Statistical and Econometric Methods for Transportation Data Analysis

Chapter 3 – Linear Regression

Example 3.9
Continuous Censored Data – RP and Bivariate 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.

Your task is to estimate two models of accident rates:

1. Using variable x83 (Number of Accidents per 100-million VMT), estimate a random parameters tobit regression.

2. Estimate a bivariate tobit regression using a) variable x3 (Number of single vehicle accidents per 100-million VMT) and a new varaible you create "mcrash" (x83-x3, which is the number of multivehicle crashes). Your bivariate tobit regression will simultaneously capture the relationship between single and multivehicle crashes.

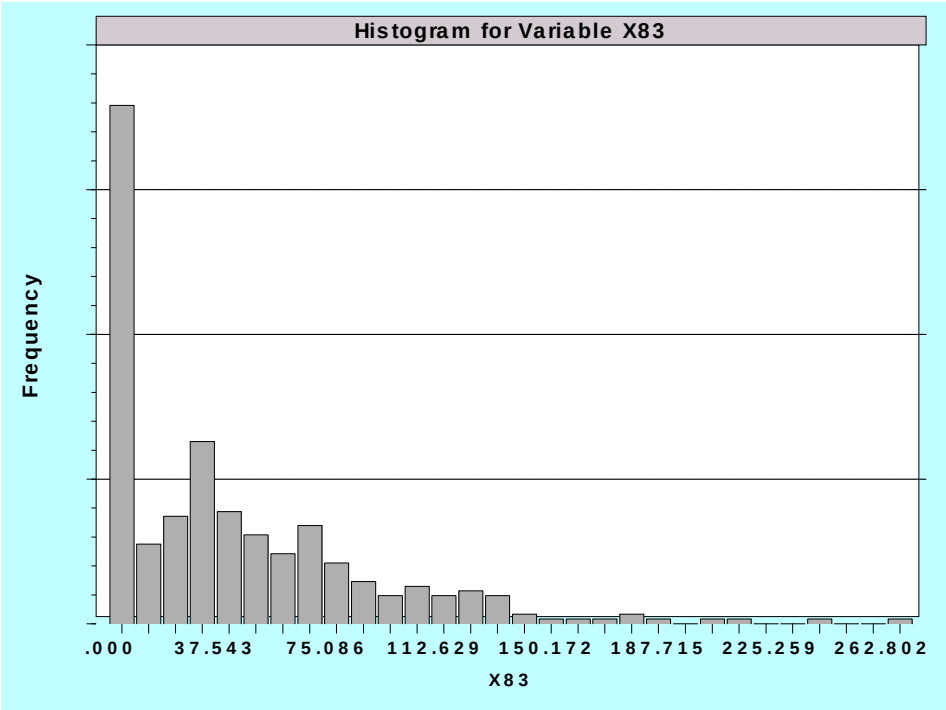
Provide the results of your best model specification and a discussion of the logical process that led you to the selection of your final specification. (discuss the theory behind the inclusion of your selected variables). Include t-statistics and justify the signs of your variables.

Variables available for your specification are (file Ex3-9.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

```
--> read;nvar=83;nobs=337;file=D:\old_drive_d\Book\Book2e-Data\Ex3-9.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,ramp,x39
    ;rpm
    ;fcn=x39(n),x32(n);pts=200;halton$

+-----------------------------------------------------+
| OLS Starting values for random parameters model     |
| Ordinary    least squares regression                |
| Model was estimated Nov 13, 2012 at 03:28:06PM      |
| LHS=X83      Mean                  =   41.29412     |
|              Standard deviation     =   47.49991     |
| WTS=none     Number of observs.     =        337     |
| Model size   Parameters             =          5     |
|              Degrees of freedom     =        332     |
| Residuals    Sum of squares         =   615737.6     |
|              Standard error of e    =   43.06543     |
| Fit          R-squared              =   .1877854     |
|              Adjusted R-squared     =   .1779996     |
| Model test   F[  4,   332] (prob)   =  19.19 (.0000) |
| Diagnostic   Log likelihood         =  -1743.700     |
|              Restricted(b=0)        =  -1778.747     |
|              Chi-sq [  4]  (prob)   =  70.09 (.0000) |
| Info criter. LogAmemiya Prd. Crt.   =   7.540169     |
|              Akaike Info. Criter.   =   7.540167     |
+-----------------------------------------------------+
```

```
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    103.574369      18.6690804     5.548   .0000
 X6      |    -.89122594       .35435603    -2.515   .0119   30.4816024
 RAMP    |    24.9168308      6.35231757     3.922   .0001    .16617211
 X39     |   -2.85638391      1.35478699    -2.108   .0350   11.2811361
 X32     |   -43.8816293      6.45321793    -6.800   .0000    .16023739

Normal exit from iterations. Exit status=0.

+----------------------------------------------+
| Random Coefficients  Tobit    Model          |
| Maximum Likelihood Estimates                 |
| Model estimated: Nov 13, 2012 at 03:28:23PM.|
| Dependent variable                  X83      |
| Weighting variable                 None      |
| Number of observations              337      |
| Iterations completed                 21      |
| Log likelihood function       -1308.802      |
| Number of parameters                  8      |
| Info. Criterion: AIC =          7.81485      |
|   Finite Sample: AIC =          7.81615      |
| Info. Criterion: BIC =          7.90553      |
| Info. Criterion:HQIC =          7.85099      |
| Sample is  1 pds and     337 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|    149.981213      34.2654190     4.377   .0000
 X6      |   -1.52834793       .55621860    -2.748   .0060   30.4816024
 RAMP    |    34.7932322      9.85654887     3.530   .0004    .16617211
---------+Means for random parameters
 X39     |   -5.97765651      2.58736695    -2.310   .0209   11.2811361
 X32     |   -219.181361      53.6372682    -4.086   .0000    .16023739
---------+Scale parameters for dists. of random parameters
 X39     |     .29915460       .28769902     1.040   .2984
 X32     |    88.3894229      32.7427440     2.700   .0069
---------+Variance parameter given is sigma
 Std.Dev.|    53.4647878      1.55576133    34.366   .0000


Implied standard deviations of random parameters

Matrix S.D_Beta has  2 rows and  1 columns.
               1
        +--------------
       1|     .29915
       2|   88.38942
```

```
--> btobit;lhs=x3,mcrash
    ;rh1=one,x32,x6,ramp,x39
    ;rh2=one,x32,x6,ramp,x39$
Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| Maximum likelihood ests.: Bivariate Tobit   |
| First  equation LHS variable: Y1 = X3       |
| Second equation LHS variable: Y2 = MCRASH   |
| Dependent variable               X3:MCR     |
| Weighting variable                 None     |
| Number of observations              337     |
| Iterations completed                 24     |
| Log likelihood function       -2091.623     |
| Number of parameters                 13     |
| Info. Criterion: AIC =         12.49034     |
|   Finite Sample: AIC =         12.49369     |
| Info. Criterion: BIC =         12.63770     |
| Info. Criterion:HQIC =         12.54908     |
| Nonlimit observations: X3      --   224.0   |
| Nonlimit observations: MCRASH  --   178.0   |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
---------+Equation (RHS) for X3
 Constant|    78.5266390      30.3545277     2.587   .0097
 X32     |   -101.418512      34.6735680    -2.925   .0034    .16023739
 X6      |   -1.04942311       .47991761    -2.187   .0288   30.4816024
 RAMP    |    20.2514581      10.8777368     1.862   .0626    .16617211
 X39     |   -1.93219884      2.31475974     -.835   .4039   11.2811361
---------+Equation (RHS) for MCRASH
 Constant|    67.8141967      16.1358624     4.203   .0000
 X32     |   -46.9679753      17.3580975    -2.706   .0068    .16023739
 X6      |    -.62591638       .26708930    -2.343   .0191   30.4816024
 RAMP    |    20.8416289      4.70799445     4.427   .0000    .16617211
 X39     |   -4.11068854      1.16166845    -3.539   .0004   11.2811361
---------+          Disturbance Variances and Correlation
 Sigma(1)|    45.1132174      1.38033237    32.683   .0000
 Sigma(2)|    24.4321580      1.30624632    18.704   .0000
 RHO(1,2)|     .54907087       .04608831    11.913   .0000
```