Statistical and Econometric Methods for Transportation Data Analysis

Chapter 14 – Ordered Probability Models

Example 14.2a
Ordered Discrete Data – Ordered Probit with Random Parameters

As in Example 14.2, a survey of 56 subjects was conducted on freeways in the Seattle area. Each subject drove a vehicle over 40 freeway segments (thus each subject can generate as many as 40 observations if there is no missing data). As they drove over the test segments, they were asked: "How would you rank the roughness of the road on a scale from one to five – with one being the smoothest (or the best) and five being the roughest (or the worst)?" Data were collected on the type of vehicle being used (minivan, pickup, etc.), in-vehicle-cabin noise (dBA), vehicle speed (km/h), socioeconomic information, IRI measurement, age of the roadway surface, information on patching, and the Pavement Structural Condition (PSC). This last term is calculated separately for flexible and rigid pavements based on the amount and severity of various distresses and its values range from 100 (excellent pavement condition) to zero (completely deteriorated pavement).

Your task is to estimate a model of the ordered response of roughness perception while accounting for repeat observations from individual subjects. However, unlike Example 14.2 in the text, you are now to estimate a random parameters model. Note that variable number 3 in Table 14.4 gives the number of observations from each respondent. This is used as "pds" in Limdep as was done for Example 14.2 for using random effects. Using the "pds" in a random parameters model means one parameter is estimated for each person. Leaving this "pds" out means that one parameter is estimated for each observation. Leaving it in would make the most sense in this case. Please provide:

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. (e.g. Discuss the theory behind the inclusion of your selected variables). Include *t*-statistics and justify the sign of your variables.

Available distributions:

n = normal; l = lognormal; u = uniform; t = triangular; d = dome; e = Erlang; w = Weibull
p = exponential; c = nonstochastic (constant)

Variables available for your specification are (in file Ex14-2.txt):

Variable Number	Explanation
1	Individual number
2	Roadway segment number
3	pds for Limdep random effects
4	Roughness ranking: 1 = very smooth; 5 = very rough
5	Sedan: 1 if yes, 0 if no
6	Sport utility vehicle: 1 if yes, 0 if no
7	Pickup: 1 if yes, 0 if no
8	Minivan: 1 if yes, 0 if no
9	Nosie dBA reading
10	Speed in miles per hour
11	Level of service: A=1, B=2, C=3,D=4,E=5,F=6
12	User regularly uses I-5: 1 if yes, 0 if no
13	User regularly uses I-90: 1 if yes, 0 if no
14	User regularly uses I-405: 1 if yes, 0 if no
15	User regularly uses SR-520: 1 if yes, 0 if no
16	Female: 1 if yes, 0 if no
17	Married: 1 if yes, 0 if no
18	Age: 0 = Less than 21; 1 = 21 - 25; 2 = 26-30; 3=31-35; 4 = 36 - 40; 5 = 41 - 45; 6 = 46 - 50; 7 = 51 - 55; 8 = 56 - 60; 9 = 61 - 65; 10 = 66 - 70; 11 = Over 70
19	Income: 0 = no income; 1 = under $15,000; 2 = $15,000 - $24,999; 3 = $25,000 - $34,999; 4 = $35,000 - $44,999; 5 = $45,000 - $54,999; 6 = $55,000 - $64,999; 7 = $65,000 - $74,999; 8 = $75,000 - $84,999; 9 = $85,000 - $99,999; 10 = $100,000 - $150,000; 11 = over $150,000

20	Education: 1 = some high school; 2 = high school diploma; 3 = technical college degree (AA); 4 = college degree (BS or BA) 5 = post-graduate degree
21	Vehicle type normally driven: (miscoded, do not use)
22	Number of household vehicles
23	Household size
24	Number of household infants
25	Number of household children
26	Number of workers
27	International roughness index (IRI) in m/km
28	Roadway surface age
29	Visible wear: 1 if yes, 0 if no
30	Visible joints: 1 if yes, 0 if no
31	Visible patching: 1 if yes, 0 if no
32	Bridge in section: 1 if yes, 0 if no
33	Surface type: 1 if concrete, 0 if asphalt
34	Rut depth in mm
35	Pavement structural condition index (PSC)
36	Section length in miles
37	Number of lanes
38	Cracking present: 1 if yes, 0 if no
39	Scaling present: 1 if yes, 0 if no
40	Faulting present: 1 if yes, 0 if no
41	Spalling present: 1 if yes, 0 if no
42	IRI change from last section (m/km)
43	Nosie change from last section (dBA)

With PDS:

```
--> RESET
Initializing NLOGIT Version 4.0.1 (January 1, 2007).
--> read;nvar=43;nobs=2240;file=D:\old_drive_d\new_laptop\CE697M\pavement-pds...
--> histogram;rhs=x4$
Histogram for X4           NOBS=  2179, Too low:     0, Too high:     0
Bin  Lower limit   Upper limit       Frequency     Cumulative Frequency
=============================================================================
  0        1.000         2.000        344 ( .1579)         344( .1579)
  1        2.000         3.000        769 ( .3529)        1113( .5108)
  2        3.000         4.000        601 ( .2758)        1714( .7866)
  3        4.000         5.000        351 ( .1611)        2065( .9477)
  4        5.000         6.000        114 ( .0523)        2179(1.0000)
--> dstat;rhs=x4$
Descriptive Statistics
All results based on nonmissing observations.
================================================================================
Variable     Mean       Std.Dev.     Minimum      Maximum        Cases Missing
================================================================================
--------------------------------------------------------------------------------
All observations in current sample
--------------------------------------------------------------------------------
X4       |  2.59706      1.09222      1.00000      5.00000         2179      61
--> create;if(x4=1)xx4=0$
--> create;if(x4=2)xx4=1$
--> create;if(x4=3)xx4=2$
--> create;if(x4=4)xx4=3$
--> create;if(x4=5)xx4=4$
--> reject;x4=-999$
--> ordered;lhs=xx4;rhs=one,x16,x27,x28,x43
    ;pds=x3; ;rpm;pts=200;halton
    ;fcn=x16(n),x27(n);marginal effects;parameters$
Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| Random Coefficients  OrdProbs Model         |
| Maximum Likelihood Estimates                |
| Model estimated: Nov 24, 2014 at 10:32:51AM.|
| Dependent variable                  XX4     |
| Weighting variable                 None     |
| Number of observations             2179     |
| Iterations completed                 43     |
| Log likelihood function       -2372.827     |
| Number of parameters                 10     |
| Info. Criterion: AIC =          2.18708     |
|   Finite Sample: AIC =          2.18713     |
| Info. Criterion: BIC =          2.21318     |
| Info. Criterion:HQIC =          2.19662     |
| Unbalanced panel has      56 individuals.   |
| Ordered probability model                   |
| Ordered probit (normal) model               |
| LHS variable = values 0,1,..., 4            |
| Simulation based on 200 Halton draws        |
+---------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Nonrandom parameters					
Constant	-.28353462	.04522951	-6.269	.0000	
X28	.01770372	.00276548	6.402	.0000	18.1404314
X43	.06036166	.00980584	6.156	.0000	.02707664
Means for random parameters					
X16	-.87499473	.05323097	-16.438	.0000	.40339605
X27	.89804819	.04621051	19.434	.0000	1.93683800
Scale parameters for dists. of random parameters					
X16	.80047786	.04387765	18.243	.0000	
X27	.26265348	.01222951	21.477	.0000	
Threshold parameters for probabilities					
MU(1)	1.51794266	.03611482	42.031	.0000	
MU(2)	2.85896541	.03624918	78.870	.0000	
MU(3)	4.40601796	.04734359	93.065	.0000	

Implied standard deviations of random parameters

Matrix S.D_Beta has 2 rows and 1 columns.

	1
1	.80048
2	.26265

Summary of Marginal Effects for Ordered Probability Model (probit)

Variable	Y=00	Y=01	Y=02	Y=03	Y=04	Y=05	Y=06	Y=07
X28	-.0026	-.0045	.0045	.0024	.0001			
X43	-.0087	-.0153	.0154	.0083	.0003			
*X16	.1454	.1879	-.2186	-.1107	-.0041			
X27	-.1297	-.2271	.2285	.1240	.0042			

Without PDS:

```
--> reject;x4=-999$
--> ordered;lhs=xx4;rhs=one,x16,x27,x28,x43
    ;rpm;pts=200;halton
    ;fcn=x16(n),x27(n);marginal effects;parameters$
```

Random Coefficients OrdProbs Model	
Maximum Likelihood Estimates	
Model estimated: Nov 24, 2014 at 11:05:10AM.	
Dependent variable	XX4
Weighting variable	None
Number of observations	2179
Iterations completed	20
Log likelihood function	-2626.411
Number of parameters	10
Sample is 1 pds and 2179 individuals.	
Ordered probability model	
Ordered probit (normal) model	
LHS variable = values 0,1,..., 4	
Simulation based on 200 Halton draws	

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Nonrandom parameters					
Constant	-.31048538	.06749943	-4.600	.0000	
X28	.01715816	.00240993	7.120	.0000	18.1404314
X43	.05160382	.00926410	5.570	.0000	.02707664
Means for random parameters					
X16	-.20189407	.04850043	-4.163	.0000	.40339605
X27	.80059411	.04545172	17.614	.0000	1.93683800
Scale parameters for dists. of random parameters					
X16	.61738028	.03939213	15.673	.0000	
X27	.12578770	.01127763	11.154	.0000	
Threshold parameters for probabilities					
MU(1)	1.43463466	.04436853	32.335	.0000	
MU(2)	2.65483218	.05904585	44.962	.0000	
MU(3)	3.95509959	.07930608	49.871	.0000	

Implied standard deviations of random parameters

Matrix S.D_Beta has 2 rows and 1 columns.

	1
1	.61738
2	.12579

Summary of Marginal Effects for Ordered Probability Model (probit)								
Variable	Y=00	Y=01	Y=02	Y=03	Y=04	Y=05	Y=06	Y=07
X28	-.0023	-.0045	.0034	.0031	.0003			
X43	-.0070	-.0136	.0104	.0093	.0009			
*X16	.0281	.0523	-.0413	-.0356	-.0035			
X27	-.1082	-.2110	.1606	.1439	.0146			