Ordered Dependent Variable Models in EViews

Ass.Prof. Andriy Stavytskyy

Data

- Data were drawn in the article "Sleep in Mammals: Ecological and Constitutional Correlates" by Allison, T. and Cicchetti, D. (1976), _Science_, November 12, vol. 194, pp. 732-734. Includes brain and body weight, life span, gestation time, time sleeping, and predation and danger indices for 62 mammals.

- File: Order.wf1

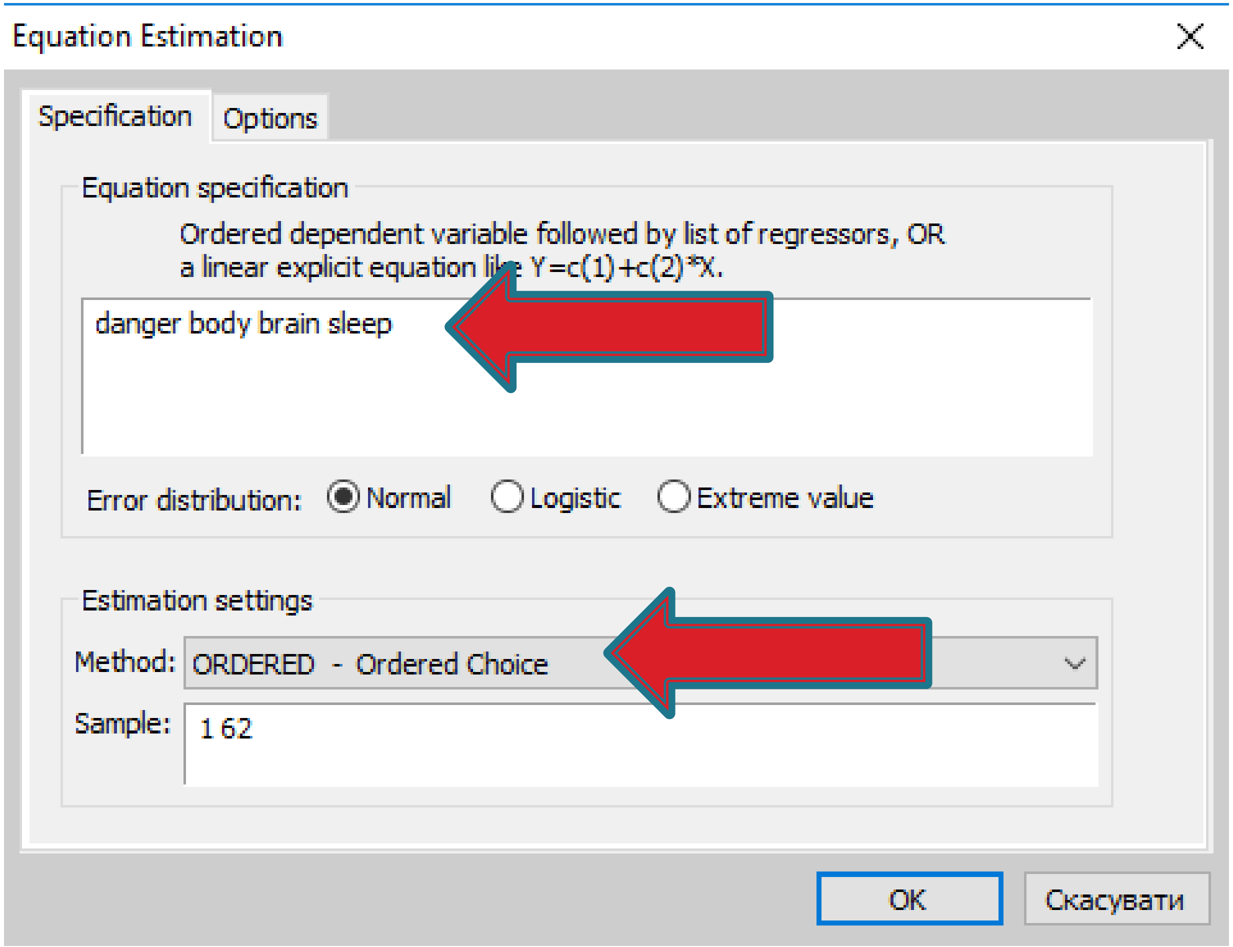
Variables

- species of animal
- body weight in kg
- brain weight in g
- slow wave ("nondreaming") sleep (hrs/day)
- paradoxical ("dreaming") sleep (hrs/day)
- total sleep (hrs/day) (sum of slow wave and paradoxical sleep)
- maximum life span (years)
- gestation time (days)
- predation index (1-5) 1 = minimum (least likely to be preyed upon) - 5 = maximum (most likely to be preyed upon)
- sleep exposure index (1-5) 1 = least exposed (e.g. animal sleeps in a well-protected den) - 5 = most exposed
- overall danger index (1-5) (based on the above two indices and other information) 1 = least danger (from other animals) - 5 = most danger (from other animals)
- Note: Missing values denoted by -999.0

View

African elephant	6654.000	5712.000	-999.0	-999.0	3.3	38.6	645.0	3	5	3
African giant pouched rat	1.000	6.600	6.3	2.0	8.3	4.5	42.0	3	1	3
Arctic Fox	3.385	44.500	-999.0	-999.0	12.5	14.0	60.0	1	1	1
Arctic ground squirrel	.920	5.700	-999.0	-999.0	16.5	-999.0	25.0	5	2	3
Asian elephant	2547.000	4603.000	2.1	1.8	3.9	69.0	624.0	3	5	4
Baboon	10.550	179.500	9.1	.7	9.8	27.0	180.0	4	4	4
Big brown bat	.023	.300	15.8	3.9	19.7	19.0	35.0	1	1	1
Brazilian tapir	160.000	169.000	5.2	1.0	6.2	30.4	392.0	4	5	4
Cat	3.300	25.600	10.9	3.6	14.5	28.0	63.0	1	2	1
Chimpanzee	52.160	440.000	8.3	1.4	9.7	50.0	230.0	1	1	1
Chinchilla	.425	6.400	11.0	1.5	12.5	7.0	112.0	5	4	4
Cow	465.000	423.000	3.2	.7	3.9	30.0	281.0	5	5	5
Desert hedgehog	.550	2.400	7.6	2.7	10.3	-999.0	-999.0	2	1	2
Donkey	187.100	419.000	-999.0	-999.0	3.1	40.0	365.0	5	5	5
Eastern American mole	.075	1.200	6.3	2.1	8.4	3.5	42.0	1	1	1
Echidna	3.000	25.000	8.6	.0	8.6	50.0	28.0	2	2	2
European hedgehog	.785	3.500	6.6	4.1	10.7	6.0	42.0	2	2	2
Galago	.200	5.000	9.5	1.2	10.7	10.4	120.0	2	2	2
Genet	1.410	17.500	4.8	1.3	6.1	34.0	-999.0	1	2	1
Giant armadillo	60.000	81.000	12.0	6.1	18.1	7.0	-999.0	1	1	1
Giraffe	529.000	680.000	-999.0	.3	-999.0	28.0	400.0	5	5	5
Goat	27.660	115.000	3.3	.5	3.8	20.0	148.0	5	5	5
Golden hamster	.120	1.000	11.0	3.4	14.4	3.9	16.0	3	1	2
Gorilla	207.000	406.000	-999.0	-999.0	12.0	39.3	252.0	1	4	1
Gray seal	85.000	325.000	4.7	1.5	6.2	41.0	310.0	1	3	1
Gray wolf	36.330	119.500	-999.0	-999.0	13.0	16.2	63.0	1	1	1
Ground squirrel	.101	4.000	10.4	3.4	13.8	9.0	28.0	5	1	3
Guinea pig	1.040	5.500	7.4	.8	8.2	7.6	68.0	5	3	4
Horse	521.000	655.000	2.1	.8	2.9	46.0	336.0	5	5	5
Jaguar	100.000	157.000	-999.0	-999.0	10.8	22.4	100.0	1	1	1
Kangaroo	35.000	56.000	-999.0	-999.0	-999.0	16.3	33.0	3	5	4
Lesser short-tailed shrew	.005	.140	7.7	1.4	9.1	2.6	21.5	5	2	4
Little brown bat	.010	.250	17.9	2.0	19.9	24.0	50.0	1	1	1

Model estimation



EViews will ignore any constant term in specification!!!

Note!

- EViews requires the dependent variable to be integer valued, otherwise you will see an error message, and estimation will stop.
- This is not, however, a serious restriction, since you can easily convert the series into an integer using @round, @floor or @ceil in an auto-series expression.

Outcome

Equation: EQ_ORDER Workfile: ORDER::Order\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: DANGER
Method: ML - Ordered Probit (Quadratic hill climbing / EViews legacy)
Date: 09/13/18 Time: 10:13
Sample (adjusted): 1 61
Included observations: 58 after adjustments
Number of ordered indicator values: 5
Convergence achieved after 5 iterations
Coefficient covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
BODY	0.000247	0.000421	0.587475	0.5569
BRAIN	-0.000397	0.000418	-0.950366	0.3419
SLEEP	-0.199508	0.041641	-4.791138	0.0000
Limit Points				
LIMIT_2:C(4)	-2.798449	0.514784	-5.436166	0.0000
LIMIT_3:C(5)	-2.038945	0.492198	-4.142527	0.0000
LIMIT_4:C(6)	-1.434567	0.473679	-3.028563	0.0025
LIMIT_5:C(7)	-0.601211	0.449109	-1.338675	0.1807

Pseudo R-squared	0.147588	Akaike info criterion	2.890028
Schwarz criterion	3.138702	Log likelihood	-76.81081
Hannan-Quinn criter.	2.986891	Restr. log likelihood	-90.10996
LR statistic	26.59830	Avg. log likelihood	-1.324324
Prob(LR statistic)	0.000007		

Coefficient β_j

- The sign of coefficient β_j shows the direction of the change in the probability of falling in the endpoint rankings ($y=0$ or $y=1$) when x_{ij} changes.
- $Pr(y=0)$ changes in the opposite direction of the sign of β_j and $Pr(y=1)$ changes in the same direction as the sign of β_j.

Limit Points

- Section presents the estimates of the coefficients and the associated standard errors and probability values:

Estimation Problems

- EViews currently has a limit of 750 total coefficients in an ordered dependent variable model. Thus, if you have 25 right-hand side variables, and a dependent variable with 726 distinct values, you will be unable to estimate your model using EViews.
- You may run into identification problems and estimation difficulties if you have some groups where there are very few observations. If necessary, you may choose to combine adjacent groups and re-estimate the model.

View-Dependent Variable Frequencies

- It computes a one-way frequency table for the ordered dependent variable for the observations in the estimation sample. EViews presents both the frequency table and the cumulative frequency table in levels and percentages.

Equation: EQ_ORDER Workfile: ORDER::Order\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable Frequencies
Equation: EQ_ORDER
Date: 09/13/18 Time: 10:27

			Cumulative	
Dep. Value	Count	Percent	Count	Percent
1	18	31.03	18	31.03
2	14	24.14	32	55.17
3	10	17.24	42	72.41
4	9	15.52	51	87.93
5	7	12.07	58	100.00

View-Prediction Evaluation

- It classifies observations on the basis of the predicted response. EViews performs the classification on the basis of the category with the maximum predicted probability.
- Each row represents a distinct value for the dependent variable. The “Obs” column indicates the number of observations with that value. Of those, the number of “Correct” observations are those for which the predicted probability of the response is the highest. Thus, 10 of the 18 individuals with a DANGER value of 1 were correctly specified. Overall, 43% of the observations were correctly specified for the fitted model versus 31% for the constant probability model.
- The bottom portion of the output shows additional statistics measuring this improvement.

Equation: EQ_ORDER   Workfile: ORDER::Order\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Prediction Evaluation for Ordered Specification
Equation: EQ_ORDER
Date: 09/13/18   Time: 10:27

Estimated Equation

Dep. Value	Obs.	Correct	Incorrect	% Correct	% Incorrect
1	18	10	8	55.556	44.444
2	14	6	8	42.857	57.143
3	10	0	10	0.000	100.000
4	9	3	6	33.333	66.667
5	7	6	1	85.714	14.286
Total	58	25	33	43.103	56.897

Constant Probability Spec.

Dep. Value	Obs.	Correct	Incorrect	% Correct	% Incorrect
1	18	18	0	100.000	0.000
2	14	0	14	0.000	100.000
3	10	0	10	0.000	100.000
4	9	0	9	0.000	100.000
5	7	0	7	0.000	100.000
Total	58	18	40	31.034	68.966

Gain over Constant Prob. Spec.

Dep. Value	Obs.	Equation % Incorrect	Constant % Incorrect	Total Gain*	Pct. Gain**
1	18	44.444	0.000	-44.444	NA
2	14	57.143	100.000	42.857	42.857
3	10	100.000	100.000	0.000	0.000
4	9	66.667	100.000	33.333	33.333
5	7	14.286	100.000	85.714	85.714
Total	58	56.897	68.966	12.069	17.500

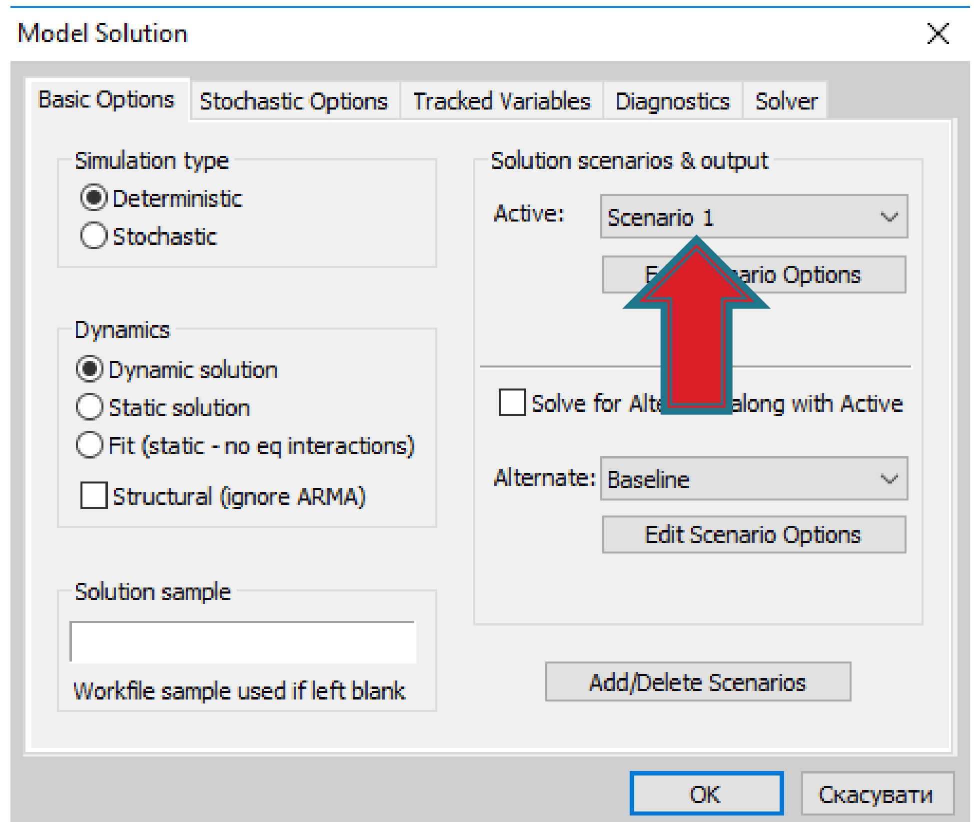
*Change in "% Correct" from default (constant probability) specification
**Percent of incorrect (default) prediction corrected by equation

Forecasting using Models – 1

- You cannot forecast directly from an estimated ordered model since the dependent variable represents categorical or rank data.
- EViews allows to forecast the probability associated with each category. To forecast these probabilities, you must first create a model. Choose **Proc/Make Model** and EViews will open an untitled model window containing a system of equations, with a separate equation for the probability of each ordered response value.

Forecasting using Models – 2

- To forecast from this model, simply click the Solve button in the model window toolbar. If you select Scenario 1 as your solution scenario, the default settings will save your results in a set of named series with "_1" appended to the end of the each underlying name.



Note!

- For this example, the series I_DANGER_1 will contain the fitted linear index $X^t_i\beta$. The fitted probability of falling in category 1 will be stored as a series named DANGER_1_1, the fitted probability of falling in category 2 will be stored as a series named DANGER_2_1, and so on.
- For each observation, the fitted probability of falling in each of the categories sums up to one.

Sum of columns

Group: UNTITLED Workfile: ORDER::Order\

View | Proc | Object | Print | Name | Freeze | Default | Sort | Edit+/- | Smpl+/- | Compare+/-

	DANGER_1_1	DANGER_2_1	DANGER_3_1	DANGER_4_1	DANGER_5_1
	DANGER_1_1	DANGER_2_1	DANGER_3_1	DANGER_4_1	DANGER_5_1
1	0.064752	0.159906	0.214865	0.312597	0.247880
2	0.127111	0.224620	0.236784	0.266248	0.145237
3	0.386767	0.294680	0.177618	0.112836	0.028100
4	0.689867	0.205389	0.073255	0.027945	0.003544
5	0.205657	0.269593	0.230944	0.209342	0.084464
6	0.219296	0.274692	0.228185	0.200411	0.077417
7	0.871178	0.099542	0.022998	0.005847	0.000435
8	0.062525	0.156818	0.213148	0.313947	0.253564
9	0.541326	0.264682	0.122894	0.060403	0.010696
10	0.241547	0.281646	0.222997	0.186471	0.067339
11	0.381266	0.295023	0.179531	0.115140	0.029041
12	0.024575	0.088989	0.159553	0.317815	0.409068
13	0.228833	0.277872	0.226056	0.194341	0.072898
14	0.019709	0.077042	0.146492	0.311419	0.445338

Make Residual Series

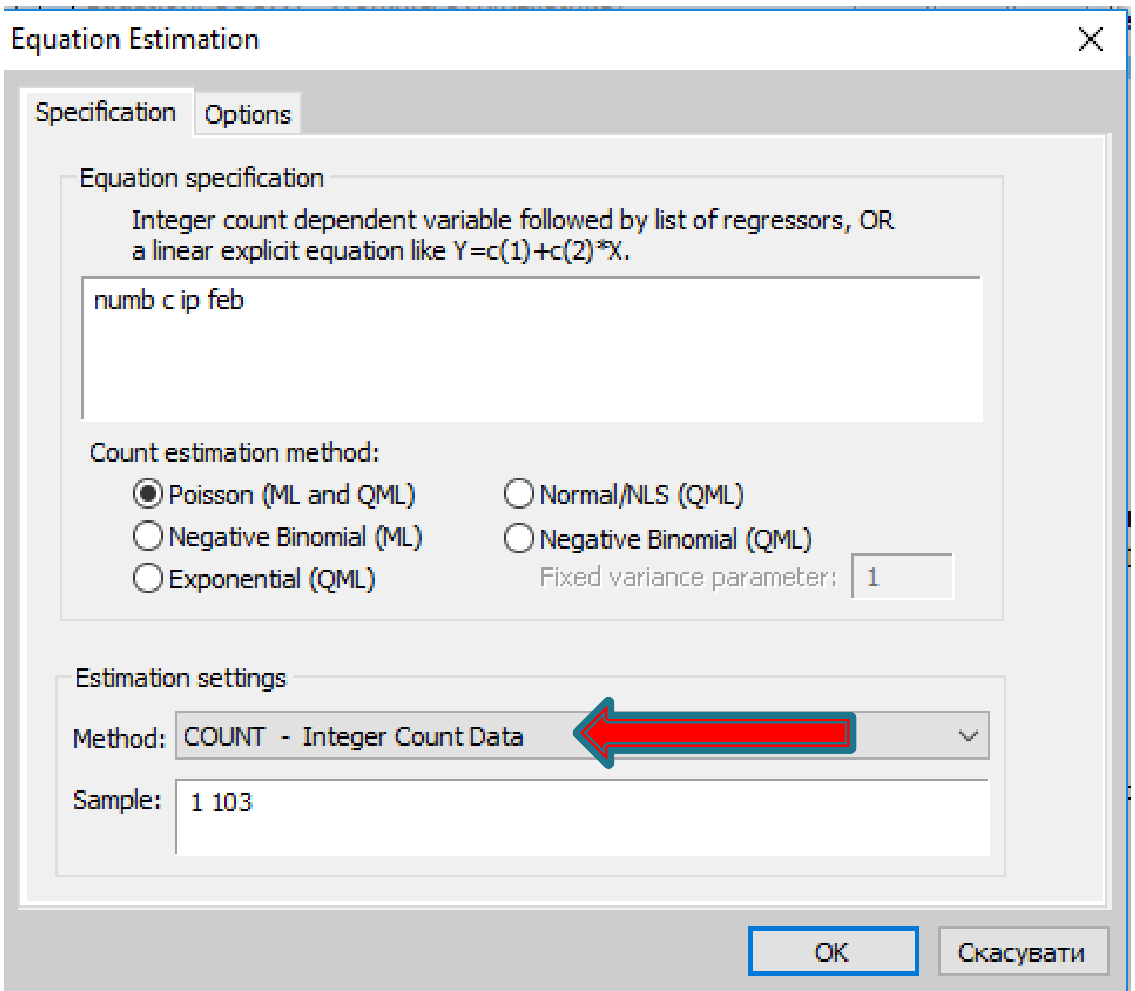
- To create a series containing the generalized residuals, select View/Make Residual Series…, enter a name or accept the default name, and click OK.

Count models

- File: Strike.WF1

- NUMB is the number of strikes,
- IP is a measure of industrial production,
- FEB is a February dummy variable

Quick/Estimate Equation…



Outcome

Equation: COUNT   Workfile: STRIKE::Strike\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: NUMB
Method: ML/QML - Poisson Count (Newton-Raphson / Marquardt steps)
Date: 09/13/18 Time: 18:47
Sample: 1 103
Included observations: 103
Convergence achieved after 3 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	1.725630	0.043656	39.52764	0.0000
IP	2.775334	0.819104	3.388254	0.0007
FEB	-0.377407	0.174520	-2.162539	0.0306

R-squared	0.064502	Mean dependent var	5.495146
Adjusted R-squared	0.045792	S.D. dependent var	3.653829
S.E. of regression	3.569190	Akaike info criterion	5.583421
Sum squared resid	1273.912	Schwarz criterion	5.660160
Log likelihood	-284.5462	Hannan-Quinn criter.	5.614503
Restr. log likelihood	-292.9694	LR statistic	16.84645
Avg. log likelihood	-2.762584	Prob(LR statistic)	0.000220

A Specification Test for Overdispersion – 1

- Estimate the Poisson model and obtain the fitted values of the dependent variable.
- Build an auxiliary regression

$$\varepsilon_i^2 - y_i = \beta y_i^2 + \nu_i$$

Command in Eviews:

```
equation testeq.ls (numb-numb_f)^2-numb numb_f^2
```

A Specification Test for Overdispersion – 2

Equation: TESTEQ Workfile: STRIKE::Strike\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: (NUMB-NUMB_F)^2-NUMB
Method: Least Squares
Date: 09/13/18 Time: 18:56
Sample: 1 103
Included observations: 103

Variable	Coefficient	Std. Error	t-Statistic	Prob.
NUMB_F^2	0.238874	0.052115	4.583571	0.0000

R-squared	0.043930	Mean dependent var	6.872929
Adjusted R-squared	0.043930	S.D. dependent var	17.65726
S.E. of regression	17.26506	Akaike info criterion	8.544908
Sum squared resid	30404.41	Schwarz criterion	8.570488
Log likelihood	-439.0628	Hannan-Quinn criter.	8.555269
Durbin-Watson stat	1.711805		

The t-statistic of the coefficient is highly significant, leading us to reject the Poisson restriction. Moreover, the estimated coefficient is significantly positive, indicating overdispersion in the residuals.

Questions?

Self study