

QS301 – Applying Quantitative Methods

Week 5: Models for ordinal outcomes II

Course structure

Term 2	Lecture	Seminar
Week 1	Recap of term 1 and introduction to data source.	Theory driven empirical research, and logistic/logit regression.
Week 2	Models for nominal outcomes I: Introducing the multinomial logit model, interpreting and reporting the results.	Multinomial logit model – from model to output.
Week 3	Models for nominal outcomes II: Testing theoretically relevant model assumptions and model modifications.	Multinomial logit model – testing theoretically relevant assumptions.
Week 4	Models for ordinal outcomes I: Introducing the ordinal logit model, interpreting and reporting the result.	Ordered logit model – from model to output.
Week 5	Models for ordinal outcomes II: Testing theoretically relevant model assumptions and model modifications.	Ordered logit model – testing theoretically relevant assumptions.
Week 6	Reading Week.	
Week 7	Introducing multilevel modelling.	Getting familiar with multilevel data structures.
Week 8	Building a multilevel model I: Comparing groups and random intercept models.	Multilevel model – Specifying an “empty” and random intercept model, interpreting and presenting results.
Week 9	Building a multilevel model II: Random slopes/coefficient models and contextual effects.	Multilevel model – Specifying a random slope model, interpreting ...
Week 10	Multilevel models for binary data and outlook.	

Outline

Task, task, task

Some useful tests and explanatory power

Another excursus on residuals and outliers

Parallel regression assumption

Comparing linear regression, ordered logit and ordered probit models

What to do (a suggestion only) ...

Outline

Task, task, task

Some useful tests and explanatory power

Another excursus on residuals and outliers

Parallel regression assumption

Comparing linear regression, ordered logit and ordered probit models

What to do (a suggestion only) ...

Task Week 5

Have a look at the variable V61 ("Men make better political leaders") and analyse it using an ordered logit model. Try to explain differences in response behaviour by including independent variables such as gender, age, etc., and use predicted probabilities, marginal effects, etc. to illustrate your results. What do you find?

tab V61

Men make better political leaders	Freq.	Percent	Cum.
1. Agree strongly	39	4.16	4.16
2. Agree	146	15.57	19.72
3. Disagree	598	63.75	83.48
4. Strongly disagree	155	16.52	100.00
Total	938	100.00	

gen menpol=5-V61

["Men make better political leaders", 1=strongly disagree; 4=strongly agree]

```
. ologit menpol woman age eduhigh income mixed postmat
```

[...]

```
Ordered logistic regression                Number of obs   =           705
                                           LR chi2(6)      =           52.19
                                           Prob > chi2     =           0.0000
Log likelihood = -702.30185                Pseudo R2      =           0.0358
```

menpol	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
woman	-.9258304	.1606507	-5.76	0.000	-1.2407	-.6109608
age	.0102414	.0046173	2.22	0.027	.0011916	.0192912
eduhigh	.0222321	.1611094	0.14	0.890	-.2935365	.3380006
income	-.0712237	.0312623	-2.28	0.023	-.1324966	-.0099508
mixed	-.2089897	.1309641	-1.60	0.111	-.4656745	.0476951
postmat	-.6817726	.2888381	-2.36	0.018	-1.247885	-.1156604
/cut1	-2.552836	.4131382			-3.362572	-1.7431
/cut2	.5492798	.398852			-.2324558	1.331015
/cut3	2.305913	.4259993			1.47097	3.140856

["Men make better political leaders ", 1=strongly disagree; 4=strongly agree]

```
. mtable, at(woman=(0(1)1) eduhigh=0 mixed=0 postmat=(0(1)1)) atmeans
```

Expression: Pr(menpol), predict(outcome())

	woman	postmat	1	2	3	4
1	0	0	0.071	0.558	0.279	0.093
2	0	1	0.131	0.639	0.181	0.049
3	1	0	0.161	0.649	0.151	0.039
4	1	1	0.275	0.619	0.086	0.020

Specified values of covariates

	age	eduhigh	income	mixed
Current	46.2	0	6.31	0

Outline

Task, task, task

Some useful tests and explanatory power

Another excursus on residuals and outliers

Parallel regression assumption

Comparing linear regression, ordered logit and ordered probit models

What to do (a suggestion only) ...

Some useful tests

(Long and Freese 2014)

$$H_0: \beta_{\text{woman}} = \beta_{\text{age}} = \beta_{\text{eduhigh}} = \beta_{\text{income}} = 0$$

LR test

```
. lrtest fullmodel dropsocio
```

Likelihood-ratio test

```
LR chi2(4) = 7.50
```

```
(Assumption: dropsocio nested in fullmodel)
```

```
Prob > chi2 = 0.1118
```

Wald test

```
. test woman age eduhigh income
```

```
( 1) [happiness]woman = 0
```

```
( 2) [happiness]age = 0
```

```
( 3) [happiness]eduhigh = 0
```

```
( 4) [happiness]income = 0
```

```
chi2( 4) = 7.42
```

```
Prob > chi2 = 0.1152
```

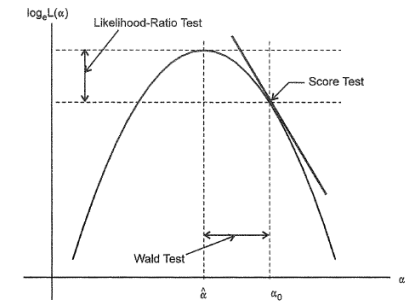


Figure D.16. Tests of the hypothesis $H_0: \alpha = \alpha_0$: The likelihood-ratio test compares $\log_e L(\hat{\alpha})$ with $\log_e L(\alpha_0)$; the Wald test compares $\hat{\alpha}$ with α_0 ; and the score test examines the slope of $\log_e L(\alpha)$ at $\alpha = \alpha_0$.

Interpretation (example): The hypothesis that all the effects of gender, age, education and income are simultaneously equal to zero cannot be rejected at the .10 level (LR $\chi^2 = 7.50$, $df = 4$, $p=0.112$).

... explanatory power (I)

```

ologit happiness woman age eduhigh income i.V55 i.V241 shealth if hmiss==0
fitstat, save
ologit happiness                i.V55 i.V241 shealth if hmiss==0

. fitstat, diff
  
```

	Current	Saved	Difference

Log-likelihood			
Model	-677.160	-673.411	-3.749
Intercept-only	-739.839	-739.839	0.000

Chi-square			
D (df=792/788/4)	1354.320	1346.821	7.499
LR (df=13/17/-4)	125.359	132.858	-7.499
p-value	0.000	0.000	0.112

R2			
McFadden	0.085	0.090	-0.005
McFadden (adjusted)	0.063	0.063	0.000
McKelvey & Zavoina	0.173	0.184	-0.010
Cox-Snell/ML	0.144	0.152	-0.008
Cragg-Uhler/Nagelkerke	0.171	0.181	-0.009
Count	0.608	0.606	0.001
Count (adjusted)	0.215	0.213	0.002

IC			
AIC	1386.320	1386.821	-0.501
AIC divided by N	1.716	1.716	-0.001
BIC (df=16/20/-4)	1461.433	1480.712	-19.280

Variance of			
e	3.290	3.290	0.000
y-star	3.980	4.030	-0.050

Quantities used in other measures

Log-likelihood based measures. Stata begins maximum likelihood iterations by computing the log likelihood of the model with all parameters but the intercept constrained to 0, referred to as $\ln L(M_{\text{Intercept}})$. The log likelihood upon convergence, referred to as $\ln L(M_{\text{Full}})$, is also listed. This information is presented in the iteration log and in the header for the estimation results.¹⁰

LR chi-square test of all coefficients. An LR test of the hypothesis that all coefficients except the intercepts are 0 can be computed by comparing the log likelihoods: $LR = 2 \ln L(M_{\text{Full}}) - 2 \ln L(M_{\text{Intercept}})$. LR is reported by Stata as $LR \text{ chi2}(df) = \#$, where the degrees of freedom in parentheses are the number of constrained parameters. For the zip and zinb models discussed in chapter 9, LR tests that the coefficients in the count portion (not the binary portion) of the model are 0.

Deviance. The deviance compares the given model with a model that has one parameter for each observation so that the model reproduces the observed data perfectly. The deviance is defined as $D = -2 \ln L(M_{\text{Full}})$, where the degrees of freedom equals N minus the number of parameters. D does not have a chi-squared distribution.

(Long and Freese 2014: 123)

Note: Likelihood-ratio test assumes current model nested in saved model.

... explanatory power (II)

```

ologit happiness woman age eduhigh income i.V55 i.V241 shealth if hmiss==0
fitstat, save
ologit happiness i.V55 i.V241 shealth if hmiss==0
. fitstat, diff

```

	Current	Saved	Difference

Log-likelihood			
Model	-677.160	-673.411	-3.749
Intercept-only	-739.839	-739.839	0.000

Chi-square			
D (df=792/788/4)	1354.320	1346.821	7.499
LR (df=13/17/-4)	125.359	132.858	-7.499
p-value	0.000	0.000	0.112

R2			
McFadden	0.085	0.090	-0.005
McFadden (adjusted)	0.063	0.063	0.000
McKelvey & Zavoina	0.173	0.184	-0.010
Cox-Snell/ML	0.144	0.152	-0.008
Cragg-Uhler/Nagelkerke	0.171	0.181	-0.009
Count	0.608	0.606	0.001
Count (adjusted)	0.215	0.213	0.002

IC			
AIC	1386.320	1386.821	-0.501
AIC divided by N	1.716	1.716	-0.001
BIC (df=16/20/-4)	1461.433	1480.712	-19.280

Variance of			
e	3.290	3.290	0.000
y-star	3.980	4.030	-0.050

Note: Likelihood-ratio test assumes current model nested in saved model.

McFadden's Pseudo R²

$$= 1 - LL_{Full} / LL_{Intercept}$$

$$= 1 - (-677.160 / -739.839)$$

McFadden Adj. Pseudo-R²

$$= 1 - (LL_{Full} - K / LL_{Intercept})$$

K is the number of parameters.

McKelvey & Zavoina Pseudo-R² is based on a latent variable model, most similar to linear regression.

Count R² is the proportion of correct predictions.

Adjusted Count R² is the proportion of correct predictions, adjusted for largest row marginal.

Outline

Task, task, task

Some useful tests and explanatory power

Another excursus on residuals and outliers

Parallel regression assumption

Comparing linear regression, ordered logit and ordered probit models

What to do (a suggestion only) ...

Outliers or “deviant cases” so far ...

From help `leastlikely` in Stata ...

For regression models for categorical dependent variables, `leastlikely` lists the in-sample observations with the lowest predicted probabilities of observing the outcome value that was actually observed. For example, in a model with a binary dependent variable, `leastlikely` lists the observations that have the lowest predicted probability of `depvar=0` among those cases for which `depvar=0`, and it lists the observations that have the lowest predicted probability of `depvar=1` among those cases for which `depvar=1`. The least likely values represent relatively deviant cases that may warrant closer inspection.

Residuals and outliers using predict (I)

- The fit of a regression model can also be assessed by looking at residuals and outliers.
- Residuals refer to the “difference between a model’s predicted and observed outcome for each observation in the sample.
- Cases that fit poorly (i.e., have large residuals) are known as *outliers*. When an observation has a large effect on the estimated parameters, it is said to be *influential*” (Long and Freese 2003: 123).
- It is recommended to use the standardized Pearson residual (i.e. constant variance).

Residuals and outliers using predict (II)

What it is all about ... linear regression example

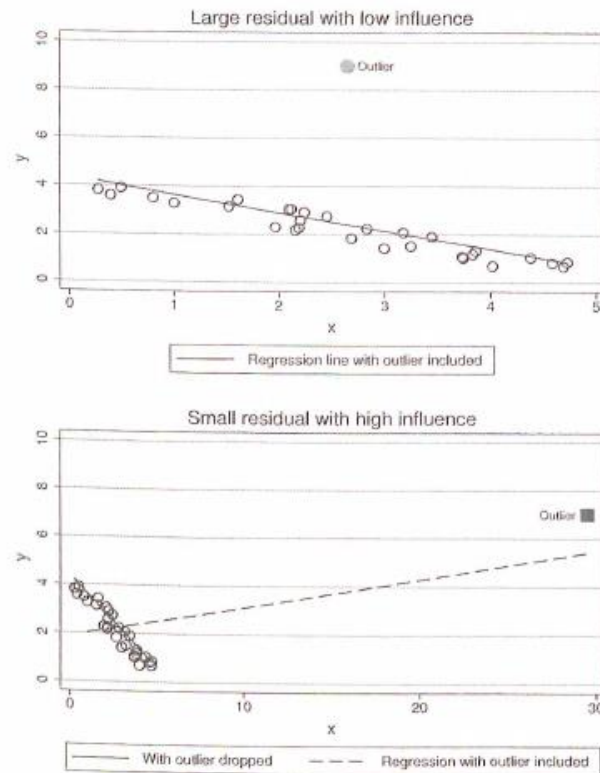


Figure 5.3. The distinction between an outlier and an influential observation

Residuals and outliers using predict (III)

What it is all about ... binary logit model example

Residuals

The predicted probability for a given set of independent variables is

$$\pi_i = \Pr(y_i = 1 \mid \mathbf{x}_i)$$

The deviations $y_i - \pi_i$ are heteroskedastic because the variance depends on the probability π_i of a positive outcome:

$$\text{Var}(y_i - \pi_i \mid \mathbf{x}_i) = \pi_i(1 - \pi_i)$$

The variance is greatest when $\pi_i = 0.5$ and decreases as π_i approaches 0 or 1. That is, a fair coin is the most unpredictable, with a variance of $0.5(1 - 0.5) = 0.25$. A coin that has a very small probability of landing head up (or tail up) has a small variance, for example, $0.01(1 - 0.01) = 0.0099$. The Pearson residual divides the residual $y - \hat{\pi}$ by its standard deviation:

$$r_i = \frac{y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}}$$

Large values of r suggest a failure of the model to fit a given observation.

Residuals and outliers using predict (IV)

What it is all about ... binary logit model example

Pregibon (1981) showed that the variance of r is not 1 because $\text{Var}(y_i - \hat{\pi}_i)$ is not exactly equal to the estimate $\hat{\pi}_i (1 - \hat{\pi}_i)$. He proposed the standardized Pearson residual

$$r_i^{\text{std}} = \frac{r_i}{\sqrt{1 - h_{ii}}}$$

where

$$h_{ii} = \hat{\pi}_i (1 - \hat{\pi}_i) \mathbf{x}_i' \widehat{\text{Var}}(\hat{\beta}) \mathbf{x}_i \quad (5.3)$$

Although r^{std} is preferred over r because of its constant variance, we find that the two residuals are often similar in practice. However, because r^{std} is simple to compute in Stata, you should use this measure.

Residuals and outliers using predict (V)

For ordered logit models, apply the method used for binary logit models with regard to the $J - 1$ categories.

- 1) Generate three binary variables for values <2 , <3 , and <4 of the happiness variable.
- 2) Run binary logit models for each binary variable and generate standardized residuals.
- 3) Sort the values of the variable that, we think, might be related to outliers, e.g. shealth.
- 4) Plotting the residuals for each observation and each of the three binary outcomes (i.e. three plots).

Outline

Task, task, task

Some useful tests and explanatory power

Another excursus on residuals and outliers

Parallel regression assumption

Comparing linear regression, ordered logit and ordered probit models

What to do (a suggestion only) ...

Parallel regression assumption or proportional-odds assumption (I)

(Long and Freese 2014: 326)

Ordered regression model with J outcome categories:

$$\Pr(y = 1 | x) = F(\tau_1 - x\beta)$$

$$\Pr(y = m | x) = F(\tau_m - x\beta) - F(\tau_{m-1} - x\beta) \text{ for } m = 2 \text{ to } J - 1$$

$$\Pr(y = J | x) = 1 - F(\tau_{J-1} - x\beta)$$

Cumulative probabilities have the form:

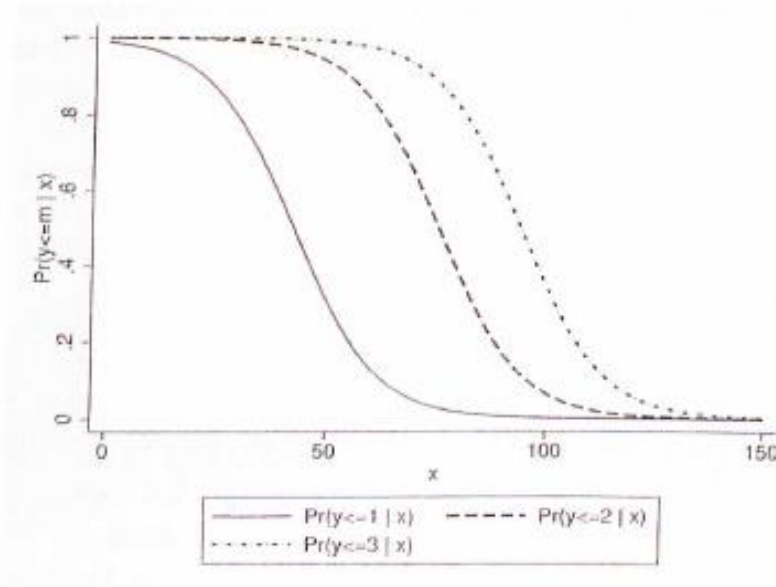
$$\Pr(y \leq m | x) = F(\tau_m - x\beta) \text{ for } m = 1 \text{ to } J - 1$$

“ β does not have a subscript m . The ordered logit model is equivalent to $J - 1$ binary regressions with the critical assumption that the slope of the coefficients are identical in each binary regression.”

Parallel regression assumption or proportional-odds assumption (II)

(Long and Freese 2014: 327)

Recall that the intercept α is not in the equation because it was assumed to equal 0 to identify the model. These equations lead to the following figure:



With 4 outcomes the cumulative probability equations are

$$\Pr(y \leq 1 | x) = F(\tau_1 - x\beta)$$

$$\Pr(y \leq 2 | x) = F(\tau_2 - x\beta)$$

$$\Pr(y \leq 3 | x) = F(\tau_3 - x\beta)$$

Each probability curve differs only in being shifted to the left or right. The curves are parallel as a consequence of the assumption that the β 's are equal for each equation.

Testing the assumption – basic logic

(Long and Freese 2014: 328)

Cumulative probabilities have the form:

$$\Pr(y \leq m|x) = F(\tau_m - x\beta) \text{ for } m = 1 \text{ to } J - 1$$

The parallel regression assumption can be tested by comparing $J - 1$ binary regression models where the β 's are allowed to differ between equations.

$$H_0: \beta_1 = \beta_2 = \dots = \beta_{J-1}$$

- *Wald test*: unconstrained model and test of restrictions in the null hypothesis.
- *LR test*: unconstrained and constrained model and comparison of log-likelihood (LL).
- *Score test*: constrained model and comparison how much the LL would change by relaxing the constraints.

Wald test, LR test, and Score test

(Long and Freese 2014; Fox 1997)

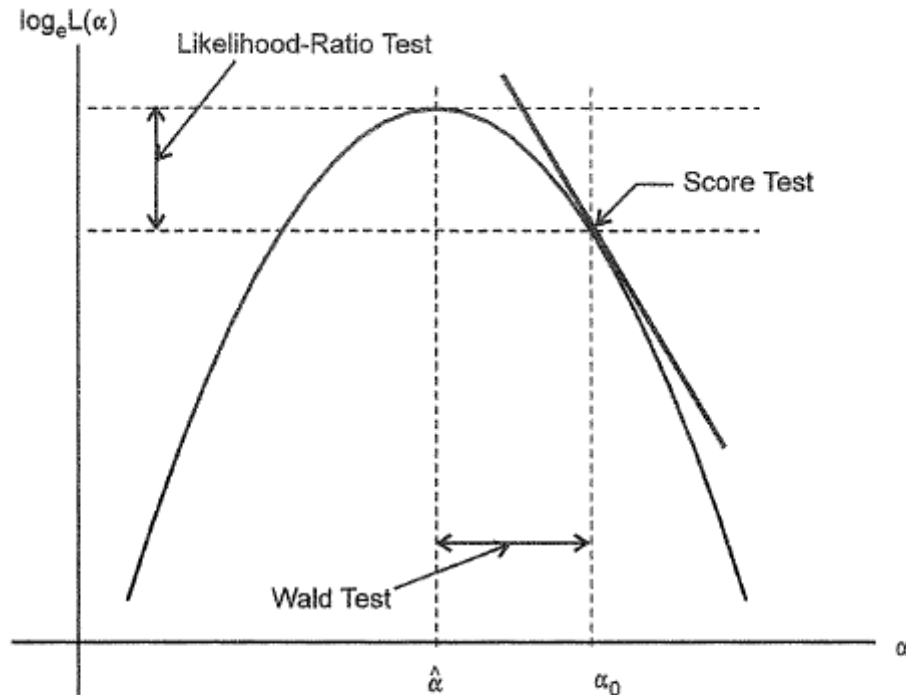
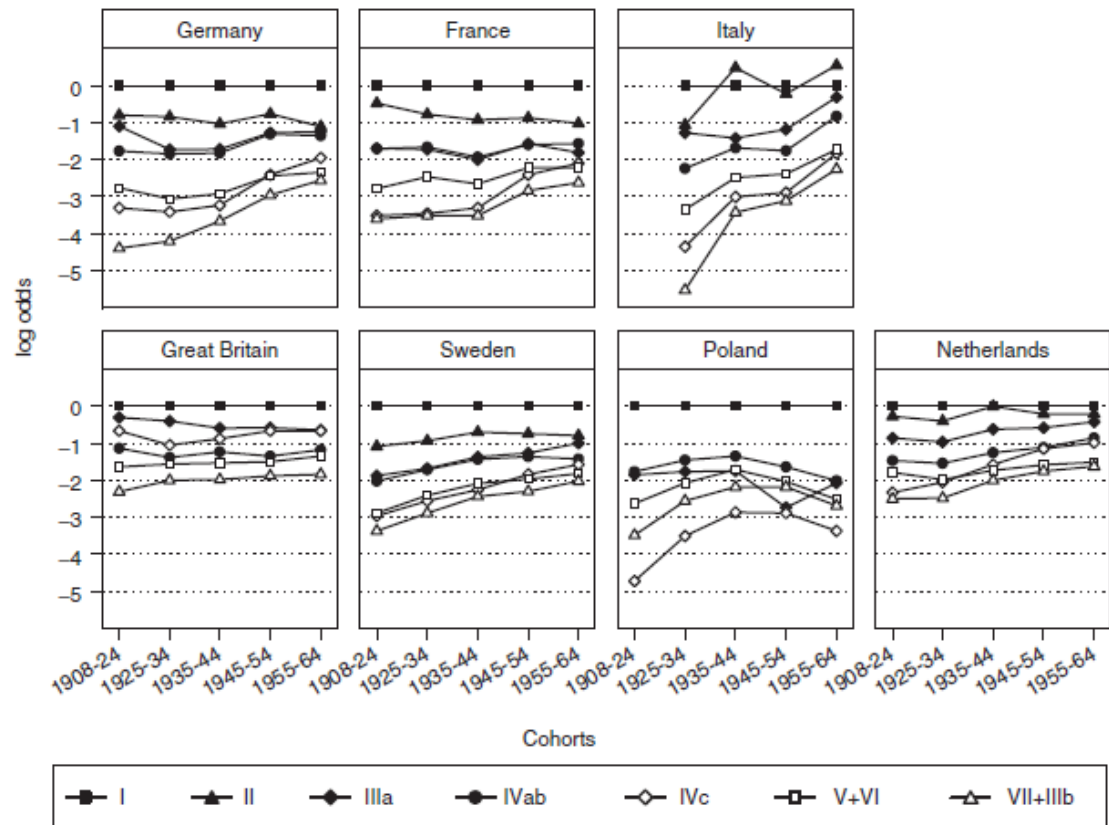


Figure D.16. Tests of the hypothesis $H_0: \alpha = \alpha_0$: The likelihood-ratio test compares $\log_e L(\hat{\alpha})$ with $\log_e L(\alpha_0)$; the Wald test compares $\hat{\alpha}$ with α_0 ; and the score test examines the slope of $\log_e L(\alpha)$ at $\alpha = \alpha_0$.

Example from the literature (I)



Class I is class I+II+IVa in Great Britain and class I+II in Poland; in Poland, classes IIIa and IIIb are combined; Educational categories 3a and 3b are merged in Italy and Poland, educational category 1c is coded into 2ab in Sweden

Figure 3 Ordered logit models for educational attainment in seven countries for women; class origin effects over cohorts, controlled for survey effects

“To examine this question we fitted models which relax the assumption of parallel slopes where this is necessary to achieve a good fit to the data. In these models threshold-specific parameters are estimated for those classes for which the odds between levels are not constant. Results are available from the authors on request. Most of the deviations are small in size and usually they are found in the older cohorts in ways that strengthen rather than weaken the conclusion of inequalities declining over time.” (Breen et al. 2010: 38)

Example from the literature (IIa)

The odds that the score is less than or equal to j versus greater than j is the ratio of the probabilities, $[\Pr(y \leq j)/\Pr(y > j)]$. In the logit model it holds, in particular, that the logarithm of the odds can be written in a simple form $\ln[\Pr(y \leq j)/\Pr(y > j)] = \beta_j - x\beta$. This model assumes that the coefficients β are the same for all categories. Therefore, the model rests on the so-called parallel regression assumption. For example, being unemployed has the same effect on the odds of having happiness level 1 vs. level 2, and on the odds of having happiness level 3 vs. level 4.

In the generalized ordered logit estimates unemployment has a significant negative effect in Equation 1 that determines the choice between the combined categories 1 and 2 and category 3 (Table I). However, it is insignificant in Equation 2, which determines the choice between categories 3 and 4. In contrast, low income has an insignificant effect in the first equation, whereas it is negative and significant in the second one. Interestingly, it seems that when

(Böckerman and Ilmakunas 2006: 163)

Example from the literature (IIb)

TABLE I

Ordered logit results for happiness and life satisfaction

	Happiness, ordered logit (4 categories)	Life satisfaction, ordered logit (10 categories)	Happiness, generalized ordered logit (3 categories)	
			Equation 1	Equation 2
Unemployed	-0.165 (1.18)	-0.371 (3.86)***	-0.807 (2.15)**	0.183 (0.71)
Low_Income	-0.475 (5.76)***	-0.531 (6.93)***	-0.317 (0.63)	-0.502 (7.29)***
Year = 1996	0.498 (8.66)***	0.216 (4.39)***	0.959 (6.81)***	0.335 (7.24)***
Year = 2000	0.440 (6.45)***	0.352 (5.18)***	0.564 (4.79)***	0.412 (6.80)***
Observations	2172	2172	2172	
Pseudo R^2	0.056	0.029	0.075	

Notes: All models include the unreported individual-level control variables as described in Appendix (A). Absolute values of robust z -statistics in parentheses (adjusted for clustering by year). * significant at 10%; ** significant at 5%; *** significant at 1%.

(Böckerman and Ilmakunas 206: 163)

Example from the literature (IIIa)

Government Checking Government: How Performance Measures Expand Distributive Politics

Anthony M. Bertelli University of Southern California and University of Manchester
Peter John University of Manchester

This paper argues that distributive politics operates in a variety of contexts in which governments seek to check the behavior of other governments. We provide a novel theoretical account of performance measurement systems as political discipline mechanisms even when measures are compiled by formally independent administrative agencies. We test the implications of our theory using a dataset of performance ratings in English local government assessed between 2002 and 2006. Results suggest that political influence favors swing voters, and local authorities sharing party affiliation with the incumbent central government are favored over those controlled by the opposition. Evidence further suggests that the independent rater in our empirical case is influenced through ties between its membership and the local authorities that it regulates. Our theoretical argument and findings have implications for many national and international contexts.

Example from the literature (IIa)

Dependent Variable

These dyads are repeatedly observed over the years in which the CPA was administered, 2002–2006. Our dependent variable is *CPA score*, the publicized rating, which takes values between one and four. The dataset includes 3,230 dyad-year observations.⁶ In 2006, the scores available were changed to a “star system” with a scale ranging from zero to four, but the basic five categories remained intact. To retain information for the entire period, we combined categories 1 and 2 at the low end of each scale into a single category (labeled 1) and the CPA’s 2002–2005 categories 3–5 were recoded to 2–4, producing the four-category scale analyzed as our dependent variable.⁷

Statistical Model and Specification

We employ a stereotype logistic regression (SLR) model (Anderson 1984)¹⁰ to estimate the conditional probability that the Audit Commission assigns a CPA rating at each of the four levels. This estimator is particularly suited to the rating problem that faces the Audit Commission when assigning CPA scores (see Audit Commission 2002, 12), where the audit commission “processes an indeterminate amount of information before providing [its] *judgement* of the grade,” our CPA score (Anderson 1984, 2, emphasis original). In assess-

¹¹Using the more parsimonious ordered logit model, we performed a likelihood-ratio test of the null hypothesis that the coefficients of the regressors are equal across categories of the dependent variable that produced a χ^2 (35 df) = 309.67, suggesting that the proportional odds assumption is inappropriate in our case.

Our example – oparallel command

```
. quietly ologit happiness woman age eduhigh income shealth  
. oparallel, ic
```

Tests of the parallel regression assumption

	Chi2	df	P>Chi2
Wolfe Gould	32.29	10	0.000
Brant	25.46	10	0.005
score	31.55	10	0.000
likelihood ratio	31.6	10	0.000
Wald	29.98	10	0.001

Information criteria

	ologit	gologit	difference
AIC	1410.08	1398.48	11.60
BIC	1447.66	1483.05	-35.39

Our example – brant command

```
. brant, detail
```

Estimated coefficients from binary logits

Variable	y_gt_1	y_gt_2	y_gt_3
woman	0.110 0.17	0.080 0.26	0.179 1.20
age	-0.031 -1.51	-0.011 -1.28	0.014 3.20
eduhigh	-0.305 -0.41	-0.203 -0.58	-0.096 -0.61
income	0.065 0.46	0.188 2.84	0.039 1.26
shealth	1.587 3.72	0.881 5.29	0.646 7.22
_cons	2.059 1.26	-0.028 -0.04	-2.886 -7.00

Brant test of parallel regression assumption

	chi2	p>chi2	df
All	25.46	0.005	10
woman	0.10	0.951	2
age	9.66	0.008	2
eduhigh	0.12	0.941	2
income	5.60	0.061	2
shealth	5.15	0.076	2

A significant test statistic provides evidence that the parallel regression assumption has been violated.

legend: b/t

Our example – gologit2 command

```
. gologit2 happiness woman age eduhigh income shealth, autofit
```

```
-----  
Testing parallel lines assumption using the .05 level of significance...
```

```
Step 1: Constraints for parallel lines imposed for woman (P Value = 0.9096)  
Step 2: Constraints for parallel lines imposed for eduhigh (P Value = 0.8960)  
Step 3: Constraints for parallel lines imposed for shealth (P Value = 0.0575)  
Step 4: Constraints for parallel lines are not imposed for  
       age (P Value = 0.00085)  
       income (P Value = 0.01856)
```

```
Wald test of parallel lines assumption for the final model:
```

```
( 1) [Not_at_all_happy]woman - [Not_very_happy]woman = 0  
( 2) [Not_at_all_happy]eduhigh - [Not_very_happy]eduhigh = 0  
( 3) [Not_at_all_happy]shealth - [Not_very_happy]shealth = 0  
( 4) [Not_at_all_happy]woman - [Quite_happy]woman = 0  
( 5) [Not_at_all_happy]eduhigh - [Quite_happy]eduhigh = 0  
( 6) [Not_at_all_happy]shealth - [Quite_happy]shealth = 0
```

```
       chi2( 6) =      6.12  
       Prob > chi2 =    0.4103
```

An insignificant test statistic indicates that the final model does not violate the proportional odds/ parallel lines assumption

. If you re-estimate this exact same model with gologit2, instead of autofit you can save time by using the parameter

pl(woman eduhigh shealth)

```
-----
Generalized Ordered Logit Estimates          Number of obs   =      811
                                             Wald chi2(9)    =     104.23
                                             Prob > chi2     =     0.0000
Log likelihood = -684.70997                 Pseudo R2       =     0.0774
-----
```

- (1) [Not_at_all_happy]woman - [Not_very_happy]woman = 0
- (2) [Not_at_all_happy]eduhigh - [Not_very_happy]eduhigh = 0
- (3) [Not_at_all_happy]shealth - [Not_very_happy]shealth = 0
- (4) [Not_very_happy]woman - [Quite_happy]woman = 0
- (5) [Not_very_happy]eduhigh - [Quite_happy]eduhigh = 0
- (6) [Not_very_happy]shealth - [Quite_happy]shealth = 0

happiness	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

Not_at_all_happy						
woman	.1776596	.1436641	1.24	0.216	-.1039168	.459236
age	-.0342668	.0171908	-1.99	0.046	-.0679601	-.0005735
eduhigh	-.1239384	.1533946	-0.81	0.419	-.4245862	.1767094
income	.1351314	.1254624	1.08	0.281	-.1107705	.3810332
shealth	.704493	.0862506	8.17	0.000	.535445	.8735411
_cons	3.429006	1.360198	2.52	0.012	.7630657	6.094946

Not_very_happy						
woman	.1776596	.1436641	1.24	0.216	-.1039168	.459236
age	-.0146565	.0083936	-1.75	0.081	-.0311076	.0017945
eduhigh	-.1239384	.1533946	-0.81	0.419	-.4245862	.1767094
income	.2091992	.0630283	3.32	0.001	.0856659	.3327324
shealth	.704493	.0862506	8.17	0.000	.535445	.8735411
_cons	.3357122	.6073673	0.55	0.580	-.8547059	1.52613

Quite_happy						
woman	.1776596	.1436641	1.24	0.216	-.1039168	.459236
age	.0146629	.0044577	3.29	0.001	.005926	.0233998
eduhigh	-.1239384	.1533946	-0.81	0.419	-.4245862	.1767094
income	.0347764	.0313145	1.11	0.267	-.0265989	.0961517
shealth	.704493	.0862506	8.17	0.000	.535445	.8735411
_cons	-3.04774	.4042524	-7.54	0.000	-3.84006	-2.25542

. If you re-estimate this exact same model with gologit2, instead of autofit you can save time by using the parameter

pl(woman eduhigh shealth)

```
-----
Generalized Ordered Logit Estimates          Number of obs   =      811
                                             Wald chi2(9)    =     104.23
                                             Prob > chi2     =     0.0000
Log likelihood = -684.70997                 Pseudo R2       =     0.0774
-----
```

- (1) [Not_at_all_happy]woman - [Not_very_happy]woman = 0
- (2) [Not_at_all_happy]eduhigh - [Not_very_happy]eduhigh = 0
- (3) [Not_at_all_happy]shealth - [Not_very_happy]shealth = 0
- (4) [Not_very_happy]woman - [Quite_happy]woman = 0
- (5) [Not_very_happy]eduhigh - [Quite_happy]eduhigh = 0
- (6) [Not_very_happy]shealth - [Quite_happy]shealth = 0

happiness	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

Not_at_all_happy						
woman	.1776596	.1436641	1.24	0.216	-.1039168	.459236
age	-.0342668	.0171908	-1.99	0.046	-.0679601	-.0005735
eduhigh	-.1239384	.1533946	-0.81	0.419	-.4245862	.1767094
income	.1351314	.1254624	1.08	0.281	-.1107705	.3810332
shealth	.704493	.0862506	8.17	0.000	.535445	.8735411
_cons	3.429006	1.360198	2.52	0.012	.7630657	6.094946

Not_very_happy						
woman	.1776596	.1436641	1.24	0.216	-.1039168	.459236
age	-.0146565	.0083936	-1.75	0.081	-.0311076	.0017945
eduhigh	-.1239384	.1533946	-0.81	0.419	-.4245862	.1767094
income	.2091992	.0630283	3.32	0.001	.0856659	.3327324
shealth	.704493	.0862506	8.17	0.000	.535445	.8735411
_cons	.3357122	.6073673	0.55	0.580	-.8547059	1.52613

Quite_happy						
woman	.1776596	.1436641	1.24	0.216	-.1039168	.459236
age	.0146629	.0044577	3.29	0.001	.005926	.0233998
eduhigh	-.1239384	.1533946	-0.81	0.419	-.4245862	.1767094
income	.0347764	.0313145	1.11	0.267	-.0265989	.0961517
shealth	.704493	.0862506	8.17	0.000	.535445	.8735411
_cons	-3.04774	.4042524	-7.54	0.000	-3.84006	-2.25542

Excursus – perfect predictions

```
. *** Perfect predictions ***
```

```
ologit happiness woman age eduhigh income shealth i.V55  
oparallel, ic
```

```
tab happiness V55
```

```
recode happiness (1 2 = 1 "Not at all / not very happy") (3 = 2 "Quite happy") (4 = 3 "Very  
happy"), pre(happiness3)
```

```
recode V55 (1 2 = 1 "As married") (3 = 2 "Divorced") (4 = 3 "Separated") (5 = 4 "Widowed")  
(6 = 5 "Single"), pre(V55NEW)
```

```
ologit happiness3 woman age eduhigh income shealth i.V55NEW, base  
oparallel, ic
```

```
. tab happiness V55
```

happiness	Marital status						Total
	1. Marrie	2. Living	3. Divorc	4. Separa	5. Widowe	6. Single	
Not at all happy	5	0	3	1	3	2	14
Not very happy	21	0	7	4	12	11	55
Quite happy	173	42	46	14	35	132	442
Very happy	297	50	18	10	43	106	524
Total	496	92	74	29	93	251	1,035

Last words on the assumption

(Long and Freese 2014: 331)

7.6.3 Caveat regarding the parallel regression assumption

In the majority of the real-world applications of the ORM that we have seen, the hypothesis of parallel regressions is rejected. Keep in mind, however, the tests of the parallel regression assumption are sensitive to other types of misspecification. Further, we have seen examples where the parallel regression assumption is violated but the predictions from `ologit` are very similar to those from the generalized ordered logit model or the multinomial logit model. When the hypothesis is rejected, consider alternative models that do not impose the constraint of parallel regressions. As illustrated in chapter 8, fitting the multinomial logit model on a seemingly ordinal outcome can lead to quite different conclusions. The generalized ordered logit model is another alternative to consider. Violation of the parallel regression assumption is not, however, a rationale for using the LRM. The assumptions implied by the application of the LRM to ordinal data are even stronger.

Outline

Task, task, task

Some useful tests and explanatory power

Another excursus on residuals and outliers

Parallel regression assumption

Comparing linear regression, ordered logit and ordered probit models

What to do (a suggestion only) ...

```
. estout ols ologit oprobit, cells(b(star fmt(3)) ///
> t(par fmt(2))) stats(r2 r2_p aic N) style(fixed) label
```

Happiness	ols b/t	ologit b/t	oprobit b/t
main			
woman	0.045 (1.02)	0.168 (1.17)	0.088 (1.04)
age	0.002 (1.65)	0.010* (2.36)	0.005 (1.93)
eduhigh	-0.033 (-0.71)	-0.105 (-0.69)	-0.058 (-0.65)
income	0.019* (2.06)	0.052 (1.71)	0.034 (1.93)
shealth	0.230*** (9.23)	0.714*** (8.29)	0.423*** (8.66)
_cons	2.512*** (21.99)		
cut1			
_cons		-1.566*** (-3.40)	-0.711** (-2.97)
cut2			
_cons		0.067 (0.18)	0.056 (0.25)
cut3			
_cons		2.979*** (7.63)	1.721*** (7.65)
r2	0.119		
r2_p		0.061	0.065
aic	1509.963	1403.470	1398.225
N	807.000	807.000	807.000

Comparing linear regression, ordered logit, and ordered probit models

Outline

Task, task, task

Some useful tests and explanatory power

Another excursus on residuals and outliers

Parallel regression assumption

Comparing linear regression, ordered logit and ordered probit models

A story about happiness revisited

A story about happiness revisited

Generalized Ordered Logit Estimates Number of obs = 811
 Wald chi2(9) = 104.23
 Prob > chi2 = 0.0000
 Log likelihood = -684.70997 Pseudo R2 = 0.0774

- (1) [Not_at_all_happy]woman - [Not_very_happy]woman = 0
- (2) [Not_at_all_happy]eduhigh - [Not_very_happy]eduhigh = 0
- (3) [Not_at_all_happy]shealth - [Not_very_happy]shealth = 0
- (4) [Not_very_happy]woman - [Quite_happy]woman = 0
- (5) [Not_very_happy]eduhigh - [Quite_happy]eduhigh = 0
- (6) [Not_very_happy]shealth - [Quite_happy]shealth = 0

happiness	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

Not_at_all_happy						
woman	.1776596	.1436641	1.24	0.216	-.1039168	.459236
age	-.0342668	.0171908	-1.99	0.046	-.0679601	-.0005735
eduhigh	-.1239384	.1533946	-0.81	0.419	-.4245862	.1767094
income	.1351314	.1254624	1.08	0.281	-.1107705	.3810332
shealth	.704493	.0862506	8.17	0.000	.535445	.8735411
_cons	3.429006	1.360198	2.52	0.012	.7630657	6.094946

Not_very_happy						
woman	.1776596	.1436641	1.24	0.216	-.1039168	.459236
age	-.0146565	.0083936	-1.75	0.081	-.0311076	.0017945
eduhigh	-.1239384	.1533946	-0.81	0.419	-.4245862	.1767094
income	.2091992	.0630283	3.32	0.001	.0856659	.3327324
shealth	.704493	.0862506	8.17	0.000	.535445	.8735411
_cons	.3357122	.6073673	0.55	0.580	-.8547059	1.52613

Quite_happy						
woman	.1776596	.1436641	1.24	0.216	-.1039168	.459236
age	.0146629	.0044577	3.29	0.001	.005926	.0233998
eduhigh	-.1239384	.1533946	-0.81	0.419	-.4245862	.1767094
income	.0347764	.0313145	1.11	0.267	-.0265989	.0961517
shealth	.704493	.0862506	8.17	0.000	.535445	.8735411
_cons	-3.04774	.4042524	-7.54	0.000	-3.84006	-2.25542

A generalized ordered logit model shows different age and income effects across happiness categories.

Outline

Task, task, task

Some useful tests and explanatory power

Another excursus on residuals and outliers

Parallel regression assumption

Comparing linear regression, ordered logit and ordered probit models

What to do (a suggestion only) ...

What to do (a suggestion only) ...

For ordered logit models ...

- 1) Test for outliers and influential cases, especially when working with small samples.
- 2) Test the parallel regression assumption. If it is violated estimate a generalized ordered logit model (`gologit2`) and evaluate how severe the violation of the assumption is.
- 3) If the parallel regression assumption is problematic present the results of the generalized ordered logit model or a multinomial logit model.
- 4) Report the log odds (logits) and at least one other estimate such as predicted probabilities, average marginal effects, etc.
- 5) Report the log-likelihood of the constant only model, the full model, a Pseudo R^2 value, AIC or BIC, and the number of respondents.

Thank you!

References

Bertelli, AM and Peter, J (2010). Government checking government: How performance measures expand distributive politics. *Journal of Politics*, 72: 545-558.

Böckerman, P and Ilmakunnas, P (2006). Elusive Effects of Unemployment on Happiness. *Social Indicators Research*, 79: 1, 159-169.

Breen, R, Luijkx, R, Müller, W and Pollak, R (2010). Long-term Trends in Educational Inequality in Europe: Class Inequalities and Gender Differences. *European Sociological Review*, 26: 1, 31-48.

Fox, J (1997). *Applied Regression Analysis, Linear Models, and Related Models*. Thousand Oaks: Sage.

Long, JS, and J Freese (2003). *Regression Models for Categorical Dependent Variables Using Stata*. Revised Edition. College Station, Texas: Stata Press.

Long, JS, and J Freese (2014). *Regression Models for Categorical Dependent Variables Using Stata*. 3rd Edition. College Station, Texas: Stata Press.