Statistical and Econometric Methods for Transportation Data Analysis

Chapter 14 – Ordered Probability Models

Example 14.3 **(Third Edition)**
Ordered Discrete Data – Hierarchical Ordered Probit Model

You are given injury-severity data from 4,999 single-vehicle highway accidents in the state of Washington. The data – collected over a three-year-period, between 2011 and 2013 – were jointly obtained from the SHRP2 Roadway Information Database and the Highway Safety Information System, and include roadway characteristics, traffic characteristics, accident-related characteristics, vehicle-, driver-, and collision-specific information. There are four possible severity outcomes: no evident injury (property-damage-only and possible injury); evident non-incapacitating injury; incapacitating injury; and fatality.

Your task is to study the factors that affect accident injury severities using ordered probability models, to account for the ordinal nature of the injury severity outcomes. One basic limitation of the traditional ordered probit/logit model is that the thresholds defining the ordered probability outcomes are considered to be fixed across the observations. The hierarchical ordered probit (HOPIT) model can to some extent address this limitation, by allowing the thresholds to vary as a function of explanatory parameters:

$$\mu_{i,j} = \mu_{i,j-1} + \exp(t_j + \mathbf{d}_j \mathbf{S}_i) \tag{1}$$

where, t is the intercept for each threshold, $\mathbf{S}$ are vectors of variables affecting the thresholds, and $\mathbf{d}$ are vectors of estimable parameters for $\mathbf{S}$.

To that end, your task is to estimate a hierarchical ordered probit model, in order to investigate the factors that determine the ordered probability outcomes (injury-severities) and the corresponding thresholds.

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 model specification. In other words, 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 (in file HOPIT.txt):

Variable	Explanation
1	HSIS case number ID
2	RID case number ID
3	Link ID
4	Injury severity (0: no injury; 1: injury; 2: serious injury; 3: fatality)
5	County (1: Pierce; 2: King; 3: Snohomish)
6	City (1: Seattle; 0: otherwise)
7	Year of crash occurrence
8	Year and month of the crash
9	Month of crash occurrence
10	Day of week of crash occurrence (1: Monday; 2: Tuesday; 3: Wednesday; 4: Thursday; 5: Friday; 6: Saturday; 7: Sunday)
11	Specific quarter of the hour when the crash occurred (1: First quarter of the hour; 2: Second quarter of the hour; 3: Third quarter of the hour; 4: Fourth quarter of the hour)
12	Time of day of crash occurrence (in decimal format – multiply with 24 to get approximate hour of the crash)
13	Time of day of crash occurrence (in 24h format)
14	Pedal cyclist involved in the crash (1: involved; 0 otherwise)
15	Pedestrian involved in the crash (1: involved; 0 otherwise)
16	Number of motor vehicles involved
17	Sobriety (1: HBD-Ability impaired; 2: HBD-Ability not impaired; 3: Sobriety unknown; 4: Had not been drinking; 5: Ability impaired (tox test); 6: Ability not impaired (tox test); 7: Had not been drinking (tox test); 9:Unknown)
18	Type of collision (0: Vehicle going straight hits pedestrian; 1: Vehicle turning right hits pedestrian; 2: Vehicle turning left hits pedestrian; 4: Vehicle hits pedestrian - all other actions; 23: From same direction - all others; 31: Not stated; 45: Vehicle-pedal cyclist; 48: Domestic animal (cat, dog, etc.); 50: Fixed object; 51: Other object; 52: Vehicle overturned; 53: Person fell, jumped or was pushed from vehicle; 54: Fire started in vehicle; 56: Breakage of any part of the vehicle resulting in injury or in further property damage; 57: All other non-collision; 85: Vehicle stroke deer; 86: Vehicle stroke elk; 87: Vehicle stroke all other non-domestic animals)
19	Junction presence (0: Roundabout; 1: At intersection and intersection related; 2: Intersection related but not at intersection; 3: At driveway; 4: Not at intersection and not intersection related; 5: At intersection but not intersection related; 7: Driveway related but not at Driveway)
20	Weather conditions (0: Unknown; 1: Clear or partly cloudy; 2: Overcast; 3: Raining; 4: Snowing; 5: Fog or smog or smoke; 6: Sleet or hail or freezing rain; 7: Severe crosswind; 8: Blowing sand or dirt or snow; 9: Other)
21	Lighting condition (1: Daylight; 2: Dawn; 3: Dusk; 4: Dark-street with lights on; 5: Dark-street with lights off; 6: Dark-no street lights; 7: Other; 9: Unknown)
22	Location characteristics (1: Parking lot; 2: Bridge or overpass; 3: Underpass or tunnel;

	4: Rest area or turnout; 8: School zone)
23	Roadway curvature characteristics (1: Straight and level alignment; 2: Straight alignment with grade; 3: Straight alignment at hill; 4: Straight alignment with sag curve; 5: Horizontal curve at level profile; 6: Horizontal curve with grade; 7: Horizontal curve at hill; 8: Horizontal curve with sag curve; 9: Unknown)
24	Stolen vehicle indicator (1: The vehicle was stolen after the crash; 0 otherwise)
25	Driver's age
26	Driver's gender (0: female; 1:male)
27	Ejection status (1: Not ejected; 2: Partially ejected; 3: Totally ejected; 4: Unknown if ejected)
28	Restraining system (1: Lap and shoulder restraint used; 2: Lap belt restraint used; 3: No restraints used; 4: Shoulder belt restraint used; 5: Unknown)
29	Blood alcohol concentration test results (in g/dL)
30	Driver's license status (1: No license; 2: License; 3: Not stated; 4: Unknown)
31	Vehicle type (1: Passenger car; 2: Pickup, panel truck or vanette under 10,000 lbs; 3: Truck (flatbed, van, etc.); 4: Truck and trailer; 5: Truck tractor; 6: Truck tractor and semi-trailer; 7: Truck - double trailer combinations; 9: Taxi; 12: Motorcycle; 14: Other)
32	Vehicle model-year
33	Traffic control (1: Flashing amber; 2: Flashing red; 3: No traffic control; 4: Officer/Flagger; 5: Other traffic control; 6: Railroad signal; 7: Traffic signal; 8: Stop sign; 9: Yield; 10: Unknown)
34	Posted speed limit in mph (0 reflects unknown posted speed limit)
35	Roadway type (0: Unknown; 1: One-way; 2: Two-way undivided; 3: Two-way divided with barrier; 4: Two-way divided, with no barrier; 5: Reversible road; 6: Interchange ramp; 8: Continuous two-way left turn center lanes)
36	State (1: Washington; 0: Other)
37	Vehicle action during crash (1: Braking; 2: Changing lane; 3: Going straight ahead; 4: Going the wrong way on divided highway; 5: Hit a legally parked, unoccupied vehicle; 6: Making left turn; 7: Making right turn; 8: Making U-turn; 9: Merging (entering traffic); 10: Other; 11: Overtaking and passing; 12: Slowing down; 13: Entering the traffic stream)
38	Number of axles (for trucks)
39	Crash occurrence on- or off-roadway (0: Off-roadway; 1: On-roadway)
40	Annual average daily traffic (AADT) in vehicles per day
41	Lane width (in feet)
42	Median width (in feet)
43	Number of lanes
44	Shoulder width (in feet)
45	Segment length (in miles)
46	Roadway classification (1: Urban; 2: Rural)
47	Horizontal curve length (in feet)
48	Horizontal curve radius (in feet)
49	Access control (1: Yes; 2: No)

```
|-> read; nvar=49;nobs=4999;file=U:\00Work-Purdue\Book\Ex14-3(3E).txt$
|-> skip
|-> histogram;rhs=x4;list$

Histogram for X4        NOBS =    4999, Too low:     0, Too high:     0
Bin  Value of X4       Frequency     Cumulative Frequency
=============================================================================
  0       0           3576 ( .7153)        3576 ( .7153)
  1       1           1323 ( .2647)        4899 ( .9800)
  2       2             72 ( .0144)        4971 ( .9944)
  3       3             28 ( .0056)        4999 (1.0000)

|-> create;if(x9=12|x9=1|x9=2)winter1=1$
|-> create;if(x18=50)fixobj=1$
|-> create;if(x25<25)young=1$
|-> create;xx40=x40/10000$
|-> dstat;rhs=young,fixobj,xx40,x36,winter1,X26$
--------+-------------------------------------------------------------------
        |                  Standard                                      Missing
Variable|         Mean    Deviation      Minimum      Maximum    Cases  Values
--------+-------------------------------------------------------------------
   YOUNG|      .344069      .475111          0.0          1.0     4999       0
  FIXOBJ|      .826565      .378661          0.0          1.0     4999       0
    XX40|      11.2802     6.226863          0.0      22.9101     4094     905
     X36|      .941988      .233789          0.0          1.0     4999       0
 WINTER1|      .285857      .451867          0.0          1.0     4999       0
     X26|       .60212      .489509          0.0          1.0     4999       0
--------+-------------------------------------------------------------------

|-> ordered;lhs=x4;rhs=one,young,fixobj,xx40,x36$
------------------------------------------------------------
Deleted    905 observations with missing data. N is now   4094
------------------------------------------------------------
Iterative procedure has converged
Normal exit:  12 iterations. Status=0, F=    .2798481D+04
+---------------------------------------------------------------------+
|                CELL FREQUENCIES FOR ORDERED CHOICES                 |
+---------------------------------------------------------------------+
|               Frequency        Cumulative  < =    Cumulative  > =  |
|Outcome      Count    Percent   Count    Percent   Count    Percent |
|----------- ------- ---------  ------- ---------  ------- --------- |
|X4=00          2926   71.4704     2926   71.4704     4094  100.0000 |
|X4=01          1079   26.3556     4005   97.8261     1168   28.5296 |
|X4=02            65    1.5877     4070   99.4138       89    2.1739 |
|X4=03            24     .5862     4094  100.0000       24     .5862 |
+---------------------------------------------------------------------+

------------------------------------------------------------------------------
Ordered Probability Model
Dependent variable                   X4
Log likelihood function     -2798.48065
Restricted log likelihood   -2814.26557
Chi squared [  4](P= .000)     31.56985
Significance level               .00000
McFadden Pseudo R-squared      .0056089
Estimation based on N =   4094, K =   7
Inf.Cr.AIC  =   5611.0 AIC/N =    1.371
Underlying probabilities based on Normal
```

```
--------+------------------------------------------------------------------
        |                 Standard           Prob.      95% Confidence
     X4 |  Coefficient       Error       z    |z|>Z*         Interval
--------+------------------------------------------------------------------
        |Index function for probability.....................................
Constant|    -.69790***      .10041    -6.95  .0000     -.89469   -.50111
   YOUNG|    -.03453         .04320     -.80  .4240     -.11920    .05013
  FIXOBJ|    -.14770***      .05062    -2.92  .0035     -.24692   -.04848
    XX40|     .01574***      .00327     4.81  .0000      .00933    .02215
     X36|     .08548         .08784      .97  .3305     -.08669    .25765
        |Threshold parameters for index.....................................
  Mu(01)|    1.46161***      .04358    33.54  .0000     1.37620   1.54702
  Mu(02)|    1.96794***      .07173    27.43  .0000     1.82735   2.10854
--------+------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Apr 12, 2017 at 00:48:24 PM
--------------------------------------------------------------------------------

|-> ordered;lhs=x4;rhs=one,young,fixobj,xx40,x36;HO1=winter1,X26;margin$
------------------------------------------------------------
Deleted    905 observations with missing data. N is now   4094
------------------------------------------------------------
Iterative procedure has converged
Normal exit:  15 iterations. Status=0, F=    .2787472D+04
--------------------------------------------------------------------------------
Ordered Probability Model
Dependent variable                   X4
Log likelihood function     -2787.47219
Restricted log likelihood   -2814.26557
Chi squared [  4](P= .000)     53.58677
Significance level               .00000
McFadden Pseudo R-squared      .0095206
Estimation based on N =   4094, K =   9
Inf.Cr.AIC  =   5592.9 AIC/N =    1.366
Underlying probabilities based on Normal
HOPIT (covariates in thresholds) model
--------+------------------------------------------------------------------
        |                 Standard           Prob.      95% Confidence
     X4 |  Coefficient       Error       z    |z|>Z*         Interval
--------+------------------------------------------------------------------
        |Index function for probability.....................................
Constant|    -.70903***      .10194    -6.96  .0000     -.90882   -.50923
   YOUNG|    -.02869         .04345     -.66  .5091     -.11386    .05648
  FIXOBJ|    -.13848***      .05084    -2.72  .0065     -.23812   -.03884
    XX40|     .01564***      .00329     4.75  .0000      .00918    .02209
     X36|     .08862         .08808     1.01  .3144     -.08401    .26125
        |Estimates of t(j) in mu(j)=exp[t(j)+d*z].....................
Theta(1)|     .56206***      .05415    10.38  .0000      .45593    .66819
Theta(2)|     .87792***      .06304    13.93  .0000      .75437   1.00146
        |Threshold covariates mu(j)=exp[t(j)+d*z].....................
 WINTER1|     .03654         .06671      .55  .5839     -.09422    .16730
     X26|    -.28484***      .06282    -4.53  .0000     -.40797   -.16171
--------+------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Apr 12, 2017 at 00:48:25 PM
--------------------------------------------------------------------------------
```

```
------------------------------------------------------------------------------
Marginal effects for ordered probability model
M.E.s for dummy variables are Pr[y|x=1]-Pr[y|x=0]
Names for dummy variables are marked by *.
--------+---------------------------------------------------------------------
        |     Partial                              Prob.      95% Confidence
     X4 |      Effect    Elasticity      z    |z|>Z*         Interval
--------+---------------------------------------------------------------------
        |--------------[Partial effects on Prob[Y=00] at means]--------------
  *YOUNG|     .00970         .01331      .47  .6384     -.03075    .05015
 *FIXOBJ|     .04803         .06590      .71  .4781     -.08466    .18071
   XX40 |    -.00530        -.08205     -.07  .9453     -.15678    .14618
   *X36 |    -.02935        -.04027    -1.18  .2396     -.07827    .01957
        |--------------[Partial effects on Prob[Y=01] at means]--------------
  *YOUNG|    -.00834        -.03294     -.56  .5726     -.03732    .02063
 *FIXOBJ|    -.04079        -.16109    -1.22  .2233     -.10645    .02486
   XX40 |     .00455         .20284      .07  .9431     -.12045    .12956
   *X36 |     .02544*        .10048     1.75  .0806     -.00310    .05399
        |--------------[Partial effects on Prob[Y=02] at means]--------------
  *YOUNG|    -.00099        -.07062     -.21  .8333     -.01017    .00820
 *FIXOBJ|    -.00520        -.37222     -.23  .8191     -.04972    .03932
   XX40 |     .00054         .43795      .06  .9547     -.01814    .01922
   *X36 |     .00286         .20467      .27  .7890     -.01807    .02378
        |--------------[Partial effects on Prob[Y=03] at means]--------------
  *YOUNG|    -.00037        -.09235     -.15  .8777     -.00510    .00436
 *FIXOBJ|    -.00204        -.50602     -.16  .8718     -.02677    .02270
   XX40 |     .00020         .57468      .05  .9589     -.00759    .00800
   *X36 |     .00105         .26103      .18  .8567     -.01035    .01245
--------+---------------------------------------------------------------------
z, prob values and confidence intervals are given for the partial effect
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Apr 12, 2017 at 00:48:25 PM
------------------------------------------------------------------------------


|-> ordered;lhs=x4;rhs=one,young,fixobj,xx40,x36
    ;rpm;pts=200;halton
    ;fcn=fixobj(n),xx40(n);margin$

------------------------------------------------------------
Deleted    905 observations with missing data. N is now   4094
------------------------------------------------------------
Iterative procedure has converged
Normal exit:  23 iterations. Status=0, F=    .2798098D+04

------------------------------------------------------------------------------
Random Coefficients  OrdProbs Model
Dependent variable                   X4
Log likelihood function     -2798.09785
Estimation based on N =   4999, K =   9
Inf.Cr.AIC  =   5614.2 AIC/N =    1.123
Sample is  1 pds and   4999 individuals
Simulation  based on   200 Halton draws
Missing data: Skipped      905 individuals.
Ordered probability model
Ordered probit (normal) model
LHS variable = values 0,1,..., 3
--------+---------------------------------------------------------------------
```

X4	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Nonrandom parameters					
Constant	-.69137***	.10076	-6.86	.0000	-.88885	-.49389
YOUNG	-.03555	.04349	-.82	.4136	-.12079	.04969
X36	.09318	.08775	1.06	.2883	-.07882	.26517
	Means for random parameters					
FIXOBJ	-.15095***	.05055	-2.99	.0028	-.25003	-.05187
XX40	.01365***	.00333	4.10	.0000	.00713	.02017
	Scale parameters for dists. of random parameters					
FIXOBJ	.02201	.02280	.97	.3344	-.02268	.06670
XX40	.01780***	.00160	11.11	.0000	.01466	.02094
	Threshold parameters for probabilities					
Mu(01)	1.50369***	.04559	32.99	.0000	1.41434	1.59304
Mu(02)	2.02807***	.07632	26.57	.0000	1.87849	2.17765

***, **, * ==> Significance at 1%, 5%, 10% level.
Model was estimated on Apr 14, 2020 at 11:12:26 AM

Marginal effects for ordered probability model
M.E.s for dummy variables are Pr[y|x=1]-Pr[y|x=0]
Names for dummy variables are marked by *.

X4	Partial Effect	Elasticity	z	Prob. |z|>Z*	95% Confidence Interval	
	[Partial effects on Prob[Y=00] at means]					
*YOUNG	.01192	.01655	.82	.4120	-.01656	.04040
*X36	-.03058	-.04246	-1.09	.2755	-.08554	.02439
*FIXOBJ	.05210***	.07233	2.92	.0035	.01711	.08708
XX40	-.00459***	-.07194	-4.10	.0000	-.00679	-.00240
	[Partial effects on Prob[Y=01] at means]					
*YOUNG	-.01034	-.03955	-.82	.4126	-.03506	.01439
*X36	.02671	.10221	1.08	.2797	-.02172	.07515
*FIXOBJ	-.04457***	-.17053	-2.96	.0031	-.07412	-.01502
XX40	.00398***	.17164	4.11	.0000	.00208	.00588
	[Partial effects on Prob[Y=02] at means]					
*YOUNG	-.00113	-.08080	-.83	.4089	-.00380	.00155
*X36	.00276	.19820	1.15	.2505	-.00195	.00747
*FIXOBJ	-.00527***	-.37849	-2.68	.0074	-.00913	-.00141
XX40	.00044***	.35374	3.95	.0001	.00022	.00065
	[Partial effects on Prob[Y=03] at means]					
*YOUNG	-.00046	-.10244	-.82	.4097	-.00156	.00064
*X36	.00110	.24511	1.17	.2415	-.00074	.00295
*FIXOBJ	-.00225***	-.50038	-2.59	.0096	-.00396	-.00055
XX40	.00018***	.45033	3.73	.0002	.00009	.00027

z, prob values and confidence intervals are given for the partial effect
***, **, * ==> Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
Model was estimated on Apr 14, 2020 at 11:12:26 AM