PURDUE
UNIVERSITY

**School of Civil Engineering**

# CE615 – Statistical and Econometric Methods II

## Assignment #11
## Continuous Censored Data
## Tobit Model and Random Parameters Tobit Model

You are given vehicle accident data from 337 rural interstate road sections in the state of Indiana for a 5-year period (1995 to 1999). The use of accidents per vehicle-miles traveled has an intuitive appeal in highway safety – providing a standardized measure of the relative safety of roadway segments that is more easily interpreted than the number of accidents per some time period. Because accident rates on specific highway segments are assessed over some finite time period, there is the likelihood that many highway segments will have no accidents reported during the analysis period. Thus, modeling accident rates by standard OLS would result in biased and inconsistent parameter estimates. The solution to this is to consider accident rates as a censored dependent variable (censored at zero) and apply a tobit model. For the accident-rate considered, the data will be left-censored with a clustering at zero (zero accidents per 100-million vehicle miles traveled) because accidents may not be observed on all roadway segments during the period of observation. For model estimation, the accident rate (number of accidents per 100-million VMT) was calculated as:

$$Accident\ Rate_i = \frac{\sum_{Year=1}^{5} Accidents_{Year,i}}{\left[\sum_{Year=1}^{5} AADT_{Year,i} \times L_i \times 365\right] \Big/ 100{,}000{,}000}$$

where $Accident\ Rate_i$ is the number of accidents per 100-million VMT on roadway segment $i$, $Year$ denotes the year (1995 to 1999), $Accidents_{Year,i}$ is the number of accidents, $AADT_{Year,i}$ the average annual daily traffic, $L_i$ the length of roadway segment $i$. The model's overall fit can be measured with Maddala's $R^2$:

$$R^2 = 1 - \exp(-LRT / N)$$

where $N$ is the number of observations, and $LRT=2[LL(\beta)–LL(0)]$ is the likelihood ratio statistic.

Your task is to estimate a model of accident rates using tobit regression. Your solution to this problem should include:

1. The results of your best model specification.

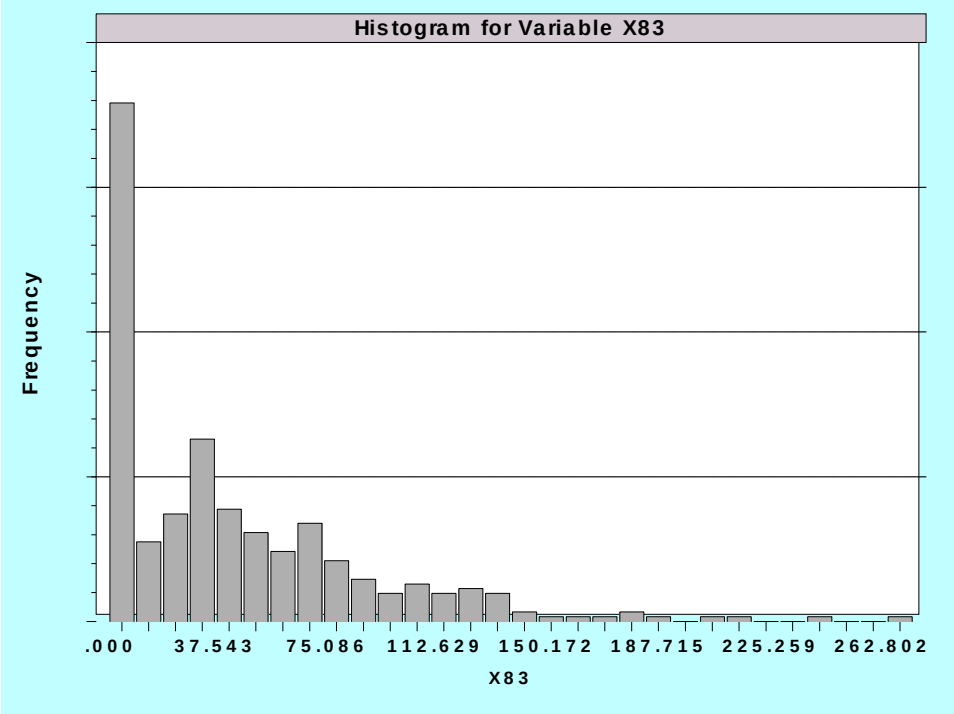
2. A discussion of the logical process that led you to the selection of your final specification. (e.g. Discuss the theory behind the inclusion of your selected variables). Include t-statistics and justify the sign of your variables.

Variables available for your specification are (file *tob337.txt*) :

| Variable | Explanation |
|---|---|
| x1 | ID |
| x2 | Number of observations for each interstate |
| x3 | Number of single vehicle accidents per 100-million VMT |
| x4 | Interstate (64: I-64, 65: I-65, 70: I-70, 74: I-74, and 164: I-164) |
| x5 | Average Friction in the road section over the 5-year period (measured at 40MPH) |
| x6 | Minimum Friction reading in the road section over the 5-year period |
| x7 | Maximum Friction reading in the road section over the 5-year period |
| x8 | Standard Deviation of the Friction readings in the road section over the 5-year period |
| x9 | Age of the pavement in 1999 |
| x10 | Average IRI in the road section over the 5-year period |
| x11 | Minimum IRI reading in the road section over the 5-year period |
| x12 | Maximum IRI reading in the road section over the 5-year period |
| x13 | Standard Deviation of the IRI readings in the road section over the 5-year period |
| x14 | Average Rutting (in inches) in the road section over the 5-year period |
| x15 | Minimum Rut (in inches) reading in the road section over the 5-year period |
| x16 | Maximum Rut (in inches) reading in the road section over the 5-year period |
| x17 | Standard Deviation of the Rut (in inches) readings in the road section over the 5-year period |
| x18 | Average PCR in the road section over the 5-year period |
| x19 | Minimum PCR in the road section over the 5-year period |
| x20 | Maximum PCR in the road section over the 5-year period |
| x21 | Standard Deviation of the PCR in the road section over the 5-year period |
| x22 | Average PQI in the road section over the 5-year period |
| x23 | Summation of AADT over the 5 years |
| x24 | Section length (in miles) |
| x25 | Total number of ramps in the opposite direction |
| x26 | Total number of ramps in the viewing direction |
| x27 | Number of lanes |
| x28 | Pavement surface type (1: asphalt, 0: concrete) |
| x29 | Median configuration (1: depressed, 2: depressed with bumps, 3: berms, 4: flush, 5: sloped, and 6: rock wall) |
| x30 | Median surface (0: concrete, 1: asphalt, 2: grass, 3: paved, 4: grass with trees, 5: grass with bushes, 6: trees, and 7: rock) |
| x31 | Median width (in feet) |
| x32 | Presence of median barrier (1: present, 0: absent) |
| x33 | Median barrier type (1: wbeam, 2: concrete, 3: brifen, 4: cable, 5: box-beam, 6: rock wall) |
| x34 | Median barrier location (0: left, 1: middle left, 2: middle, 3: middle right, 4: right) |
| x35 | Presence of interior shoulder (1: present, 0 absent) |
| x36 | Interior shoulder width (in feet) |
| x37 | Interior shoulder surface (0: concrete, 1: asphalt) |
| x38 | Interior rumble strips (1: present, 0: absent) |
| x39 | Outside shoulder width (in feet) |
| x40 | Outside shoulder surface (0: concrete, 1: asphalt) |
| x41 | Outside rumble strips (1: present, 0: absent) |
| x42 | Outside barrier type (1: wbeam, 2: concrete, 3: brifen, 4: cable, 5: box-beam, 6: rock wall) |

| | |
|---|---|
| x43 | Outside barrier location (1: less than 15 feet, 2: greater than 15 feet) |
| x44 | Average AADT over the 5 years |
| x45 | Average AADT of trucks over the 5 years |
| x46 | Percentage of single unit trucks (average daily) |
| x47 | Percentage of combination trucks (average daily) |
| x48 | Speed limit of the road section |
| x49 | State speed limit |
| x50 | Number of bridges in the road section |
| x51 | Horizontal curve 1 type (1: inside, 2: outside) |
| x52 | Length of horizontal curve 1 |
| x53 | Radius of horizontal curve 1 |
| x54 | Horizontal curve 2 type (1: inside, 2: outside) |
| x55 | Length of horizontal curve 2 |
| x56 | Radius of horizontal curve 2 |
| x57 | Horizontal curve 3 type (1: inside, 2: outside) |
| x58 | Length of horizontal curve 3 |
| x59 | Radius of horizontal curve 3 |
| x60 | Horizontal curve 4 type (1: inside, 2: outside) |
| x61 | Length of horizontal curve 4 |
| x62 | Radius of horizontal curve 4 |
| x63 | Horizontal curve 5 type (1: inside, 2: outside) |
| x64 | Length of horizontal curve 5 |
| x65 | Radius of horizontal curve 5 |
| x66 | Average radius per horizontal curve in the road section |
| x67 | Number of horizontal curves in the road section |
| x68 | Length of vertical curve 1 |
| x69 | Vertical curve 1 type (1: crest, 2: sag) |
| x70 | K parameter for vertical curve 1 |
| x71 | Length of vertical curve 2 |
| x72 | Vertical curve 2 type (1: crest, 2: sag) |
| x73 | K parameter for vertical curve 2 |
| x74 | Length of vertical curve 3 |
| x75 | Vertical curve 3 type (1: crest, 2: sag) |
| x76 | K parameter for vertical curve 3 |
| x77 | Number of vertical curves in the road section |
| x78 | Pavement surface change in the road section (1: change, 0: no change) |
| x79 | Changes in vertical profile (1: change, 0: no change) |
| x80 | Number of bridges per mile |
| x81 | Number of horizontal curves per mile |
| x82 | Number of vertical curves per mile |
| x83 | Number of Accidents per 100-million VMT |

```
--> RESET
--> read;nvar=83;nobs=337;file=D:\new_laptop\CE697N-disk\TOB337.TXT$
--> skip
--> histogram;rhs=x83$
```




```
--> create;if(x38=1&x41=1)rumblstr=1$
--> create;if(x26>0)ramp=1$
--> create;if(x38=1&x41=1)rumblstr=1$
--> create;if(x26>0)ramp=1$
--> tobit;lhs=x83;rhs=one,x32,x6,x12,x16,x18,ramp,x39,x47,x50
    ;marginal effects$


 ************************************************************************
 * NOTE: Deleted      8 observations with missing data. N is now    329 *
 ************************************************************************

Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| Limited Dependent Variable Model - CENSORED |
| Maximum Likelihood Estimates                |
| Model estimated: Nov 10, 2010 at 09:44:50AM.|
| Dependent variable                    X83   |
| Weighting variable                   None   |
| Number of observations                329   |
| Iterations completed                    6   |
| Log likelihood function       -1298.850     |
| Threshold values for the model:             |
| Lower=     .0000     Upper=+infinity        |
| LM test [df] for tobit=     21.168[ 10]     |
| ANOVA  based fit measure =     .187353      |
| DECOMP based fit measure =     .204418      |
+---------------------------------------------+
```

```
+---------+-------------+----------------+--------+---------+----------+
|Variable | Coefficient | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+-------------+----------------+--------+---------+----------+
         Primary Index Equation for Model
 Constant      67.2956685     88.5463578      .760   .4473
 X32         -123.528829      15.8944130    -7.772   .0000     .15805471
 X6           -1.56553606       .53937372   -2.903   .0037    30.5592705
 X12            .16192574       .12362880    1.310   .1903    101.386018
 X16           9.33598386     29.7236801      .314   .7535     .21806837
 X18            .79781145       .90663051      .880   .3789    94.4201460
 RAMP          30.8421934      8.42597294     3.660   .0003     .17021277
 X39          -5.75386157      2.25809145    -2.548   .0108    11.2983066
 X47          -35.7374539     27.6789876    -1.291   .1967     .23107196
 X50          -6.18365987      4.41955008    -1.399   .1618     .34042553
         Disturbance standard deviation
 Sigma         54.0516387      2.62217276    20.613   .0000
+-------------------------------------------+
| Partial derivatives of expected val. with |
| respect to the vector of characteristics. |
| They are computed at the means of the Xs. |
| Observations used for means are All Obs.  |
| Conditional Mean at Sample Point  35.3801 |
| Scale Factor for Marginal Effects   .6687 |
+-------------------------------------------+
+---------+-------------+----------------+--------+---------+----------+
|Variable | Coefficient | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+-------------+----------------+--------+---------+----------+
 Constant      45.0032188     59.3544926      .758   .4483
 X32          -82.6085103      9.49254349   -8.702   .0000     .15805471
 X6            -1.04693457       .36040609   -2.905   .0037    30.5592705
 X12            .10828601       .08259238    1.311   .1898    101.386018
 X16           6.24333384     19.8774926      .314   .7535     .21806837
 X18            .53352740       .60535146      .881   .3781    94.4201460
 RAMP          20.6253687      5.66169987     3.643   .0003     .17021277
 X39          -3.84782999      1.50841794    -2.551   .0107    11.2983066
 X47          -23.8990190     18.4803827    -1.293   .1959     .23107196
 X50          -4.13525275      2.95394518    -1.400   .1615     .34042553
 Sigma            .000000   ......(Fixed Parameter).......
--> tobit;lhs=x83;rhs=one,x32,x6,x12,x16,x18,ramp,x39,x47,x50
    ;rpm;pts=200;halton
    ;fcn=x32(n)$
+-----------------------------------------------------+
| OLS Starting values for random parameters model     |
| Ordinary    least squares regression                |
| Model was estimated Nov 10, 2010 at 09:44:50AM      |
| LHS=X83      Mean                 =   42.13068      |
|              Standard deviation   =   47.68223      |
| WTS=none     Number of observs.   =        329      |
| Model size   Parameters           =         10      |
|              Degrees of freedom   =        319      |
| Residuals    Sum of squares       =   598786.6      |
|              Standard error of e  =   43.32521      |
| Fit          R-squared            =   .1970563      |
|              Adjusted R-squared   =   .1744027      |
| Model test   F[  9,   319] (prob) =   8.70 (.0000)  |
| Diagnostic   Log likelihood       =  -1701.667      |
|              Restricted(b=0)      =  -1737.770      |
|              Chi-sq [  9]  (prob) =  72.21 (.0000)  |
| Info criter. LogAmemiya Prd. Crt. =   7.567412      |
|              Akaike Info. Criter. =   7.567393      |
+-----------------------------------------------------+
```

| Variable | Coefficient | Standard Error | b/St.Er. | P[\|Z\|>z] | Mean of X |
|---|---|---|---|---|---|
| Constant | 100.811773 | 66.0926175 | 1.525 | .1272 | |
| X6 | -.98811938 | .38712311 | -2.552 | .0107 | 30.5592705 |
| X12 | .07364387 | .09179010 | .802 | .4224 | 101.386018 |
| X16 | 3.60567884 | 22.3421459 | .161 | .8718 | .21806837 |
| X18 | .10795692 | .66272420 | .163 | .8706 | 94.4201460 |
| RAMP | 23.3676259 | 6.53211971 | 3.577 | .0003 | .17021277 |
| X39 | -3.47679430 | 1.47889203 | -2.351 | .0187 | 11.2983066 |
| X47 | -13.8628122 | 20.3557843 | -.681 | .4959 | .23107196 |
| X50 | -3.94485022 | 2.93460531 | -1.344 | .1789 | .34042553 |
| X32 | -44.7964161 | 7.36136187 | -6.085 | .0000 | .15805471 |

Normal exit from iterations. Exit status=0.

| Random Coefficients Tobit Model | |
|---|---|
| Maximum Likelihood Estimates | |
| Model estimated: Nov 10, 2010 at 09:45:32AM. | |
| Dependent variable | X83 |
| Weighting variable | None |
| Number of observations | 337 |
| Iterations completed | 23 |
| Log likelihood function | -1293.256 |
| Sample is 1 pds and | 337 individuals. |
| Missing data: Skipped | 8 individuals. |
| TOBIT (censored) regression model | |
| (Lower) censoring limit is | .00 |
| Simulation based on 200 Halton draws | |

| Variable | Coefficient | Standard Error | b/St.Er. | P[\|Z\|>z] | Mean of X |
|---|---|---|---|---|---|
| | Nonrandom parameters | | | | |
| Constant | 96.0801252 | 57.3498022 | 1.675 | .0939 | |
| X6 | -1.84762264 | .33746849 | -5.475 | .0000 | 30.5592705 |
| X12 | .20862415 | .08479648 | 2.460 | .0139 | 101.386018 |
| X16 | 5.66105038 | 18.8205540 | .301 | .7636 | .21806837 |
| X18 | .71420105 | .59636375 | 1.198 | .2311 | 94.4201460 |
| RAMP | 32.2730332 | 5.32003298 | 6.066 | .0000 | .17021277 |
| X39 | -7.32638069 | 1.57636793 | -4.648 | .0000 | 11.2983066 |
| X47 | -29.9633733 | 15.8914149 | -1.886 | .0594 | .23107196 |
| X50 | -5.97285462 | 3.19924852 | -1.867 | .0619 | .34042553 |
| | Means for random parameters | | | | |
| X32 | -301.937993 | 50.1580519 | -6.020 | .0000 | .15805471 |
| | Scale parameters for dists. of random parameters | | | | |
| X32 | 150.616895 | 29.1109333 | 5.174 | .0000 | |
| | Variance parameter given is sigma | | | | |
| Std.Dev. | 52.4137073 | 1.03155116 | 50.811 | .0000 | |

Implied standard deviations of random parameters

Matrix S.D_Beta has 1 rows and 1 columns.

| | 1 |
|---|---|
| 1 | 150.61689 |