LIMITED DEPENDENT VARIABLE MODELS

(CENSORED AND TRUNCATED)

Amrapali Roy Barman

Apurva Dey

Jessica Pudussery

Vasundhara Rungta

INTRODUCTION

The effect of truncation occurs when sample data are drawn from a subset of a larger population of interest. We are concerned with inferring the characteristics of a full population from a sample drawn from a restricted part of that population. Censoring is a more common problem in recent studies. When the dependent variable is censored, values in a certain range are all transformed to (or reported as) a single value. The censoring of a range of values of the variable of interest introduces a distortion into conventional statistical results that is similar to that of truncation.

In economics, such a model was first suggested in a pioneering work by Tobin [1958]. He analysed household expenditure on durable goods using a regression model which took account of the fact that the expenditure (the dependent variable of his regression model) cannot be negative. Tobin called his model the model of limited dependent variables. It and its various generalizations are known popularly among economists as Tobit models, a phrase coined because of similarities to Probit models.

Between 1958, when Tobin's article appeared, and 1970, the Tobit model was used infrequently in econometric applications, but since the early 1970's numerous applications rangingover a wide area of economics have appeared and continue to appear. This phenomenon is clearly due to a recent increase in the availability of micro sample survey data which the Tobit model analyses well.

Some other examples that have appeared in the empirical literature are as follows:

1. The number of extramarital affairs [Fair (1977, 1978)]

2. Charitable contributions [Reece (1979)]

3.The number of arrests after release from prison [Witte (1980)]

4.Annual marketing of new chemical entities [Wiggins (1981)]

5.The number of hours worked by a woman in the labour force [Quester and Greene (1982)]

AIM

The objective of our project is to study empirically the working of the Tobit model - censored and truncated.

For our first model we aim to study the factors affecting the level of pension received, namely- age, experience in years, tenure (years with current employer), education level measured by years of schooling and number of dependents. We use the censored regression model here because in the data many people receive no pension. Since our dependent variable pension takes the value 0 in a large number of cases, the censored regression model is appropriate here.

For our second model, we aim to study the factors affecting the achievement score received by a number of students who have managed to enter a special GATE (gifted and talented education) program which requires a minimum score of 40. This achievement score is modelled as depending on scores received in separate language and mathematics tests taken by all students, even those who have not managed to enter the program. The data is truncated because the only students surveyed are the ones who have entered the program, ignoring those who have scored below 40. Thus the truncated regression model is appropriate here.

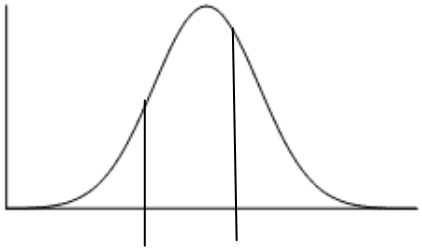
METHODOLOGY AND THEORY

A limited dependent variable Y is defined as a dependent variable whose range is substantively restricted.

$Y_i{}^* = \beta' X_i + u_i$, where $Y_i{}^*$ is a latent variable

$Ui \sim N(0, \boldsymbol{\sigma^2})$

Up till **c** we observe the latent variable and after that, one starts observing the variable itself.



$c\beta' X_i$

In the usual linear regression model we write

$Y_i = \beta' X_i + u_i$ Where $u_i \sim N(0, \boldsymbol{\sigma^2})$

$\Rightarrow X_i \sim N(\beta' X_i, \boldsymbol{\sigma^2})$

$\Rightarrow -\infty < Y_i < \infty$

However in many economics applications Y_i may not or do not satisfy this restriction.Mostly we have $Y_i \geq 0$

Example: Working hours,where $0 \leq Y_i \leq 24$ .More generally, $a \leq Y_i \leq b$.

To handle this problem, there are two methods:

1. **Non linear specification** :
   We write $Y_i = e^{\beta' Xi + ui}$
   Ln Yi= β'Xi + ui
   Now Yi>0
   But $-\infty <$Ln Yi $< \infty$
   we assume$u_i \sim N(0, \sigma^2)$
   $\Rightarrow$Ln $Y_i \sim N(\beta' X_i, \sigma^2)$

   However, inference is a problem in this method because

   $E(Yi) \neq e^{E(\hat{\beta}' Xi + ui)}$

2. **Latent Variable Framework :**
   Write $Y_i = \beta' X_i + u_i$ if $\beta' X_i + u_i > 0$

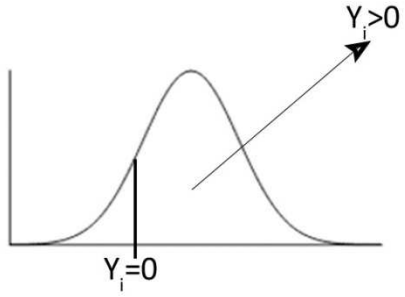
   $= 0$ if $\beta' X_i + u_i \leq 0$

   This can be written in a **latent variable framework**

$Y_i^* = \beta' X_i + u_i$, $ui \sim N(0, \sigma^2 u)$

$Y_i = Yi^*$ if $Y_i^* > 0$

$= 0$ if $Y_i^* \leq 0$



The two types of limited dependent variable models are :

Censoring

Occurs when the values of the dependent variable are restricted to a range of values ie; we observe both $Y_i = 0$ and $Y_i > 0$.

However, there is information (the independent variables) about the whole sample.A model commonly used to deal with censored data is the **tobit model**.

- Censored regression models are used for data where only the value for the dependent variable (hours of work in the example above) is unknown while the values of the independent variables (age, education, family status) are still available.
- When data is censored the distribution that applies to the sample data is a mixture of discrete and continuous distribution. The total probability is 1 as required, but instead of scaling the 2nd part we simply assign the full probability in the censored region to the censoring point ,in this case 0.

Truncation

The effect of truncation occurs when the observed data in the sample are only drawn from a subset of a larger population. The sampling of the subset is based on the value of the dependent variable.ie; we observe only $Y_i > 0$.

Truncated regression models are used for data where whole observations are missing so that the values for the dependent and the independent variables are unknown.

EXAMPLES:

1. A study of the determinants of incomes of the poor. Only households with income below a certain poverty line are part of the sample.
2. Suppose we have a sample of AIEEE rejects-those who scored below the 30th percentile. We wish to estimate an IQ equation: AIEEE=f(education,age,socio economic characteristics etc). We will need to take into account the fact that dependent variable is truncated.

CENSORED MODEL ESTIMATION :

- Estimation of the equation $Y_i = \beta' X_i + u_i$ by OLS generates inconsistent estimates of β

Mathematical explanation :

➢ $$E(Y_i|Y_i \quad E(Y_i|Y_i>0) = \beta' x_i + \sigma \frac{\phi(\beta' x_i/\sigma)}{\Phi(\beta' x_i/\sigma)} \neq \beta' x_i$$

Inverse Mills Ratio

➢ $$E(Y_i|X_i) = \Phi\left(\frac{\beta' x_i}{\sigma}\right)\beta' x_i + \sigma\phi\left(\frac{\beta' x_i}{\sigma}\right)$$

$\therefore E(Y_i|X_i) \neq \beta' x_i$

and in particular, depends on X_i

=> OLS inconsistent

Intuitive explanation :

INSERT FIGURE HERE

The regression line of Yi*(The original unobserved latent variable)
Blue line-Regression line of Yi (The observed variable)
Basically for the observations of Yi*≤ 0, Yi takes the value 0.Hence, the entire regression line changes. It will be downward biased.

As we can also see from the figure, the resulting intercept and slope coefficients are bound to be different than if all the observations were taken into account

The **Inverse Mills Ratio** or the hazard rate:

1. It is the ratio of the probability density function to the cumulative distribution function of a distribution.
2. A common application of the inverse Mills ratio to take account of a possible selection bias. If a dependent variable is censored it causes a concentration of observations at zero values. This problem was first acknowledged by Tobin (1958), who showed that if this is not taken into consideration in the estimation procedure, an ordinary least squares estimation (OLS) will produce biased parameter estimates

1. **<u>Maximum likelihood(ML)</u>** :
   Censored regression models are usually estimated bythe Maximum Likelihood (ML) method. Assuming that the disturbance term follows a normal distribution with mean 0 and variance σ^2

	L_i
$\psi_1 = \{i \mid Y_i = 0\}$	$1 - \Phi\left(\beta' X_i / \sigma\right)$
$\psi_2 = \{i \mid Y_i > 0\}$	$\frac{1}{\sigma}\phi\left(\frac{Y_i - \beta' X_i}{\sigma}\right)$

   Therefore,

$$L = \prod_{i \in \psi_1}\left[1 - \Phi\left(\beta' X_i / \sigma\right)\right] \prod_{i \in \psi_2}\left[\frac{1}{\sigma}\phi\left(\frac{Y_i - \beta' X_i}{\sigma}\right)\right]$$

   <u>Observations</u>:

- Li is a mixture of probability and density
- Depends on β and σ
- Likelihood function is globally concave only if there is one side censoring.

2. **<u>Non linear estimation</u>** :

$$Y_i = E(Y_i \mid X_i) + \eta_i$$

$$Y_i = \Phi\left(\frac{\beta' X_i}{\sigma}\right)\beta' X_i + \sigma\phi\left(\frac{\beta' X_i}{\sigma}\right) + \eta_i$$

$$SS = \sum \left[Y_i - E(Y_i \mid X_i)\right]^2$$

$$\frac{\partial SS}{\partial \beta} = 0, \quad \frac{\partial SS}{\partial \sigma} = 0$$

   <u>Difference between maximum likelihood and non linear estimation:</u>

- In ML we assume $u_i \sim N(0, \sigma^2)$
- In NLE we assume only independence of error, no assumption on distribution of errors.

HECKMAN'S 2 STEP PROCEDURE

A popular alternative to maximum likelihood estimation of the tobit model is Heckman's two-step, or correction, method.

Step 1: Use the probit estimate to compute estimate of inverse mills ratio.
Step 2: For positive observations of Y, run a regression of Y_i on X_{1i} and X_{2i} and inverse mills ratio estimate.

$$Y_i = E(Y_i \mid X_i) + \eta_i$$
$$= \Phi\left(\left(\frac{\beta}{\sigma}\right)' X_i\right) X_i \beta' + \sigma \phi\left(\left(\frac{\beta}{\sigma}\right)' X_i\right) + \eta_i$$

$$X_{1i} = \phi\left(\left(\frac{\hat{\beta}}{\sigma}\right)' x_i\right) x_i$$

$$X_{2i} = \phi\left(\left(\frac{\hat{\beta}}{\sigma}\right)' x_i\right)$$

$$Y_i = \beta' X_{1i} + \sigma X_{2i} + (\eta_i + v_i)$$

We get consistent estimates but not efficient.

APPLICATION OF HECKMAN'S 2 STEP PROCEDURE :

For several decades criminologists have recognized the widespread threat of sample selection bias in criminological research. Sample selection issues arise when a researcher is limited to information on a non-random sub-sample of the population of interest. Specifically, when observations are selected in a process that is not independent of the outcome of interest, selection effects may lead to biased inferences regarding a variety of different criminological outcomes. In criminology, one common approach to this problem is Heckman's (1976) two-step estimator.

TRUNCATED MODEL ESTIMATION

1. Regression Yi on Xi produces inconsistent β because of omitted variable bias.

$$E(Y_i \mid X_i) = E(Y_i^* \mid Y_i^* > 0)$$
$$= \beta' X_i + \sigma \frac{\phi\left(\beta' x_i / \sigma\right)}{\Phi\left(\beta' x_i / \sigma\right)}$$
$$Y_i = E(Y_i \mid X_i) + \eta_i$$
$$\Rightarrow Y_i = \beta' X_i + \sigma \frac{\phi\left(\beta' x_i / \sigma\right)}{\Phi\left(\beta' x_i / \sigma\right)} + \eta_i$$

Thus, E(Yi)≠β'Xi

2. Heckman's 2 step not possible as we cannot turn this into a probit model and get $(\hat{\beta}/\boldsymbol{\sigma})$ and inverse mills ratio estimate
3. Non linear estimation is also difficult to do

Maximum Likelihood:

Step 1: Calculation of likelihood function

$$L_i = f(Y_i) = f(Y_i^* \mid Y_i^* > 0)$$

$$L_i = \frac{\frac{1}{\sigma}\phi\left(\frac{Y_i - \beta' X_i}{\sigma}\right)}{\Phi\left(\beta' X_i / \sigma\right)}$$

$$\text{now, } L = \prod_{i=1}^{n} L_i$$

$$\therefore L = \prod_{i=1}^{n} \frac{\frac{1}{\sigma}\phi\left(\frac{Y_i - \beta' X_i}{\sigma}\right)}{\Phi\left(\beta' X_i / \sigma\right)}$$

Step 2: MaximiseΣ(Log N-Log D)

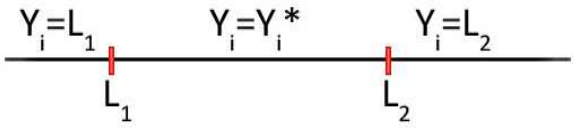
Step 3: Iterate to convergence

TWO LIMIT TOBIT MODEL :

The two limit tobit Model is a special case of censored model.

$Y_{i=}\ \beta' X_i + u_i$

Yi = L1if Yi*< L1
Yi = Yi*if L1 ≤ Yi ≤ L2
Yi=L2 if Yi* ≥ L2



$$E(Y_i \mid X_i) = L_1 P(Y_i = L_1) + L_2 P(Y_i = L_2) + E(Y_i \mid L_1 \le Y_i \le L_2) P(L_1 \le Y_i \le L_2)$$

Likelihood Function:

$$L_i(Y_i = L_1) = \Phi\left(\frac{L_1 - \beta' X_i}{\sigma}\right) \Rightarrow \Psi_1$$

$$L_i(Y_i = L_2) = 1 - \Phi\left(\frac{L_2 - \beta' X_i}{\sigma}\right) \Rightarrow \Psi_2$$

$$L_i(Y_i = Y_i^*) = \frac{1}{\sigma}\phi\left(\frac{Y_i - \beta' X_i}{\sigma}\right) \Rightarrow \Psi_3$$

$$L = \prod_{i \in \Psi_1} L_i \prod_{i \in \Psi_2} L_i \prod_{i \in \Psi_3} L_i$$

SAS MODEL SPECIFICATION

Censored Model

$$\text{Pension} = \alpha_1 + \alpha_2.\text{exper} + \alpha_3.\text{age} + \alpha_4.\text{tenure} + \alpha_5.\text{depends} + u_i$$

Dependent variable- **pension:** $ value of employee pension

Explanatory variables- **exper:** years of work experience

age: age in years

tenure : years with current employer

educ : years schooling

depends: number of dependents

The sample is censored with the lower boundary being at 0 and has 616 observations.

Truncated Model

$$\text{Achiv} = \beta_1 + \beta_2.\text{langscore} + \beta_3.\text{mathscore} + u_i$$

Dependent variable- **achiv**: This is the achievement score of the students in the GATE program.

Explanatory variables- **langscore** : score received in language test

mathscore :score received in the mathematics test.

Students are required to have a minimum achievement score of 40 to enter the special program. Thus, the sample is truncated at an achievement score of 40 as the lower boundary. The sample has 178 observations.

SAS COMMANDS USED

PROC QLIM

The QLIM (qualitative and limited dependent variable model) procedure analyzes limited dependent variable models in which dependent variables take discrete values or dependent variables are continuous and observed only in a limited range of values.

The QLIM procedure offers a class of models in which the dependent variable is censored or truncated from below or above or both.

When a continuous dependent variable is observed only within a certain range and values outside this range are not available, the QLIM procedure offers a class of models that adjust for truncation.

In case of censoring, the dependent variable is continuous only in a certain range and all values outside this range are reported as being on its boundary. For example, if it is not possible to observe negative values, the value of the dependent variable is reported as equal to zero. Because the data are censored, ordinary least squares (OLS) results are inconsistent, and it cannot be guaranteed that the predicted values from the model fall in the appropriate region.

The QLIM procedure uses maximum likelihood methods.

The standard Tobit model is estimated by specifying the endogenous variable to be truncated or censored. The limits of the dependent variable can be specified with the CENSORED or TRUNCATED option in the ENDOGENOUS or MODEL statement when the data are limited by specific values or variables.

The **lb=** option on the **endogenous** statement indicates the value at which the left truncation takes place (ie. the lower bound). There is also a**ub=** option to indicate the value of the right truncation (ie. the upper bound), which was not needed in this example.

Heckman's TwoStep Procedure

We have specified exactly two MODEL statements.One of the models is a binary probit model; therefore, we have specified the DISCRETE option in the MODEL. We could have also specified it in the ENDOGENOUS statement. We base the selection on the binary probit model for the second model; therefore, we have specified the SELECT option for this model.

Procsgplot procedure can be used to draw a histogram which is a distribution plot.Showbin specifies that the midpoints of the value bins are used to create the tickmarks for the horizontal axis. We have used this to show the truncation in the data.

We have used the**Proc means** command to get the summary statistics of the dependent variable achiv in the truncated model. This gives us the mean and standard error of the dependent variable and we can also see that it is truncated at 40 because the minimum value is 41.

COMMANDS

/*censored regression-maximum likelihood estimation*/

```
datasasuser.censoreddata;
procqlim data=sasuser.censoreddata;
model pension=exper age tenure educ depends;
endogenouspension~censored(lb=0);
run;
```

/*heckmans two step procedure-censored model*/

```
data sasuser.heck1;
setsasuser.censoreddata;
sel = (pension~=0);
run;
procqlim data=sasuser.heck1;
modelsel=exper age tenure educ depends/discrete;
model pension=exper age tenure educ depends/select (sel=1) ;
run;
```

/*summary statistics*/

```
proc means data = sasuser.truncateddata;
varachivlangscoremathscore;
run;
```

/* drawing a histogram for truncated data*/

```
procsgplot data = sasuser.truncateddata;
histogramachiv / scale = count showbins;
densityachiv;
run;
```

/*truncated regression –maximum likelihood estimation*/

```
datasasuser.truncateddata;
procqlim data=sasuser.truncateddata;
modelachiv= langscoremathscore;
endogenousachiv~truncated(lb=40);
run;
```

OUTPUT TABLES

ML for Censored Model

The QLIM Procedure

Summary Statistics of Continuous Responses

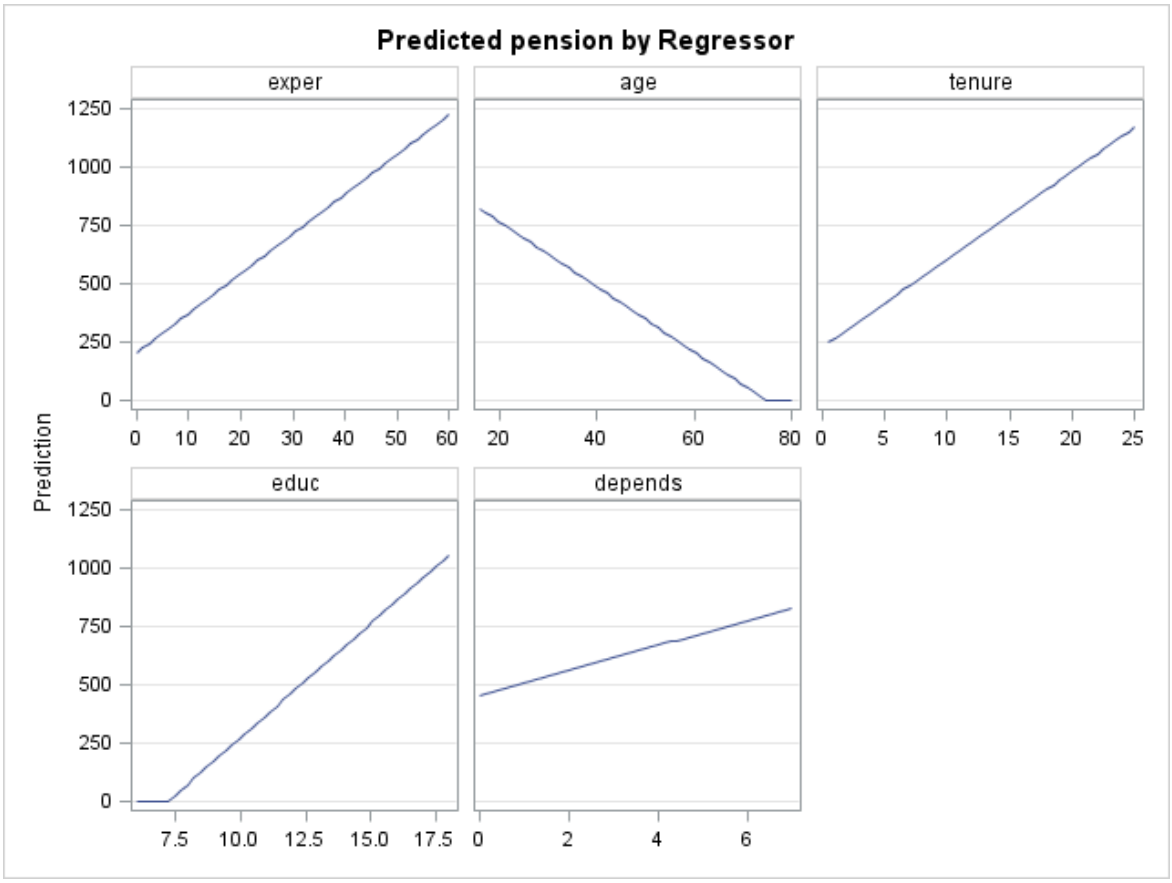
Variable	Mean	Standard Error	Type	Lower Bound	Upper Bound	N Obs Lower Bound	N Obs Upper Bound
pension	652.3368	619.119944	Censored	0		172	

Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	pension
Number of Observations	616
Log Likelihood	-3686
Maximum Absolute Gradient	4.79507E-6
Number of Iterations	28
Optimization Method	Quasi-Newton
AIC	7386
Schwarz Criterion	7417

Algorithm converged.

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr> |t|
Intercept	1	-850.322041	201.635278	-4.22	<.0001
exper	1	17.006833	5.625877	3.02	0.0025
age	1	-14.009404	5.458372	-2.57	0.0103
tenure	1	37.670954	4.662240	8.08	<.0001
educ	1	97.871188	11.129362	8.79	<.0001
depends	1	53.305029	20.899637	2.55	0.0108
_Sigma	1	697.625593	24.856554	28.07	<.0001


Predicted pension by Regressor
exper
age
tenure
educ
depends
Prediction
1250
1000
750
500
250
0
0 10 20 30 40 50 60
20 40 60 80
0 5 10 15 20 25
7.5 10.0 12.5 15.0 17.5
0 2 4 6

Heckmans Two Step Procedure for Censored Model

The QLIM Procedure

Summary Statistics of Continuous Responses

Variable	N	Mean	Standard Error	Type	Lower Bound	Upper Bound	N Obs Lower Bound	N Obs Upper Bound
pension	444	905.0439	550.369625	Regular				

Discrete Response Profile of sel

Index	Value	Total Frequency
1	0	172
2	1	444

Model Fit Summary

Number of Endogenous Variables	2
Endogenous Variable	sel pension
Number of Observations	616
Log Likelihood	-3709
Maximum Absolute Gradient	0.0001782
Number of Iterations	30
Optimization Method	Quasi-Newton
AIC	7447
Schwarz Criterion	7509

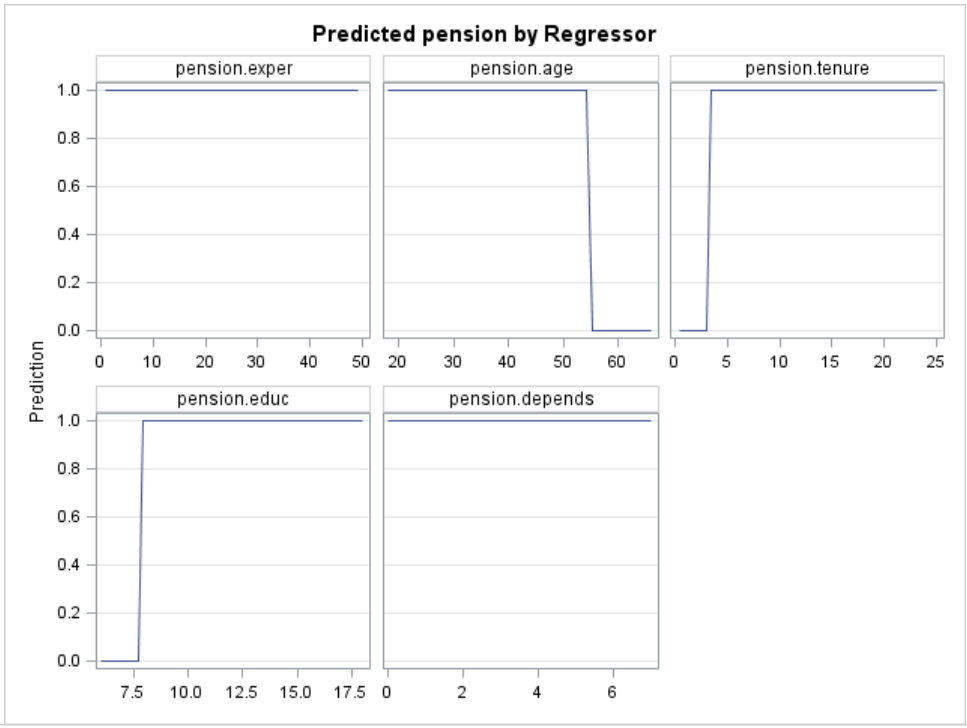
Algorithm converged.

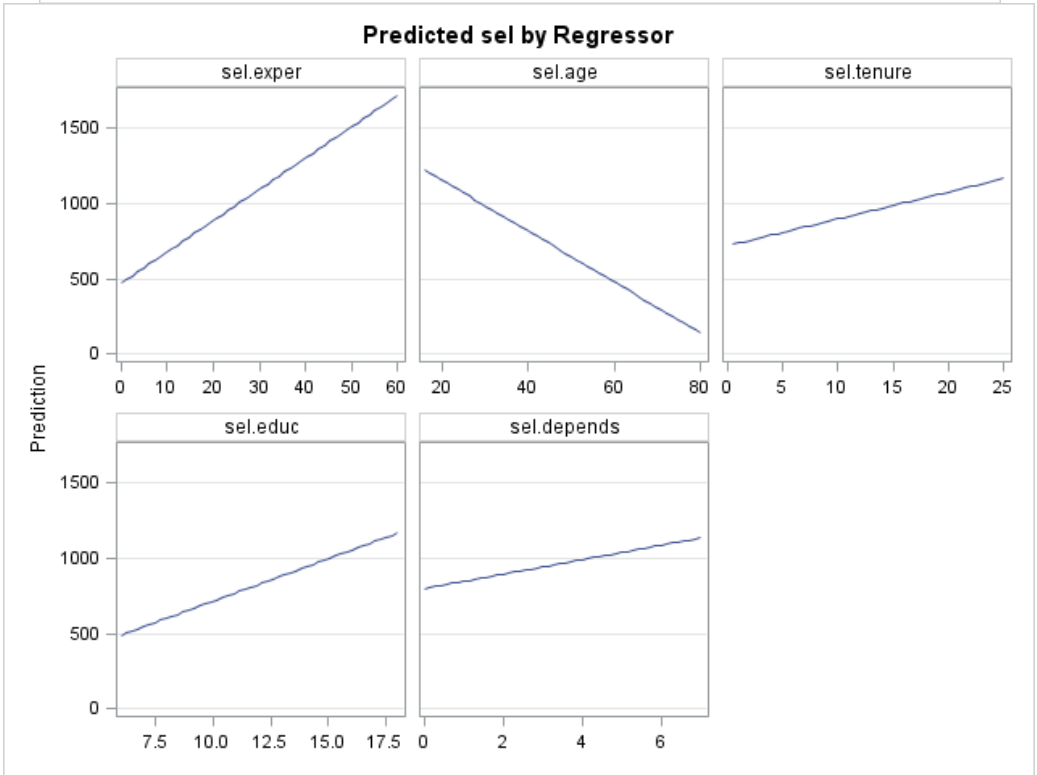
Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr> |t|
pension.Intercept	1	205.671877	.	.	.
pension.exper	1	20.817162	4.899971	4.25	<.0001
pension.age	1	-17.023148	4.767911	-3.57	0.0004
pension.tenure	1	17.960797	.	.	.
pension.educ	1	56.586989	.	.	.
pension.depends	1	48.240345	16.549817	2.91	0.0036
_Sigma.pension	1	495.407269	16.624905	29.80	<.0001
sel.Intercept	1	-1.366816	0.384213	-3.56	0.0004
sel.exper	1	0.007565	0	.	.

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr> |t|
sel.age	1	-0.006920	0	.	.
sel.tenure	1	0.061465	0	.	.
sel.educ	1	0.131079	0	.	.
sel.depends	1	0.036328	0.040747	0.89	0.3726
_Rho	1	-0.000058659	.	.	.


Predicted pension by Regressor
pension.exper
pension.age
pension.tenure
pension.educ
pension.depends
Prediction
1.0
0.8
0.6
0.4
0.2
0.0
0 10 20 30 40 50
20 30 40 50 60
0 5 10 15 20 25
7.5 10.0 12.5 15.0 17.5
0 2 4 6

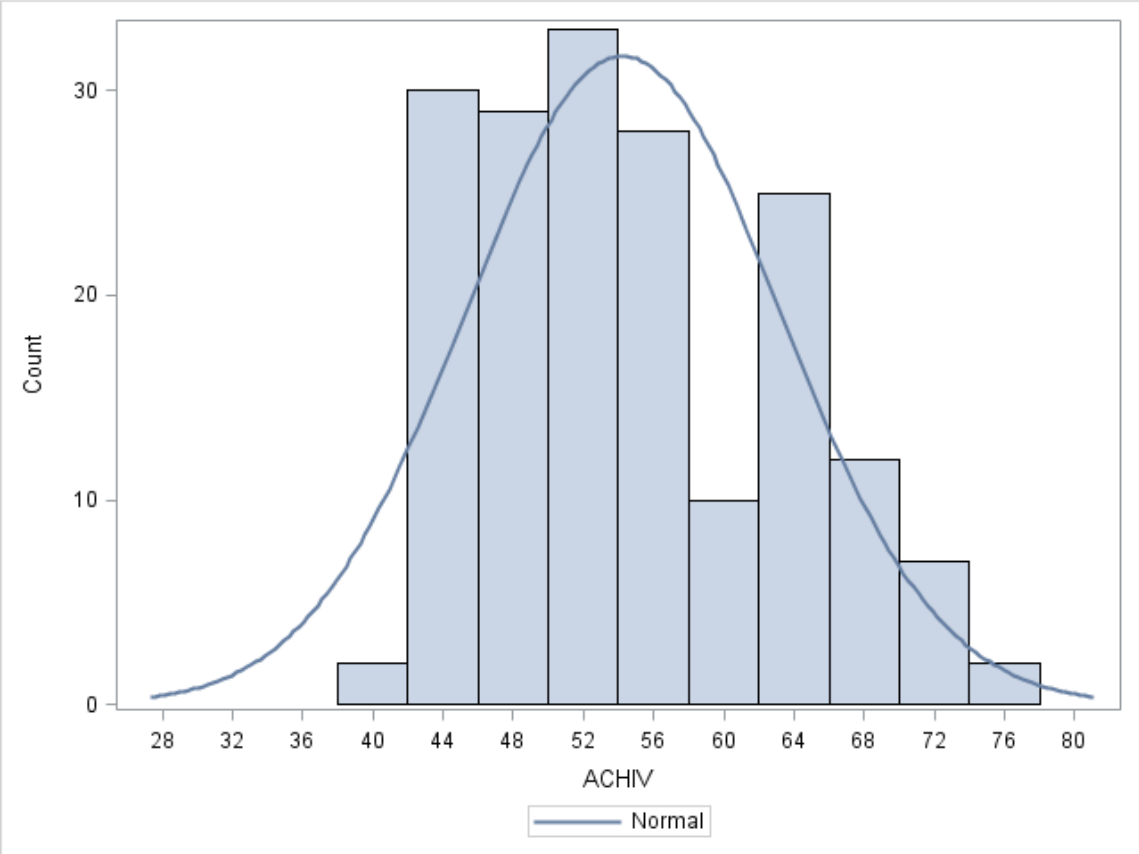

Predicted sel by Regressor
sel.exper
sel.age
sel.tenure
sel.educ
sel.depends
Prediction
1500
1000
500
0
0 10 20 30 40 50 60
20 40 60 80
0 5 10 15 20 25
7.5 10.0 12.5 15.0 17.5
0 2 4 6

Summary Statistics for Truncated Model

The MEANS Procedure

Variable	Label	N	Mean	StdDev	Minimum	Maximum
ACHIV	ACHIV	178	54.2359551	8.9632299	41.0000000	76.0000000
LANGSCORE	LANGSCORE	178	5.4011236	0.8944896	3.0999999	6.6999998
MATHSCORE	MATHSCORE	178	5.3028090	0.9483515	3.0999999	7.4000001

Histogram for Truncated Model:



Maximum Likelihood Estimation for Truncated Model

The QLIM Procedure

Summary Statistics of Continuous Responses

Variable	Mean	Standard Error	Type	Lower Bound	Upper Bound	N Obs Lower Bound	N Obs Upper Bound
ACHIV	54.23596	8.963230	Truncated	40			

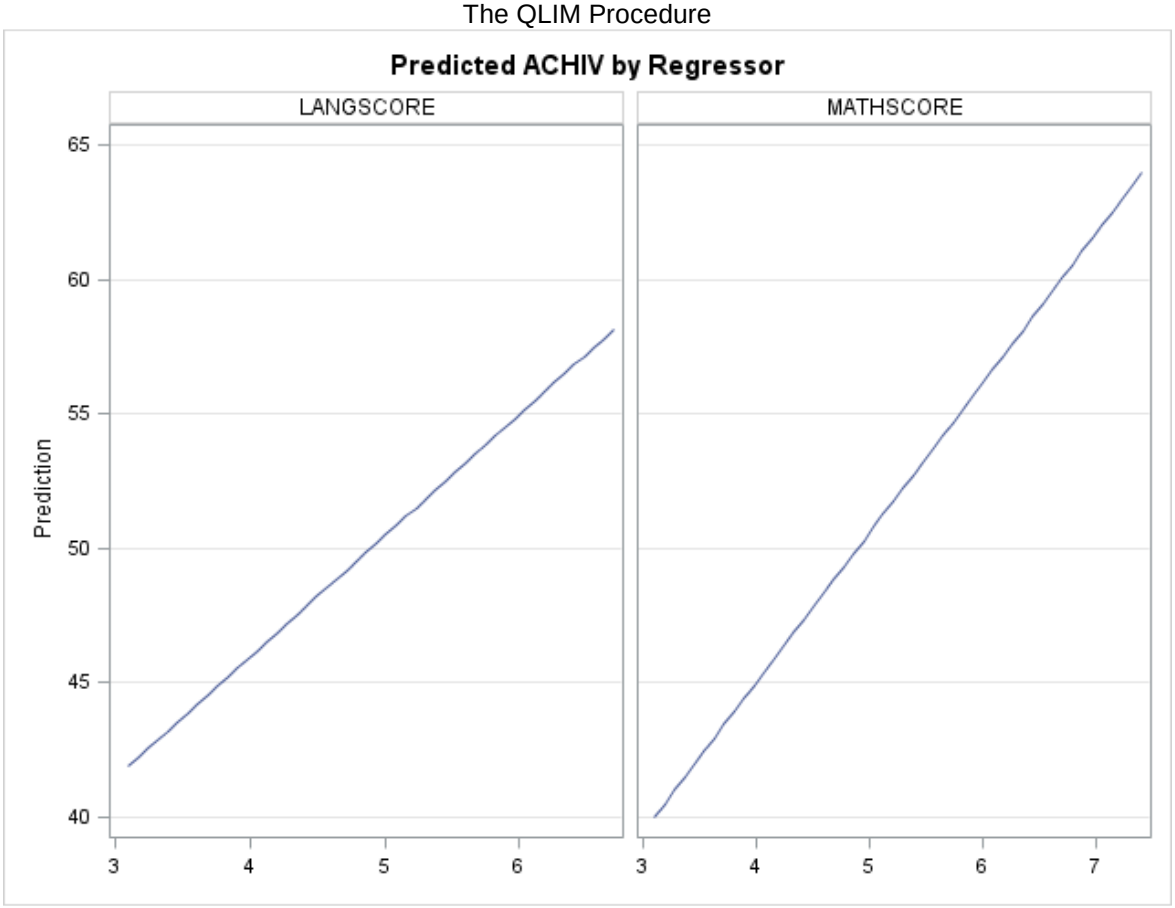
Model Fit Summary	
Number of Endogenous Variables	1
Endogenous Variable	ACHIV
Number of Observations	178
Log Likelihood	-575.70323
Maximum Absolute Gradient	1.50252E-6
Number of Iterations	10
Optimization Method	Quasi-Newton
AIC	1159
Schwarz Criterion	1172

Algorithm converged.

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Approx Pr> |t|
Intercept	1	-1.614073	6.307643	-0.26	0.7980
LANGSCORE	1	4.512715	0.975717	4.63	<.0001
MATHSCORE	1	5.568134	0.913297	6.10	<.0001
_Sigma	1	7.814077	0.555477	14.07	<.0001

The QLIM Procedure



INTERPRETATION OF RESULTS

INTERPRETATION OF RESULTS FOR TRUNCATED MODEL

- The output begins with summary statistics of the continuous outcome variableachiv. The summary includes the mean of the dependent variable **achiv**, as well as the standard error of the dependent variable. We can see that **achiv** is truncated at the value of 40 since the minimum is 41.

- The Model Fit Summary table gives information about the model, including the log likelihood.

- In the table called Parameter Estimates, we have the truncated regression coefficients, the standard error of the coefficients, the t-values, and the p-value associated with each t-value.

- **The null hypothesis says that the variables are statistically insignificant.**

- The p value is defined as **the lowest significance level at which a null hypothesis is rejected**

- The probability of obtaining a t value of 4.63 or greater for the coefficient of the variable langscore is <.0001. Thus the variable **langscore** is statistically significant. A unit increase in language score leads to a 4.51 unit increase in predicted achievement.

- The probability of obtaining a t value of 6.10 or greater for the coefficient of the variable mathscore is <.0001. Thus the variable **mathscore** is statistically significant. A unit increase in mathscore leads to a 5.57 unit increase in predicted achievement. .

  Thus we can see that the coefficients of all the explanatory variables in our model have the signs expected a-priori from the theory and are statistically significant at 1% level of significance. We see that an increase in the language score and mathematics score of an individual leads to an increment in the achievement level.

INTERPRETATION OF RESULTS FOR CENSORED MODEL

Maximum Likelihood

- The Model Fit Summary table gives information about the model, including the log likelihood.

- In the table called Parameter Estimates, we have the censored regression coefficients, the standard error of the coefficients, the t-values, and the p-value associated with each t-value.

- .The intercept term has no economic interpretation.

- The probability of obtaining a t value of 3.02 or greater for the coefficient of the variable exper is =0.0025. Thus the variable **exper** is statistically significant at 5% level of significance.A unit increase in experience holding all other variables constant leads to a 17.01 unit increase in expectedpension.

- The p-value of the coefficient of the variable age is =0.0103. Thus the variable **age** is statistically significant at 5% level of significance.A unit increase in age holding all other variables constantleads to a 14.01 unit decrease in expected pension.

- The probability of obtaining a t value of 8.08 or greater for the coefficient of the variable tenure is <0.0001. Thus the variable **tenure** is statistically significant at 5% level of significance.A unit increase in tenure holding all other

variables constant leads to a 37.67 unit increase in expected pension.

- The probability of obtaining a t value of 8.79 or greater for the coefficient of the variable educ is <0.0001. Thus the variable **educ**is statistically significant at 5% level of significance.A unit increase in education holding all other variables constant leads to a 97.87 unit increase in expected pension.

- The probability of obtaining a t value of 2.55 or greater for the coefficient of the variable depends is =0.0108. Thus the variable **depends** is statistically significant at 5% level of significance.A unit increase in number of dependentsholding all other variables constant leads to a 53.30 unit increase in expected pension.

Thus we can see that the coefficients of all the explanatory variables in our model have the signs expected a-priori from the theory and are statistically significant at 5% level of significance. We see that the an increase in the experience of an individual, his/her level of educational attainment, years of tenure and the number of dependents, all lead to an increase in expected pension. It is the increase in education which leads to the maximum increase in expected pension and increase in experience which leads to least increase. An increase in age leads to a decrease in expected value of pension received.

Heckman Two Step Procedure

- A unit increase in **exper** holding all other variables constant leads to a 20.82 unit increase in expected pension.
- A unit increase in **age** holding all other variables constant leads to a 17.02 unit decrease in expected pension.
- A unit increase in **tenure** holding all other variables constant leads to a 17.96 unit increase in expected pension.
- A unit increase in **educ** holding all other variables constant leads to a 56.59 unit increase in expected pension.
- A unit increase in **depends** holding all other variables constant leads to a 48.24 unit increase in expected pension.

Thus we can see that the coefficients of all the explanatory variables in our model have the signs expected a-priori from the theory.

Applications:

Dividend Payment Model

Dividend is the return that a shareholder gets from a company, out of its profits, on his shareholding. Equity investors receive returns in the form of dividends.

$Yi^*=B'Xi+ui$

Yi=0 if $Yi^*<Lo$
=Lo if $Lo<=Yi^*<L1$
=L1 if $L1<=Yi^*$

Likelihood Function :

$$L_i(Y_i=0)= \Phi\left(\frac{L_0-\beta'X_i}{\sigma}\right) => \Psi_1$$

$$L_i(Y_i=L_0)=\Phi\left(\frac{L_1-\beta'X_i}{\sigma}\right)-\Phi\left(\frac{L_0-\beta'X_i}{\sigma}\right) => \Psi_2$$

$$L_i(Y_i=L_1)= 1-\Phi\left(\frac{L_1-\beta'X_i}{\sigma}\right) => \Psi_3$$

$$L=\prod_{i\epsilon\Psi_1} L_i \prod_{i\epsilon\Psi_2} L_i \prod_{i\epsilon\Psi_3} L_i$$

Expected Value of Yi:

$$E(Y_i|X_i) = L_0\left[\Phi\left(\frac{L_1-\beta'X_i}{\sigma}\right)-\Phi\left(\frac{L_0-\beta'X_i}{\sigma}\right)\right] + L_1\left[1-\Phi\left(\frac{L_1-\beta'X_i}{\sigma}\right)\right]$$

Asset Holding Model of Rosset

Here we make use of the fact that because of transaction costs,investors donot change their portfolio in response to small gains.

Yi*=B'Xi+ui

$$Y_i = \begin{cases} Y_i^* - \alpha_1 & \text{if } Y_i^* \le \alpha_1 \\ 0 & \text{if } \alpha_1 < Y_i^* \le \alpha_2 \\ Y_i^* - \alpha_2 & \text{if } Y_i^* > \alpha_2 \end{cases}$$

$$\text{Given: } \alpha_1 < 0$$
$$\alpha_2 > 0$$

Likelihood Function

$$L_i(Y_i \epsilon \psi_1) = \frac{1}{\sigma} \phi\left(\frac{Y_i - \beta' X_i + \alpha_1}{\sigma}\right)$$

$$L_i(Y_i \epsilon \psi_2) = \Phi\left(\frac{\alpha_2 - \beta' X_i}{\sigma}\right) - \Phi\left(\frac{\alpha_1 - \beta' X_i}{\sigma}\right)$$

$$L_i(Y_i \epsilon \psi_3) = \frac{1}{\sigma} \phi\left(\frac{Y_i - \beta' X_i + \alpha_2}{\sigma}\right)$$

$$L = \prod_{i \epsilon \psi_1} L_i \prod_{i \epsilon \psi_2} L_i \prod_{i \epsilon \psi_3} L_i$$

Expected value of Yi

Similarly,the expected value of Yi can be calculated and it will turn out not equal to B'Xi.