) explains the results of Tobit quantile regression model at quantile ratio (0.25).

The Tobit quantile ratio =0.25

Variables	**Coefficients**	**Std. Error**	**t value**	**Pr(>|t|)**
Intercept	1.19874	3.15160	0.38036	0.70396
X_1	0.05511	0.01151	4.78803	0.00000
X_2	0.02181	0.00468	4.65587	0.00000
X_3	0.43958	0.14203	3.09503	0.00216
X_4	0.34028	0.10948	3.10813	0.00207
X_5	0.10780	0.05443	1.98044	0.04860
X_6	0.00013	0.00018	0.71527	0.00018
X_7	0.08701	0.02545	3.41924	0.00072
X_8	0.01501	0.01194	1.25746	0.20960
X_9	0.33974	0.18144	1.87244	0.06215
X_{10}	0.28274	0.17817	1.58687	0.11363
The pseudo-R square = 0.3737121				

Through the value of The pseudo-R squar , we find the tenth independent variables can intrpert (37.37%) of variation in response variables (the number of abortion), In other words, these independent variables cannot explain (62.63%), in variation of independent variable thus the Tobit quantile regression line at (0.25)ratio is weak in interpretation of data under this study . Also, we find that all these ten variables have a correlation relation with response variable (the number of abortion cases). We find the number of proper variables are eight (8) independent variables of ten , as a result there are two proper variables(X_8 progesterone hormone and X_{10} Toxoplasmosis).

- Analyzing Tobit quantile regression model at a quantile ratio (0.37).
- Table no. (3) shows the results of Tobit quantile regression model at a quantile ratio (0.37).

The Tobit quantile ratio =0.37				
Variables	**Coefficients**	**Std. Error**	**t value**	**Pr(>|t|)**
Intercept	4.92199	0.43579	11.29449	0.00000
X_1	-0.00549	0.00833	-0.65900	0.51042
X_2	0.00022	0.00406	0.05502	0.95616
X_3	0.13450	0.13636	0.98629	0.32481
X_4	0.09830	0.12863	0.76424	0.44535
X_5	0.15607	0.06011	2.59645	0.00990
X_6	0.00009	0.00017	0.53134	0.59559
X_7	-0.00344	0.02127	-0.16182	0.87156
X_8	-0.00015	0.00889	-0.01733	0.98619
X_9	-0.09851	0.18323	-0.53761	0.59126
X_{10}	-0.11485	0.15462	-0.74279	0.45821
The pseudo-R square = 0.24184094				

Through the value of The pseudo-R square we find the ten independent variables can interpret (21.18%) of variation in response variables (the number of abortion cases) , In other words, these independent variables cannot explain (75.82%), in variation of independent variable thus the Tobit quantile regression line at (0.37)ratio is weak in interpretation of data under this study . Also we find the variables (X_1 mothers age, X_7 the number of mother's hours working during the pregnancy, X_8 hormone progesterone, X_9 using the wrong medication and X_{10} Toxoplasmosis) are variables with proper effect , and the number of non-proper variables are(4) variables as the following table .

-analyzing Tobit quantile regression model at quantile ratio (0.50)

Table No. (5) shows the results of analyzing Tobit quantile regression model at quantile ratio (0.50)

		The Tobit quantile ratio =0.50		
Variables	**Coefficients**	**Std. Error**	**t value**	**Pr(>|t|)**
Intercept	4.92979	0.76528	6.44180	0.00000
X_1	-0.00581	0.01373	-0.42346	0.67228

X_2	0.00025	0.00507	0.04968	0.96041
X_3	0.37563	0.15525	2.41947	0.01616
X_4	0.15895	0.15802	1.00594	0.31529
X_5	0.18597	0.07472	2.48893	0.01338
X_6	0.00023	0.00021	1.08185	0.28022
X_7	-0.03496	0.03273	-1.06801	0.28641
X_8	0.00733	0.01674	0.43760	0.66200
X_9	0.01709	0.23417	0.07299	0.94187
X_{10}	-0.06967	0.22200	-0.31382	0.75389
The pseudo-R square = 0.02691362				

Through the value of The pseudo-R square , we find that the ten independent variables can interpret (2.69%) of variation of response variables (the number of abortion cases), in another words this independent variable cannot interpret (97.31%)in variation of response variable , thus the Tobit quantile regression line at (0.50) ratio is very weak in interpretation of the data under study .Also we find that every three(3)variables have inverse relation with response variable (the number of abortion cases) and other variables have coroporational relation and the number of proper variables are two (2) independent variables of ten (10) variables.

- analyzing Tobit quantile regression model at quantile ratio (0.62)

- Table No. (5) shows the results of analyzing Tobit quantile regression model at quantile ratio (0.62)

		The Tobit quantile ratio =0.62		
Variables	**Coefficients**	**Std. Error**	**t value**	**Pr(>|t|)**
Intercept	5.39780	0.88588	6.09312	0.00000
X_1	-0.00863	0.01679	-0.51428	0.60745
X_2	0.00067	0.00540	0.12383	0.90154
X_3	0.26596	0.20822	1.27733	0.20251
X_4	0.16892	0.17893	0.94403	0.34594
X_5	0.21031	0.08137	2.58470	0.01024
X_6	0.00022	0.00024	0.94244	0.34676
X_7	-0.01265	0.03551	-0.35626	0.72191
X_8	-0.00321	0.01951	-0.16432	0.86959
X_9	0.04271	0.24764	0.17248	0.86318
X_{10}	-0.09505	0.26102	-0.36415	0.71601

The pseudo-R square = 0.01107621	

Through the value of The pseudo-R square , we find that the ten independent variables can interpret (1.10%) of variation of response variables (the number of abortion cases), in another words this independent variable cannot interpret (98.90%) in variation of independent variables thus the Tobit quantile regression line at (0.62) ratio is very weak in explaining the data under study . Also we find there are (4) variables have inverse relations with response variables (No. of abortion cases), and other variables have proportional relation, and the number of proper variables are (1) independent variable of (10) variables.

-analyzing Tobit quantile regression model at quantile ratio (0.75)

Table No. (5) shows the results of analyzing Tobit quantile regression model at quantile ratio (0.75)

			The Tobit quantile ratio =0.75	
Variables	**Coefficients**	**Std. Error**	**t value**	**Pr(>|t|)**
Intercept	4.76758	1.18678	4.01725	0.00008
X_1	0.00019	0.02332	0.00836	0.99334
X_2	0.00770	0.01019	0.75603	0.45025
X_3	0.61738	0.25367	2.43378	0.01555
X_4	0.19964	0.26928	0.74137	0.45907
X_5	0.23500	0.11298	2.08003	0.03840
X_6	0.00049	0.00034	1.45775	0.14599
X_7	-0.04603	0.05735	-0.80266	0.42283
X_8	0.00135	0.02369	0.05689	0.95467
X_9	0.38859	0.37104	1.04731	0.29583
X_{10}	-0.34562	0.33208	-1.04078	0.29885
The pseudo-R square =0.04936482				

Through the value of The Pseudo-R square, we find that the ten independent variables can interpret (4.93%) of variation of response variables (the number of abortion cases), in another words this independent variable cannot interpret (95.07%) in variation of independent variable, thus the Tobit quantile regression line at (0.75) ratio is very weak in explaining the data under study. Also we find there are (2) independent variable have inverse relation with response variables

(No. of abortion cases) and other variables have proportional relation, and the number of proper variables are (2) variables of (10) variables.

-analyzing Tobit quantile regression model at quantile ratio (0.88)

Table No. (5) shows the results of analyzing Tobit quantile regression model at quantile ratio (0.88)

The Tobit quantile ratio =0.88				
Variables	**Coefficients**	**Std. Error**	**t value**	**Pr(>|t|)**
Intercept	-0.28277	1.42393	-0.19858	0.84273
X_1	-0.00749	0.03482	-0.21523	0.82974
X_2	0.02194	0.01173	1.87038	0.06244
X_3	0.27291	0.28998	0.94111	0.34743
X_4	0.40536	0.33304	1.21716	0.22454
X_5	0.29010	0.16365	1.77262	0.07734
X_6	0.00097	0.00038	2.58155	0.01033
X_7	-0.03336	0.06707	-0.49745	0.61925
X_8	-0.01465	0.03175	-0.46142	0.64484
X_9	0.30601	0.43422	0.70474	0.48154
X_{10}	-0.18490	0.45830	-0.40345	0.68692
The pseudo-R square =0.05174905				

Through the value of The Pseudo-R square, we find that the ten independent variables can interpret (5.17%) of variation of response variables (the number of abortion cases), in another words this independent variable cannot interpret (94.93%) in variation of independent variable, thus the Tobit quantile regression line at (0.88) ratio is very weak in explaining the data under study. We find that there are (5) independent variables have inverse relation with response variable (No. of abortion cases) and the other variables have proportional relation , and the number of proper variables is one independent variable of (10) variables.

Conclusions

The pseudo-R square for the phenomenon data under study as follow:

- In the first position is Tobit quantile regression line at quantile ratio (0.13)

- In the second position is Tobit quantile regression line at quantile ratio (0.25)
- In the third position is Tobit quantile regression line at quantile ratio (0.37)
- In the fourth position is Tobit quantile regression line at quantile ratio (0.88)
- In the fifth position is Tobit quantile regression line at quantile ratio (0.75)
- In the sixth position is Tobit quantile regression line at quantile ratio (0.50)
- In the seventh position is Tobit quantile regression line at quantile ratio (0.50)

From the previous results we conclude that the best Tobit quantile regression line in representing the phenomenon data under study is at quantile ratio (0.13), where the Tobit quantile regression line has a high interpretation ability and this obvious from the value of The pseudo-R square.

Also we find the results at ratio (0.13) absolutely agree with medical logic.

But the other Tobit quantile regression lines at a various ratio were weak in interpretation the phenomenon data of studied data , and the relations of the independent variable were unstable through these various ratios , also the propriety of these variables were unstable

Resources

Fitzenberger, B. (1996): A Guide to Censored Quantile Regressions," in Handbook of Statistics,

Koenker, R. (2016). quantreg: Quantile regression. R package version 5.26.

Koenker, R. and V. D'Orey (1987). Algorithm AS 229: Computing regression quantiles. Journal of the Royal Statistical Society: Series C (Applied Statistics) 36, 383-393.

Koenker, R. (2005). *Quantile Regression. Cambridge Books*. Cambridge University Press

Greene william. (2007). Econometric analysis. Book. New York University. Seven edition.

Powell, J. (1986). Censored regression quantiles. Journal of Econometrics 32, 143-155.

Tobin, James (1958)"Estimation of Relation ships for limited dependent Variables " Econometrica, January 26, pp24-36 .