Statistical and Econometric Methods for Transportation Data Analysis

Chapter 11 – Count Data Models

Example 11.6 (should be Example 11.8, typo in book)
Count Data – Zero-Inflated Models and RPZIP

You are given 204 observations from a travel survey conducted in the Seattle metropolitan area. The purpose of the survey was to study the number of times (per week) commuters' delayed their departure time on their work-to-home trip to avoid traffic congestion. The data are non-negative integers and are thus well suited to the Poisson regression approach. You are estimating a parameter vector β such that:

$$\lambda = EXP(\beta X)$$

where λ is the Poisson parameter that in this case is the expected number of departure changes per week.

Recall that in Example 11.3 we used variable X2 (see next page) to eliminate people in the sample that said they never delayed (leaving just 96 observations). In this assignment, use the full 204 observations and explore various zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB) and random parameters (RPZIP) possibilities.

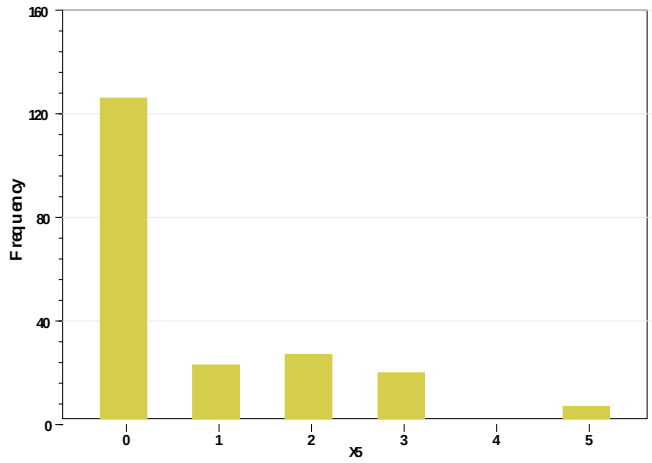
Include a discussion of your findings with the various model forms (including random parameters) and a select and discuss what you believe to be your best model.

Variables available for your specification are (file Ex11-6.txt)

Variable Number	Explanation
x1	Household number
x2	Do you ever delay work-to-home departure to avoid traffic congestion? 1-yes, 0-no
x3	If sometimes delay, on average how many minutes do you delay?
x4	If sometimes delay, do you 1-perform additional work, 2-engage in non-work activities, or 3-do both?
x5	If sometimes delay, how many times have you delayed in the past week?
x6	Mode of transportation used work-to-home: 1-car SOV, 2-carpool, 3-vanpool, 4-bus, 5 other.
x7	Primary route (work-to-home): 1-I90, 2-I5, 3-SR520, 4-I405, 5-other
x8	Do you generally encounter traffic congestion on you work-to-home trip? 1-yes, 2-no
x9	Age: 1-(<25), 2-(26-30), 3-(31-35), 4-(36-40), 5-(41-45), 6-(46-50), 7-(>50)
x10	Gender: 1-male, 0-female
x11	Number of cars in household
x12	Number of children in household
x13	Income: 1 - less than 20000, 2 - 20000 to 29999, 3 - 30000 to 39999, 4 - 40000 to 49999, 5 - 50000 to 59999, 6 - >60000
x14	Do you have flexible work hours? 1-yes, 0-no
x15	Distance from work to home (in miles)
x16	Face LOS D or worse? 1-yes, 0-no
x17	Ratio of actual travel time to free-flow travel time
x18	Population of work zone
x19	Retail employment in work zone
x20	Service employment in work zone
x21	Size of work zone (in acres)

```
|-> read;nvar=21;nobs=204;file=U:\00Work-Purdue\Book\Book2e-Data\Ex11-6.txt$
Reading data file as space delimited format.
|-> create;if(x7=3)sr520=1$
|-> create;if(x7=2)I5=1$
|-> create;if(x13=6)highinc=1$
|-> create;if(x6=1)sov=1$
|-> dstat;rhs=x5$

--------+---------------------------------------------------------------------
        |                  Standard                                    Missing
Variable|         Mean    Deviation      Minimum      Maximum    Cases  Values
--------+---------------------------------------------------------------------
      X5|      .862745     1.313359          0.0          5.0      204       0
--------+---------------------------------------------------------------------
|-> histogram;rhs=x5$
```




```
Descriptive Statistics for   1 variables
DSTAT results are matrix LASTDSTA in current project.
|-> poisson;lhs=x5
    ;rhs=one,sr520,x14,x15,x17
    ;rh2=one,sov
    ;zip
    ;limit=6;truncation;upper$

Iterative procedure has converged
Normal exit:  16 iterations. Status=0, F=    .2208137D+03

------------------------------------------------------------------------------
Zero Inflated Poisson      Regression Model
Logistic distribution used for splitting model.
ZIP term in probability is F[tau x Z(i)     ]
Comparison of estimated models
            Pr[0|means]       Number of zeros        Log-likelihood
Poisson          .56606   Act.=   126 Prd.=   115.5      -226.14085
Z.I.Poisson      .56590   Act.=   126 Prd.=   115.4      -220.81374
Note, the ZIP log-likelihood is not directly comparable.
ZIP model with nonzero Q does not encompass the others.
Vuong statistic for testing ZIP vs. unaltered model is      1.2675
Distributed as standard normal. A value greater than
+1.96 favors the zero altered Z.I.Poisson model.
A value less than -1.96 rejects the ZIP model.
```

```
--------+------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     X5 |  Coefficient       Error       z    |z|>Z*         Interval
--------+------------------------------------------------------------------
        |Poisson/NB/Gamma regression model.................................
Constant|   -3.49286***      .41472    -8.42  .0000    -4.30569  -2.68002
   SR520|    -.38554*        .22534    -1.71  .0871     -.82720    .05612
     X14|    -.22509         .14615    -1.54  .1235     -.51155    .06137
     X15|     .03696**       .01742     2.12  .0339      .00282    .07111
     X17|    1.91280***      .20606     9.28  .0000     1.50893   2.31667
        |Zero inflation model.............................................
Constant|    -.37721         .46587     -.81  .4181    -1.29029    .53588
     SOV|   -2.32491        1.44804    -1.61  .1084    -5.16302    .51320
--------+------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Feb 09, 2016 at 10:04:25 AM
--------------------------------------------------------------------------------

|-> poisson;lhs=x5
    ;rhs=one,sr520,x14,x15,x17
    ;rh2=one,sov
    ;zip
    ;model=N
    ;limit=6;truncation;upper$
Normal exit:  13 iterations. Status=0, F=    .2206845D+03
Iterative procedure has converged
Normal exit:  39 iterations. Status=0, F=    .2184568D+03

--------------------------------------------------------------------------------
Zero Inflated Neg.Binomial Regression Model
Logistic distribution used for splitting model.
ZIP term in probability is F[tau x Z(i)     ]
Comparison of estimated models
            Pr[0|means]       Number of zeros        Log-likelihood
Poisson          .56606   Act.=   126 Prd.=   115.5      -226.14085
Neg. Bin.        .69509   Act.=   126 Prd.=   141.8      -220.68452
Z.I.Neg_Bin      .56133   Act.=   126 Prd.=   114.5      -218.45677
Note, the ZIP log-likelihood is not directly comparable.
ZIP model with nonzero Q does not encompass the others.
Vuong statistic for testing ZIP vs. unaltered model is       .9539
Distributed as standard normal. A value greater than
+1.96 favors the zero altered Z.I.Neg_Bin model.
A value less than -1.96 rejects the ZIP model.
--------+------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     X5 |  Coefficient       Error       z    |z|>Z*         Interval
--------+------------------------------------------------------------------
        |Poisson/NB/Gamma regression model.................................
Constant|   -3.94630***      .53113    -7.43  .0000    -4.98731  -2.90530
   SR520|    -.47449*        .28165    -1.68  .0921    -1.02652    .07753
     X14|    -.22298         .18461    -1.21  .2271     -.58481    .13885
     X15|     .03830*        .02174     1.76  .0781     -.00431    .08091
     X17|    2.12042***      .28515     7.44  .0000     1.56153   2.67932
        |Dispersion parameter.............................................
   Alpha|     .30167         .34165      .88  .3773     -.36795    .97128
        |Zero inflation model.............................................
Constant|    -.78630         .67128    -1.17  .2415    -2.10199    .52939
     SOV|   -28.6576      .9160D+12      .00 1.0000 ***********  ***********
--------+------------------------------------------------------------------
nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
***, **, * ==>  Significance at 1%, 5%, 10% level.
```

```
Model was estimated on Feb 09, 2016 at 10:04:25 AM
------------------------------------------------------------------------------

|-> poisson;lhs=x5
    ;rhs=one,sr520,x14,x15,x17
    ;rh2=one,x17
    ;zip=normal
    ;rpm;pts=200;halton
    ;fcn=x14(n)
    ;limit=6;truncation;upper$


Iterative procedure has converged
Normal exit:  22 iterations. Status=0, F=    .1956229D+03

------------------------------------------------------------------------------
Random Coefficients  PoissnRE Model
Dependent variable                  X5
Log likelihood function      -195.62293
Restricted log likelihood    -317.38151
Chi squared [  1](P= .000)    243.51717
Significance level               .00000
McFadden Pseudo R-squared      .3836348
Estimation based on N =    204, K =   8
Inf.Cr.AIC  =    407.2 AIC/N =    1.996
Sample is  1 pds and    204 individuals
Simulation  based on   200 Halton draws
Zero Inflated Poisson Regression
(Upper) truncation limit is        6.00
Logistic regime switch prob.
--------+---------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
      X5|  Coefficient       Error       z    |z|>Z*         Interval
--------+---------------------------------------------------------------------
        |Nonrandom parameters.................................................
Constant|     .58136         .66979      .87  .3854     -.73140   1.89412
   SR520|    -.60140***      .22515    -2.67  .0076    -1.04268   -.16011
     X15|    -.01672         .02272     -.74  .4619     -.06125    .02782
     X17|     .16736         .30748      .54  .5862     -.43529    .77001
        |Means for random parameters.........................................
     X14|    -.27425         .17445    -1.57  .1159     -.61617    .06768
        |Scale parameters for dists. of random parameters....................
     X14|     .25511**       .10546     2.42  .0156      .04842    .46180
        |Variables in ZERO regime logit probability..........................
Constant|    11.7993***     2.61264     4.52  .0000      6.6786   16.9200
     X17|   -7.51639***     1.75283    -4.29  .0000   -10.95187  -4.08091
--------+---------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Feb 09, 2016 at 10:05:09 AM
------------------------------------------------------------------------------
```