Queensland University of Technology
Transport Data Analysis and Modeling Methodologies

Lab Session #12
(Random Parameters Count-Data Models)

You are given accident, evirnomental, traffic, and roadway geometric data from 275 segments of highway in Washington State. The data are from 1990. Your task is to estimate a count-data model for the total number of accidents on these segments. See Example 16.2 in the text.

The random parameter Poisson and negative binomial models are derived by making the estimable parameters,

$$\beta_n = \beta + \omega_n$$

where ω_n is a randomly distributed term (for example a normally distributed term with mean zero and variance σ^2). With this equation, the Poisson parameter becomes $\lambda_n|\boldsymbol{\omega}_n = EXP(\boldsymbol{\beta}_n\mathbf{X}_n)$ in the Poisson model and $\lambda_n|\boldsymbol{\omega}_n = EXP(\boldsymbol{\beta}_n\mathbf{X}_n + \varepsilon_n)$ in the negative binomial with the corresponding probabilities for Poisson or negative binomial now $P(y_i|\boldsymbol{\omega}_i)$. With this, the log-likelihood can be written as,

$$LL = \sum_{\forall n} ln \int_{\boldsymbol{\omega}_n} g(\boldsymbol{\omega}_n) P(y_n | \boldsymbol{\omega}_n) \, d\boldsymbol{\omega}_n$$

where $g(.)$ is the probability density function of the $\boldsymbol{\omega}_i$. As was the case with the mixed logit model described previously, because probability estimations are computationally cumbersome, a simulation-based maximum likelihood method is again used (with Halton draws again being an efficient alternative to random draws).

In your specification, consider random variable possibilities including constant or fixed (C), normally distributed (N) and log-normally distributed (L).

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 (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 (in file Ex16-2.txt):

Variable Number	**Explanation**
ID	Segment ID number
FREQ	Number of accidents
ROUTE	Route Number
LENGTH	Segment length in miles
INCLANES	Number of lanes in increasing milepost direction
DECLANES	Number of lanes in decreasing milepost direction
WIDTH	Total combined width of all lanes
MIMEDSH	Minimum median shoulder in feet
MXMEDSH	Maximum median shoulder in feet
SPEED	Speed limit (mi/h)
URB	Indicates urban area (1=yes, 0=no)
FC	Functional class (1=local, 2=collector, 3=arterial, 4=principal arterial, 5=interstate)
AADT	Average Annual Daily Traffic
SINGLE	Daily percentage of single unit trucks
DOUBLE	Daily percentage of tractor and trailer trucks
TRAIN	Daily percentage of tractor and two-trailer trucks
PEAKHR	Percent of daily traffic in the peak hour
GRADEBR	Number of grade breaks in the segment
MIGRADE	Minimum grade in the segment
MXGRADE	Maximum grade in the segment
MXGRDIFF	Maximum grade difference in the segment
TANGENT	Tangent length in the segment
CURVES	Number of cureves in the segment

MINRAD	Minimum radius in feet
ACCESS	Segment access control (0=none, 1=partial, 3=full)
MEDWIDTH	Median width (1=less than 30ft; 2=30 to 40ft; 3=40 to 50ft; 4=50 to 60ft to 5=high)
FRICTION	Friction value (0 to 100 with 100 being high)
ADTLANE	Average daily travel per lane
SLOPE	Segment slope (0=flat, 1=slight, 2=medium, 3=high)
INTECHAG	Indicates number of interganges in the segment
AVEPRE	Average precipitation per month in inches
AVESNOW	Average snowfall per month in inches

```
--> read;nvar=32;nobs=275;names=ID,FREQ,ROUTE,LENGTH,INCLANES,DECLANES,WIDTH,
    MIMEDSH,MXMEDSH,SPEED,URB,FC,AADT,SINGLE,DOUBLE,TRAIN,PEAKHR,GRADEBR,
    MIGRADE,MXGRADE,MXGRDIFF,TANGENT,CURVES,MINRAD,ACCESS,MEDWIDTH,
    FRICTION,ADTLANE,SLOPE,INTECHAG,AVEPRE,AVESNOW
    ;FILE=D:/old_drive_d/book/book2e-Data/Ex16-2.txt$
--> create;expose=aadt*length*365/100000000$
--> negbin;lhs=freq;rhs=one,expose
    ;rpm;pts=200;halton
    ;fcn=expose(n);marginal effects$

+---------------------------------------------+
| Poisson  Regression Start Values for FREQ   |
| Maximum Likelihood Estimates                |
| Model estimated: Oct 23, 2013 at 02:15:11PM.|
| Dependent variable                  FREQ    |
| Weighting variable                  None    |
| Number of observations               275    |
| Iterations completed                  10    |
| Log likelihood function       -13794.16     |
| Number of parameters                   2    |
| Info. Criterion: AIC =        100.33573     |
|   Finite Sample: AIC =        100.33589     |
| Info. Criterion: BIC =        100.36204     |
| Info. Criterion:HQIC =        100.34629     |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 Constant|    2.41799169       .01804768   133.978   .0000
 EXPOSE  |    1.11300618       .01856630    59.948   .0000    .25784008

Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| Random Coefficients  NegBnReg Model         |
| Maximum Likelihood Estimates                |
| Model estimated: Oct 23, 2013 at 02:15:19PM.|
| Dependent variable                  FREQ    |
| Weighting variable                  None    |
| Number of observations               275    |
| Iterations completed                  12    |
| Log likelihood function       -986.9452     |
| Number of parameters                   4    |
| Info. Criterion: AIC =           7.20687    |
|   Finite Sample: AIC =           7.20741    |
| Info. Criterion: BIC =           7.25948    |
| Info. Criterion:HQIC =           7.22799    |
| Restricted log likelihood     -13794.16     |
| McFadden Pseudo R-squared      .9284520     |
| Chi squared                   25614.44     |
| Degrees of freedom                     1    |
| Prob[ChiSqd > value] =          .0000000    |
| Sample is  1 pds and     275 individuals.   |
| Negative binomial regression model          |
| Simulation based on 200 Halton draws        |
+---------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Nonrandom parameters					
Constant	1.90434975	.07035639	27.067	.0000	
Means for random parameters					
EXPOSE	2.62050637	.23106908	11.341	.0000	.25784008
Scale parameters for dists. of random parameters					
EXPOSE	.62186133	.13914793	4.469	.0000	
Dispersion parameter for NegBin distribution					
ScalParm	1.59206464	.11345999	14.032	.0000	

Implied standard deviations of random parameters

Matrix S.D_Beta has 1 rows and 1 columns.

	1
1	.62186

Partial derivatives of expected val. with respect to the vector of characteristics.
They are computed at the means of the Xs.
Conditional Mean at Sample Point 13.1974
Scale Factor for Marginal Effects 13.1974

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	25.1325027	1.03104841	24.376	.0000	
EXPOSE	34.5839221	5.10998652	6.768	.0000	.25784008