PURDUE
UNIVERSITY

School of Civil Engineering

CE615 – Statistical and Econometric Methods II

Assignment #9
(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.

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 Random-NB-R.csv):

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
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
CUMTAN	Cumulative 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)
ACCYR	Accident year (0 for 1990)
MEDWIDTH	Median width (1=less than 30ft; 2=30 to 40ft; 3=40 to 50ft; 4=50 to 60ft to 5=high)
PRECIP	Indicates presence of a precipitation (1=yes; 0=no)
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)
ELEVDIFF	Elevation difference in the segment in feet
VEGET	Indicates presence of vegitation in the clear zone (1=yes; 0=no)
OTHEROBS	Indicates presence of other obstructions in the clear zone (1=yes; 0=no)
DITCH	Indicates presence of a ditch (1=yes; 0=no)
PILLAR	Indicates presence of a pillars in clear zone (1=yes; 0=no)
GRAVSHD	Indicates gravel shoulder (1=yes; 0=no)
CROSOVER	Indicates presence of a median crossover (1=yes; 0=no)
INTECHAG	Indicates number of interganges in the segment
EXITENTR	Indicates presence of exit/entrance ramps (1=yes; 0=no)
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:\NEW_LAPTOP\CE697M\Random-NB-R.csv"$
--> create;laneadt=aadt/(inclanes+declanes)$
--> create;lnlanadt=log(laneadt)$
--> create;lnaadt=log(aadt)$
--> create;density=laneadt/length$
--> create;if(friction<=30)lowfri=1$
--> create;if(friction>30&friction<50)medfri=1$
--> create;if(friction>=50)hifri=1$
--> create;curvmile=curves/length$
--> create;if(curvmile<=0.5)lowcvmil=1;(else)lowcvmil=0$
--> create;if(curvmile>0.5&curvmile<=2.5)medcvmil=1;(else)medcvmil=0$
--> create;if(curvmile>2.5)hicvmil=1;(else)hicvmil=0$
--> create;truck=single+double+train$
--> create;notruck=truck*aadt$
--> create;if(medwidth=1)med030=1$
--> create;if(medwidth=2)med3040=1$
--> create;if(medwidth=3)med4050=1$
--> create;if(medwidth=4)med5060=1$
--> create;if(medwidth=5)med60=1$
--> create;if(speed<=50)speed1=1$
--> create;if(speed<=55)speed2=1$
--> create;if(speed>55)speed3=1$
--> create;if(speed>=55)speed4=1$
--> create;if(fc=1)local=1$
--> create;if(fc=5)intstate=1$
--> create;if(access=0)none   =1$
--> create;if(access=1)partial=1$
--> create;if(access=2)full   =1$
--> create;if(slope=0)flat=1$
--> create;if(slope=1)slight=1$
--> create;if(slope=2)medium=1$
--> create;if(slope=0 |slope= 1)slpflat=1;(else)slpflat=0$
--> create;if(slope=2)slpmed=1;(else)slpmed=0$
--> create;if(avepre<=1.5)lowpre=1;(else)lowpre=0$
--> create;if(avepre>1.5&avepre<=2.5)medpre=1;(else)medpre=0$
--> create;if(avepre>2.5)hipre=1;(else)hipre=0$
--> create;if(avesnow<=1)norsnow=1$
--> create;if(avesnow>1)hisnow=1$
--> create;lanewid=(inclanes+declanes)/width$
--> dstat;rhs=lanewid$
Descriptive Statistics
All results based on nonmissing observations.
================================================================================
Variable        Mean         Std.Dev.        Minimum         Maximum      Cases
================================================================================
--------------------------------------------------------------------------------
All observations in current sample
--------------------------------------------------------------------------------
LANEWID   .809060949E-01  .644470645E-02  .392156863E-01  .869565217E-01    275

--> create;if(lanewid<12)nlanwid=1;(else)nlanwid=0$
--> create;if(lanewid>12)wlanwid=1;(else)wlanwid=0$
--> create;expose=aadt*length*365/100000000$
--> create;cpm=curves/length$
--> create;intpm=intechag/length$
--> create;gbrpm=gradebr/length$
--> negbin;lhs=freq
    ;rhs=one,expose,lowpre,gbrpm,intpm,cpm,hisnow,friction
    ;rpm;pts=200;halton
    ;fcn=expose(n),intpm(n),cpm(n),hisnow(n);marginal effects$
```

```
+--------------------------------------------+
| Poisson  Regression Start Values for FREQ  |
| Maximum Likelihood Estimates               |
| Model estimated: Sep 14, 2010 at 10:58:02AM.|
| Dependent variable                  FREQ   |
| Weighting variable                  None   |
| Number of observations               275   |
| Iterations completed                  10   |
| Log likelihood function       -1930.447    |
| Restricted log likelihood     -3178.572    |
| Chi squared                    2496.251    |
| Degrees of freedom                    7    |
| Prob[ChiSqd > value] =         .0000000    |
+--------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	3.47694246	.16673315	20.853	.0000	
LOWPRE	-.45972263	.04068160	-11.301	.0000	.34545455
GBRPM	.00983845	.01048671	.938	.3482	1.88675058
FRICTION	-.01498735	.00329092	-4.554	.0000	49.6541818
EXPOSE	1.04887778	.02179082	48.134	.0000	.25784008
INTPM	.26642089	.02606694	10.221	.0000	.50825186
CPM	-.21030053	.01886909	-11.145	.0000	1.44004619
HISNOW	-.16300637	.03197338	-5.098	.0000	.38545455

```
Normal exit from iterations. Exit status=0.


+--------------------------------------------+

| Random Coefficients  NegBnReg Model        |
| Maximum Likelihood Estimates               |
| Model estimated: Sep 14, 2010 at 11:00:51AM.|
| Dependent variable                  FREQ   |
| Weighting variable                  None   |
| Number of observations               275   |
| Iterations completed                  37   |
| Log likelihood function       -959.5771    |
| Restricted log likelihood     -1930.447    |
| Chi squared                    1941.740    |
| Degrees of freedom                    4    |
| 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	2.80652833	.23018347	12.193	.0000	
LOWPRE	-.41469783	.04984634	-8.320	.0000	.34545455
GBRPM	-.05174123	.01374911	-3.763	.0002	1.88675058
FRICTION	-.01285173	.00427592	-3.006	.0027	49.6541818
Means for random parameters					
EXPOSE	2.70353472	.08077514	33.470	.0000	.25784008
INTPM	.14928785	.03659369	4.080	.0000	.50825186
CPM	-.16475207	.02354290	-6.998	.0000	1.44004619
HISNOW	-.17609228	.04356986	-4.042	.0001	.38545455
Scale parameters for dists. of random parameters					
EXPOSE	.94835072	.04769983	19.882	.0000	
INTPM	.52574076	.02761391	19.039	.0000	
CPM	.09475685	.01353526	7.001	.0000	
HISNOW	.59979248	.03453036	17.370	.0000	
Dispersion parameter for NegBin distribution					
ScalParm	6.58221367	.59646021	11.035	.0000	

Implied standard deviations of random parameters

Matrix S.D_Beta has 4 rows and 1 columns.

	1
1	.94835
2	.52574
3	.09476
4	.59979

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 10.9719
Scale Factor for Marginal Effects 10.9719

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	30.7928275	2.53342719	12.155	.0000	
LOWPRE	-4.55000523	.53542273	-8.498	.0000	.34545455
GBRPM	-.56769737	.15055969	-3.771	.0002	1.88675058
FRICTION	-.14100732	.04694485	-3.004	.0027	49.6541818
EXPOSE	29.6627963	.99449202	29.827	.0000	.25784008
INTPM	1.63796495	.39969347	4.098	.0000	.50825186
CPM	-1.80763613	.26098342	-6.926	.0000	1.44004619
HISNOW	-1.93205934	.44559488	-4.336	.0000	.38545455