



School of Civil Engineering

CE614 – Statistical and Econometric Methods

Assignment #7

Ordered Probit with Random Effects

A survey of 56 subjects was conducted on freeways in the Seattle area (see text pages 352-357). 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:

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.

Variables available for your specification are (in file pavement-pds.txt):

Variable Number	Explanation
1	Individual number
2	Roadway segment number
3	pds for Limdep random effects
4	Ruffness 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	Noise change from last section (dBA)

```
--> RESET
--> read;nvar=43;nobs=2240;file=D:\old_drive_d\new_laptop\CE697M\pavement-pds.txt
--> histogram;rhs=x4$
--> 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;margin$
```

Normal exit from iterations. Exit status=0.

```
+-----+
| Ordered Probability Model
| Maximum Likelihood Estimates
| Model estimated: Oct 16, 2013 at 09:49:00AM.
| Dependent variable          XX4
| Weighting variable          None
| Number of observations      2179
| Iterations completed        13
| Log likelihood function     -2634.961
| Number of parameters        8
| Info. Criterion: AIC =      2.42585
|   Finite Sample: AIC =      2.42588
| Info. Criterion: BIC =      2.44673
| Info. Criterion:HQIC =      2.43348
| Restricted log likelihood   -3187.274
| McFadden Pseudo R-squared  .1732870
| Chi squared                 1104.626
| Degrees of freedom          4
| Prob[ChiSqd > value] =     .0000000
| Underlying probabilities based on Normal
+-----+
```

```
+-----+
| Ordered Probability Model
| Cell frequencies for outcomes
| Y Count Freq  Y Count Freq  Y Count Freq
| 0   344 .157  1   769 .352  2   601 .275
| 3   351 .161  4   114 .052
+-----+
```

```
+-----+-----+-----+-----+-----+-----+
| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
+-----+-----+-----+-----+-----+-----+
|-----+ Index function for probability
| Constant | -.29190281 | .06465815 | -4.515 | .0000 |
| X16      | -.16405495 | .04715926 | -3.479 | .0005 | .40339605
| X27      | .72777644  | .04216693 | 17.259 | .0000 | 1.93683800
| X28      | .01589304  | .00244284 | 6.506  | .0000 | 18.1404314
| X43      | .04592131  | .00885239 | 5.187  | .0000 | .02707664
|-----+ Threshold parameters for index
| Mu(1)    | 1.31575527 | .03132014 | 42.010 | .0000 |
| Mu(2)    | 2.42432574 | .03621135 | 66.949 | .0000 |
| Mu(3)    | 3.57284187 | .05623083 | 63.539 | .0000 |
```

Normal exit from iterations. Exit status=0.

```

+-----+
| Random Effects Ordered Probability Model
| Maximum Likelihood Estimates
| Model estimated: Oct 16, 2013 at 09:49:04AM.
| Dependent variable           XX4
| Weighting variable           None
| Number of observations        2179
| Iterations completed         21
| Log likelihood function       -2373.267
| Number of parameters         9
| Info. Criterion: AIC =       2.18657
|   Finite Sample: AIC =       2.18661
| Info. Criterion: BIC =       2.21006
| Info. Criterion:HQIC =       2.19516
| Restricted log likelihood     -2634.961
| McFadden Pseudo R-squared    .0993160
| Chi squared                   523.3874
| Degrees of freedom            1
| Prob[ChiSqd > value] =       .0000000
| Underlying probabilities based on Normal
| Unbalanced panel has         56 individuals.
+-----+

```

Variable	Coefficient	Standard Error	b/St.Er.	P[Z >z]	Mean of X
-----+Index function for probability					
Constant	-.09913820	.15826398	-.626	.5310	
X16	-.42592044	.17398713	-2.448	.0144	.40339605
X27	.88019502	.04952193	17.774	.0000	1.93683800
X28	.01876465	.00302688	6.199	.0000	18.1404314
X43	.05625568	.01027184	5.477	.0000	.02707664
-----+Threshold parameters for index model					
Mu(01)	1.58053102	.04305620	36.709	.0000	
Mu(02)	2.90002859	.03738015	77.582	.0000	
Mu(03)	4.29680749	.05009421	85.775	.0000	
-----+Std. Deviation of random effect					
Sigma	.65637805	.07866815	8.344	.0000	

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
*X16	.0499	.0902	-.0518	-.0738	-.0146			
X27	-.0975	-.1921	.1009	.1569	.0319			
X28	-.0021	-.0041	.0022	.0033	.0007			
X43	-.0062	-.0123	.0065	.0100	.0020			

Cross tabulation of predictions. Row is actual, column is predicted.
 Model = Probit . Prediction is number of the most probable cell.

Actual	Row Sum	0	1	2	3	4	5	6	7	8	9
0	344	0	329	11	4	0					
1	769	0	628	102	38	1					
2	601	0	286	161	148	6					
3	351	0	52	118	165	16					
4	114	0	6	27	73	8					
Col Sum	2179	0	1301	419	428	31	0	0	0	0	0