Logistic Regression: Binomial, Multinomial and Ordinal[1]

Håvard Hegre

23 September 2011

Chapter 3

Multinomial Logistic Regression

Tables 1.1 and 1.2 showed how the probability of voting SV or Ap depends on whether respondents classify themselves as supporters or opponents of the current tax levels on high incomes. We calculated odds ratios in each of these 2x2 tables to obtain a measure of the degree to which the tax variable affects voting. However, the table showing the odds of voting Ap is particularly problematic because respondents that said they would vote SV are placed in the reference category. If in fact those who wish to maintain the taxes on high incomes have higher odds of voting Ap and are even more likely to vote SV, then the odds ratio from table 1.2 will underestimate the impact of the tax variable on voting Ap. In this instance, positive attitudes towards taxes increase the probability of being in the reference category to some extent. To avoid this source of bias we have to look separately at the three possible voting choices. Table 3.1 reports voting for SV, Ap, and the Bourgeois parties, grouped by respondents' attitudes towards taxes. We will denote these voting choices S, A, and B.

---

[1] This note is a translation of chapters 3 and 4 in Hegre, Håvard. 2011. *Logistisk regresjon: binomisk, multinomisk og rangert.*

Table 3.1: Voting for Ap vs. SV vs. other parties as a function of attitudes towards current taxes

Attitude towards current ta xes	Other		Ap		SV		Total	
	No.	%	No.	%	No.	%	No.	%
Reduce taxes	207	67.9	65	21.3	33	10.8	305	100.0
Maintain taxes	158	54.3	81	27.8	52	17.9	291	100.0
Total	365	61.2	146	24.5	85	14.3	596	100.0

Source: Valgundersøkelsen 2001 (Aardal et al., 2003)

3.1 Odds Ratios for More Than Two Categories

The percentages of those who say they will vote SV and Ap are the same in table 3.1 as in the two earlier tables, but the percentage of those who say they will vote 'other parties' is different. We can nonetheless calculate many different odds and odds ratios here. It is possible to calculate the odds of voting Ap (A) vs. Bourgeois (B), SV (S) vs. B, S vs. A, S vs. A+B, and so on. We have to decide on which of these we are most interested in.

3.1.1 Reference-Outcome Odds

One possible way in which to decide on odds is to set one of the outcomes as a reference outcome (see definition 1) and calculate the odds of the remaining outcomes relative to the reference outcome. We may call it *reference-outcome odds*, and *reference-outcome logit* in the logarithmic form.[2]

Definition 8. Reference-outcome odds are the odds of one outcome Y_J relative to a reference outcome Y_0:

$$o_{J0} = \frac{p(Y = J)}{p(Y = 0)}$$

In this case, we set B as the reference outcome. We see in the row labeled 'total' in the table that 24.5% voted Ap and 61.2% voted Other. The reference outcome odds for Ap is therefore

---

[2] Agresti (2007) calls this type of logits 'baseline category logits'.

$$o_{Ap} = 0.245/0.612 = 0.40$$

The equivalent odds for SV is

$$o_{SV} = 0.143/0.612 = 0.23$$

B is a good choice as reference outcome for several reasons. First, B is the outcome with most observations. This might come to use later. More importantly, in this case we are interested in what makes voters choose parties to the left as opposed to other parties. It is therefore desirable to know that a variable increases the odds of voting Ap as opposed to Bourgeois. We can get results in this form when B is the reference outcome. This is illustrated in figure 3.1.

Note that the definition of odds from chapters 1 and 2 is just a special case of this definition. When the dependent variable has only two categories, only one reference-outcome odds is possible and we simply call it 'odds'.
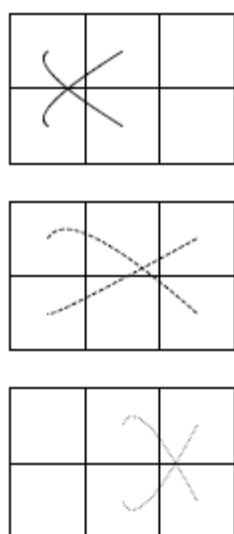

Figure 3.1: Graphical depiction of the odds ratios in a 2x3 table

3.1.2 Reference-Outcome Odds Ratio

We may calculate odds ratios using the reference-outcome odds as we did for odds in chapter 1. Odds ratio of voting Ap vs. Bourgeois for the different attitudes towards taxes is

$$O_{AB} = \frac{81 \cdot 207}{65 \cdot 158} = 1.63$$

Odds ratio of voting SV vs. Bourgeois is

$$O_{SB} = \frac{52 \cdot 207}{33 \cdot 158} = 2.06$$

We may also be interested in knowing the odds ratio of voting SV vs. Ap:

$$O_{SA} = \frac{52 \cdot 65}{33 \cdot 81} = 1.26$$

Notice that we can find the odds ratio O_{SA} if we know the other two odds ratios:

$$\frac{2.06}{1.63} = 1.26$$

This holds for any frequency we might observe. We can see this by dividing one formula by the other and simplifying:

$$\frac{O_{SB}}{O_{AB}} = \frac{52 \cdot 207}{33 \cdot 158} \Big/ \frac{81 \cdot 207}{65 \cdot 158} = \frac{52 \cdot 207}{33 \cdot 158} \cdot \frac{65 \cdot 158}{81 \cdot 207} = \frac{52 \cdot 207}{33 \cdot 158} \cdot \frac{65 \cdot 158}{81 \cdot 207} = \frac{52 \cdot 65}{33 \cdot 81} = O_{SA}$$

3.1.3 Log Reference Outcome Odds Ratio

The natural logarithms of these odds ratios are called reference-outcome logits. The reference-outcome logits are

$$ln(O_{AB}) = ln(1.63) = 0.49$$,
$$ln(O_{SB}) = ln(2.06) = 0.72$$

and

$$ln(O_{SA}) = ln(1.26) = 0.23$$.

We can see that the last logit is the difference between the two first ones:

$$ln(O_{SA}) = ln(O_{SB}) - ln(O_{AB})$$

(this follows from the logarithmic identities, see section A.2).[3]

We could have estimated the reference outcome logit for Ap vs. Bourgeois by estimating a logistic regression on those who choose one of these two categories while excluding the SV voters. We might have done the same for the other possible pairs of outcomes. But instead we want to find a model that allows us to estimate all these logits at once. That model is the multinomial logistic regression model.

3.2 Multinomial Logistic Regression

Earlier, we derived an expression for logistic regression based on the log odds of an outcome (expression 2.3):

$$ln(\frac{p}{1-p}) = \beta_0 + \beta_1 X$$

In logistic regression the dependent variable has two possible outcomes, but it is sufficient to set up an equation for the logit relative to the reference outcome,

$$ln(\frac{Pr(Y=1)}{Pr(Y=0)})$$.

3.2.1 Specifying the Multinomial Logistic Regression

Multinomial logistic regression is an expansion of logistic regression in which we set up *one equation for each logit* relative to the reference outcome (expression 3.1). '*p*' is ambiguous when there are more than two outcomes. To keep track of the different probabilities we will write *Pr(Y=S)* for the probability of voting SV, *Pr(Y=A)* for Ap, and so on.

---

[3] A note from the translator: see Stock and Watson, p. 268 (2007) or p. 308 (2012).

$$ln(\frac{Pr(Y = S)|X}{Pr(Y = B)|X}) = \beta_0^S + \beta_1^S X$$

$$ln(\frac{Pr(Y = A)|X}{Pr(Y = B)|X}) = \beta_0^A + \beta_1^A X$$

(3.1)

In this case the dependent variable has three categories, so the model has two equations. The equation for the logit of voting SV has two parameters β_0^S and β_1^S. β_0^S is an intercept like in all generalized linear models, and β_1^S is the slope coefficient of the X variable. Likewise, the equation for the logit of voting Ap has the two parameters β_0^A and β_1^A. The superscripts 'S' and 'A' indicate which outcome the parameters belong to.

When estimating a multinomial model for a dependent variable with K categories, we estimate $K - 1$ linear equations. Here, $K = 3$, so we estimate two equations, $\beta_0^S + \beta_1^S X$ and $\beta_0^A + \beta_1^A X$. Logistic regression is therefore a special case of multinomial regression where $K = 2$. The linear expression $\beta_0^S + \beta_1^S X$ tells us more precisely the probability that $Y = S$ *relative* to the probability that $Y = B$. Similarly, the expression $\beta_0^A + \beta_1^A X$ models the probability that $Y = A$ relative to the probability that $Y = B$. There is no need for an equivalent expression for the probability that $Y = B$ because it is given once we know the other two probabilities.

Table 3.2 shows the results obtained when estimating a multinomial logistic regression model with the three-category party choice variable as the dependent variable. The models in column A only have an intercept. The tax variable is added in column B. We include the income variable in column C.

The estimated intercepts in column A are the estimated log odds of voting Ap or SV relative to Bourgeois. The estimates match what we can calculate from the column totals in table 3.1: the odds of voting Ap vs. Bourgeois is $146/365 = 0.4$. The logarithm of the odds is -0.916. Similarly, the odds of voting SV vs. Bourgeois is 0.233, equivalent to a log odds of -1.457.

Table 3.2: Multinomial logistic regression of party choice: SV vs. Ap vs. other. Log odds ratios

	A	B	C
Ap			
Maintain taxes		0.490*	0.503*
		(0.197)	(0.198)
Log income (in thousands)			0.0871
			(0.126)
Constant	-0.916***	-1.158***	-1.658*
	(0.0979)	(0.142)	(0.739)
SV			
Maintain taxes		0.725**	0.695**
		(0.246)	(0.248)
Log income (in thousands)			-0.294*
			(0.115)
Constant	-1.457***	-1.836***	-0.204
	(0.120)	(0.187)	(0.655)
Observations	596	596	596
Log likelihood	-549.9	-543.7	-539.7
χ^2	-2.27e-12	12.31	20.30
Number of parameters	2	4	6

Standard errors in parantheses
* p < 0.05, ** p < 0.01, *** p < 0.001

The estimated intercepts in column B are estimates of the log odds of voting SV/Ap vs. Bourgeois for respondents whose X variable is scored 0; that is, for those who wish to reduce taxes on high incomes. The estimates for the tax variable are the log odds ratios we calculated from table 3.1: the log odds ratio of voting Ap vs. Bourgeois is 0.49 and the log odds ratio of voting SV vs. Bourgeois is 0.72.

There are two estimated effects of the tax variable in the table: one for the Ap equation (β_1^A) and one for the SV equation (β_1^S). The estimate in the Ap equation is 0.490. This means that the log odds of voting Ap vs. Bourgeois is 0.490 higher among tax supporters than among tax opponents, or that the odds are $exp(0.490) = 1.63$ times higher. The estimate in the SV equation is 0.725. The odds of voting SV is $exp(0.725) = 2.06$ times higher among tax supporters than among tax opponents, just as we calculated from table 3.1. We see in column C that these estimates hold when controlling for income.

Table 3.3: Multinomial logistic regression of party choice: SV vs. Ap vs. other. Odds ratios.

	A	B	C
Ap			
Maintain taxes		1.633*	1.653*
		(2.49)	(2.54)
Log income (in thousands)			1.091
			(0.69)
Constant	0.400***	0.314***	0.190*
	(-9.36)	(-8.15)	(-2.24)
SV			
Maintain taxes		2.064**	2.004**
		(2.94)	(2.80)
Log income (in thousands)			0.746*
			(-2.56)
Constant	0.233***	0.159***	0.816
	(-12.10)	(-9.80)	(-0.31)
Observations	596	596	596
Log likelihood	-549.9	-543.7	-539.7
χ^2	-2.27e-12	12.31	20.30

Exponentiated coefficients (odds ratios). z-statistic in parantheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.3 presents the results as odds ratios (as exponentiated coefficients). The odds of voting Ap is 1.6 times higher among tax supporters, and the odds of voting SV is 2.1 times higher — both relative to voting Bourgeois.

3.2.2 Variables vs. Parameters

In multinomial logistic regression as in other multi-equation models, we estimate several different effects of the variables. The tax variable has one effect on the odds of voting SV vs. B, estimated with the parameter β_1^S, and another effect on the odds of voting Ap vs. B, estimated with the parameter β_1^A. We call the different estimated effects 'parameters'. It is important to distinguish between parameters and variables when discussing results from a multinomial model.

3.3 Simplifying Multinomial Logit Models

Multinomial regression models tend to grow large. In table 3.4 we have grouped the parties in five groups: Frp, H, Krf/V/Sp, Ap, SV/RV. We have set Frp as reference category. We have included all the variables from table 2.3. This means we are estimating four equations with 11 parameters each, totaling 44 parameters.

It is undesirable to estimate models as complex as this. More complex models yield less precise parameter estimates. There is also a real danger that we over-fit (over-determine) the model to data when there are relatively few observations per estimated parameter. With 472 observations there are hardly more than 10 observations per parameter. That is dangerously low. How can we simplify the model? We will look at six ways in which to approach this problem:

1. Remove variables from the model;
2. Merge categories in the explanatory variables;
3. Set parameters to 0;
4. Merge categories in the outcome variable;
5. Set parameters to being identical;
6. Simplify the dependent variable.

3.3.1 Removing Variables

Generally speaking, we may remove variables from a model when they are of limited theoretical interest and when removing them does not create a danger of omitted variable bias. If we remove such variables and the estimates for the remaining variables do not change dramatically, the model is often simpler and better.

If we are unsure whether these conditions hold or not, we may formally test whether removing variables reduces how well the model fits the data. The model in table 3.4 has a log likelihood of -604.0. If we remove the age variable and re-estimate the multinomial model, the log likelihood decreases to -609.2.[4] We want to test whether a model that includes the age variable fits the data better than the null model where the age variable is excluded.

---

[4] The results are not reported here.

Table 3.4: Multinomial logistic regression of party choice: five party groupings. Log odds ratios. No restrictions on the parameters.

	H	V/Sp/KrF	Ap	SV/RV
Female	0.196	0.448	0.470	0.434
	(0.399)	(0.390)	(0.399)	(0.440)
Maintain taxes	-0.736	-0.176	-0.0659	0.860
	(0.411)	(0.397)	(0.410)	(0.457)
Log income	-0.119	-0.564	-0.344	-0.450
(in thousands)	(0.376)	(0.359)	(0.370)	(0.379)
'yes' to EU	1.366**	-0.564	1.128**	-0.295
	(0.419)	(0.432)	(0.422)	(0.477)
Environmental	0.365	0.701	1.110	1.724
issues matter	(1.157)	(1.100)	(1.112)	(1.102)
Increase	-0.146	0.284	-0.615	0.0825
foreign aid	(0.744)	(0.686)	(0.750)	(0.741)
Reduce	-0.930*	-2.303***	-2.139***	-1.957***
foreign aid	(0.426)	(0.443)	(0.449)	(0.523)
Pro choice on	0.858*	-0.0362	0.978*	1.187*
abortion	(0.404)	(0.397)	(0.410)	(0.462)
Age (at	-0.0258	-0.0203	-0.0220	-0.0494**
01.12.97)	(0.0153)	(0.0148)	(0.0153)	(0.0169)
Education	0.149	0.0721	-0.0475	0.263
	(0.146)	(0.148)	(0.151)	(0.163)
Constant	1.778	5.809*	3.689	3.413
	(2.463)	(2.400)	(2.457)	(2.498)
Observations	472			
Log likelihood	-604.0			
χ^2	249.4			
Number of parameters	44			

Standard errors in parantheses
* p < 0.05, ** p < 0.01, *** p < 0.001

The likelihood-ratio test is useful for this purpose. We saw in section 2.4 that we can use the likelihood-ratio test if (1) the null model is nested in the alternative model, (2) the structural model is the same, and (3) the data set is the same. The two first conditions are met here because we are still estimating a multinomial model with the same dependent variable and have only removed a variable from the alternative model. The last condition is important to keep in mind. It means that we must be sure to estimate the reduced (null) model on the same observations

as included in the alternative model when we conduct this test. In practice, this means we have to discard observations from the null model that have missing data on the tax variable, but that have data on all other variables.

Table 3.5: Multinomial logistic regression of party choice: five party groupings. Log odds ratios. No restrictions on the parameters, but some variables removed.

	H	V/Sp/KrF	Ap	SV/RV
Maintain taxes	-0.744	-0.108	-0.0987	0.909*
	(0.405)	(0.392)	(0.403)	(0.450)
'yes' to EU	1.310**	-0.720	1.062*	-0.389
	(0.412)	(0.424)	(0.414)	(0.469)
Environmental	0.328	0.761	1.102	1.769
issues matter	(1.153)	(1.097)	(1.108)	(1.097)
Reduce	-0.898*	-2.347***	-2.052***	-1.958***
foreign aid	(0.408)	(0.422)	(0.433)	(0.504)
Pro choice on	0.841*	-0.160	0.973*	1.115*
abortion	(0.400)	(0.391)	(0.404)	(0.456)
Age (at	-0.0210	-0.0182	-0.0182	-0.0470**
01.12.97)	(0.0143)	(0.0139)	(0.0144)	(0.0159)
Education	0.128	0.0387	-0.104	0.232
	(0.140)	(0.142)	(0.146)	(0.157)
Constant	1.036	2.961**	1.925	1.093
	(1.068)	(1.051)	(1.086)	(1.170)
Observations	472			
Log likelihood	-610.2			
χ^2	237.0			

Standard errors in parantheses. Number of parameters: 32
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The LR test is the difference between the log likelihood of the two models multiplied by 2, or 10.42 in this case.[5] There are four equations in table 3.4. Because the age variable has one estimated parameter in each of the equations, a model without the variable has four less parameters. The χ^2 distribution with four degrees of freedom indicates that values greater than 9.49 only appear in 5% of hypothetical samples if the null model is true. Because 10.42 is above this critical value we may discard the hypothesis that the null model fits the data as well as the

---

[5] $LR = -2(-609.2 - (-604.0)) = -2 \times -5.21.$

11

alternative model. This means that we wish to keep the alternative model that includes the age variable.

Similar LR tests indicate that all variables except three yield a significantly better fit to the data. These are the male/female variable, the income variable, and the 'increase foreign aid' variable. Table 3.5 shows the results when re-estimating the model without these three variables.

We run the likelihood ratio test once more to double-check that the model in table 3.5 fits the data as well as the model in table 3.4. The log likelihood has now decreased to −610.2, which is 6.2 units lower than the log likelihood of the model in table 3.4. This means that the LR test statistic is 12.4. We have reduced the model by three variables with four parameters each — a total of 12 parameters. At the 5% significance level, the critical value for chi-squared with twelve degrees of freedom is 21.03. Because 12.4 is less than 21.03, we cannot reject the null hypothesis that the model without these three variables fits as well to the data. Statistically speaking, the model in table 3.5 is as good as the larger model.

The model remains large and unwieldy. Are we sure that the variables distinguish sensibly between the different party groupings, or might we as well merge some of them? To test this, we might impose some *restrictions* on the model.

3.3.2 Merging Categories in the Explanatory Variables

The easiest way to reduce the number of parameters is to reduce the number of categories in categorical explanatory variables. We have already done this when we removed the variable 'increase foreign aid'. It is equivalent to merging the categories 'maintain current levels' and 'increase foreign aid'.

3.3.3 Setting Parameters to 0

One type of restriction is to force the statistical software to set one or more single parameters to 0. For example, we could force the estimate for tax attitude to be 0 in the equation for 'V/Sp/Krf'. This is equivalent to dropping the variable from this particular equation, or to assuming that the variable does not influence the odds of voting 'V/Sp/Krf' relative to Bourgeois.

3.3.4 Merging Categories in the Outcome Variable

We might imagine that two party groupings are so similar that all the variables in our model have the same effect on the odds of voting for them. For example, H and FrP are both considered Bourgeois, and we might think that the odds of voting for one or the other increases equally much for voters that oppose current taxes, want to reduce foreign aid, and consider environmental issues to be secondary to other concerns. If we are correct, all parameter estimates in the H equation should be close to zero. Then there is no reason to treat Frp and H as separate parties. The model would be simpler and better if we merge the two parties into one party grouping.

We can test whether the outcomes 'H' and 'FrP' are the same by setting all parameters (except the intercept) in the 'H' equation to 0, and thereafter re-estimate and compare log likelihood for the restricted model to the model in table 3.5. The LR test statistic comparing the restricted model to that in 3.5 is 34.34, and is clearly significant in the χ^2 distribution with seven degrees of freedom (one for each of the seven variables in the model). Since the LR test is significant, it gives us no reason to prefer the reduced model over the alternative model.

Similar comparisons of the other party groupings to the reference outcome yield the same results. None of the party groupings can be merged with FrP.

3.3.5 Setting Parameters to Being Identical

Another type of restriction is to force the statistics software to yield one or more identical parameters in two equations. For example, we can force the estimates for the tax variable to be identical in the equations for 'V/Sp/KrF' and 'H'. This is equivalent to setting the difference between the two estimates to 0.

We can use this procedure to check whether other party groupings can be merged. For example, we can test whether the outcomes 'H' and 'V/Sp/KrF' are equal by setting all parameters (except the intercept) in the two equations to being identical, and then re-estimate and compare log likelihood for the restricted model to the model in table 3.5. The LR test statistic for this comparison is 73.71. In other words, it suggests that this merging of party groupings fits the data even more poorly than the merging of 'H' and 'FrP'. Similar comparisons of the other pairs of

party groupings yield LR test statistics between 23.80 and 73.71. The tests suggest that none of the party groupings should be merged.

These tests give us no support for simplifying the model by merging party groupings. But we can still use restrictions on selected parameters to simplify the model. For example, it seems that different attitudes towards taxes do nothing to distinguish voters for the centrist parties (V/Sp/KrF) from Ap voters. Changes in this variable are associated with approximately the same change in the log odds of voting for one or the other party grouping (the two estimates are almost the same in table 3.5: -0.108 and -0.099). The difference is much smaller than the estimated standard errors of the parameters. In other words, they are not significantly different from 0: different attitudes towards income tax do not change the log odds relative to the reference outcome, which is voting FrP. We can therefore restrict the parameters for the tax variable to being 0 in the equations for the centrist parties and Ap. The LR test statistic comparing the model in table 3.5 with the reduced model including this last restriction yields 0.08. We have merged three parameters into one and are therefore to compare this test-statistic value with the χ^2 distribution with three degrees of freedom. 0.08 is far from being statistically significant with three degrees of freedom, so the reduced model is preferable to the larger model.

It also looks as if attitudes to EU have about the same effect on the log odds of voting either H or Ap. The difference between these estimates is less than one standard error. We introduce a new restriction setting these two parameters to being equal. The test statistic for comparing these two models is 0.65. We have only reduced the number of parameters by 1, and so we compare to the χ^2 distribution with one degree of freedom. 0.65 is not significant.

We can introduce other restrictions such that the age and environment variables have the same effect on H, the centrist parties, and Ap, but not FrP and SV/RV, and that the foreign aid variable has the same effect on the centrist parties, Ap, and SV/RV, but not FrP and H. Also, the education variable might have the same effect on H and SV/RV, but no effect on the centrist parties and Ap relative to FrP. Finally, we might assume that the abortion variable has the same effect on H, Ap, and SV/RV, and another (but equal) effect on FrP and the centrist parties.

Table 3.6: Multinomial logistic regression of party choice: five party groupings. Restrictions on some parameters. Log odds ratios.

	H	V/Sp/KrF	Ap	SV/RV
Maintain taxes	-0.695**	-0.0502		0.965**
	(0.270)	(0.265)		(0.334)
'yes' to EU	1.152**	-0.739	1.152**	-0.404
	(0.378)	(0.420)	(0.378)	(0.466)
Environmental	0.702	0.686	0.702	1.695
issues matter	(1.084)	(1.094)	(1.084)	(1.095)
Reduce	-0.855*	-2.304***	-2.008***	-1.913***
foreign aid	(0.402)	(0.417)	(0.422)	(0.500)
Pro choice on	0.876*	-0.187	0.876*	1.090*
abortion	(0.368)	(0.389)	(0.368)	(0.455)
Age (at	-0.0179	-0.0166	-0.0179	-0.0454**
01.12.97)	(0.0128)	(0.0135)	(0.0128)	(0.0156)
Education	0.201*	0.116		0.312**
	(0.0867)	(0.0936)		(0.112)
Constant	0.627	2.576**	1.472*	0.691
	(0.864)	(0.895)	(0.749)	(1.025)
Observations	472			
Log likelihood	-612.0			
χ^2	196.1			

Standard errors in parantheses. Number of parameters: 17
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.6 reports the results from re-estimating the model with these restrictions. The parameter estimates show the restrictions that apply. The tax estimates are 0 in the equations for the center parties and Ap. The estimates for EU are 1.152 in both the H and Ap equations—they are forced to being identical.

The model is not a worse fit to the data than the models in tables 3.4 and 3.5—the likelihood ratio tests of the reduced model against these are not significant. The log likelihood of the model in table 3.6 is −612.0. The log likelihood in table 3.5 was −610.2. By restricting the model we have reduced the number of parameters from 32 to 17. The change in log likelihood (multiplied by two) by four units is not significant for the much reduced model.

Restricting this model has many advantages. Models with fewer parameters tend to give clearer results. Look at the estimate for age in the SV/RV equation: the

estimate in table 3.5 was −0.0470 and the standard error 0.0159. The estimate in table 3.6 is slightly larger and the standard error smaller. The estimate for H on the tax variable is slightly smaller in the restricted model, but the standard error is much smaller and the estimate is significant at the 5% level. The estimate for 'reduce foreign aid' in the equations for the centrist parties, Ap, and SV/RV is significant at a much lower level since we use information from all these outcomes to estimate a common parameter.

Restrictions can help us avoid the 'empty cell' problem. What would have happened if no voter for the centrist parties wanted to reduce foreign aid? Then, the observed odds ratio for the centrist parties vs. FrP would have been infinitely large, and the parameter could not have been estimated. By merging outcomes, as we do here, it is possible to yield estimates even when data is sparse.

A downside to imposing restrictions is that the results depend to some extent on the restrictions we choose. Earlier, we assumed that the parameters for the education variable were identical for Ap and the centrist parties, but we might as well have assumed that they were identical in the equations for H and the centrist parties. We could define another restriction, yielding slightly different results. We may sometimes have good reasons for choosing one restriction over another. In this instance we could have referred to earlier research showing that education has about the same impact on voting centrist parties as on voting Ap. If we have no prior research or specific theoretical reasons for specifying the model in one way or another, we should try different specifications to see how robust the results are to alternative specifications.

3.3.6 Simplifying the Dependent Variable

The estimates for the environment variable in table 3.5 follow a clear pattern. Respondents who say that environmental issues influence their voting have slightly higher odds of voting H than FrP, even higher odds of voting the centrist parties, yet higher odds of voting Ap, and the highest odds of voting SV. This pattern suggests a final way of simplifying the model. If we are willing to assume that the outcomes are rank-ordered, for example along an environmental dimension, then we can merge the outcomes by assuming that they are placed in different positions along this dimension. This simplification yields the ordinal logistic regression model, which is the topic of the next chapter.

3.4 Predicted Probabilities

How should we interpret the results in table 3.6? For example, what is the substantive impact of the tax variable?

We can say quite a bit just by looking at the estimates in the table. The estimate in the H equation is −0.695. It means that the log odds of voting H relative to Frp is 0.695 lower among those who wish to maintain the tax rate on high incomes relative to those who wish to reduce taxes. Equivalently, we can say that the odds of voting H relative to Frp is 0.50 times higher (or half the size). The estimates for this variable are 0 in the equations for the center parties and Ap because we imposed a constraint assuming that the effect of tax attitudes is the same for center parties, Ap, and FrP. The estimate for the tax variable is 0.965 in the SV/RV equation. In other words, the odds of voting for the most leftist parties is *exp*(0.965) = 2.62 times higher among those who wish to maintain the tax rates on high incomes. The difference between the estimates in the SV/RV and H equations is 0.965−(−0.695) = 1.66. This means that the odds of voting SV/RV relative to H is *exp*(1.66) = 5.26 times higher among those who wish to maintain the tax levels.

We can also see these effects by looking at changes in the predicted probabilities of the five outcomes for given changes in the independent variables. In the chapter on logistic regression we established an expression (2.2) for the logistic regression model, relating it to the *probability* of an outcome:

$$Pr(Y = 1|X) = \frac{exp(\beta_0 + \beta_1 X)}{1 + exp(\beta_0 + \beta_1 X)}$$

We can do the same for multinomial logistic regression:

$$Pr(Y = S|X) = \frac{exp(\beta_0^S + \beta_1^S X)}{1 + exp(\beta_0^S + \beta_1^S X) + exp(\beta_0^A + \beta_1^A X)}$$

$$Pr(Y = A|X) = \frac{exp(\beta_0^A + \beta_1^A X)}{1 + exp(\beta_0^S + \beta_1^S X) + exp(\beta_0^A + \beta_1^A X)}$$

(3.2)

Table 3.7: Descriptive statistics for the variables in table 3.6

	FrP	H	V/Sp/KrF	Ap	SV/RV		
Pr(y|x)	0.071	0.239	0.292	0.264	0.135		
	Age	Education	Tax attitude	'yes' to EU	Environmental issues matter	Reduce for. aid	Pro choice
Mean of x	45.8	4.17	0.494	0.445	0.102	0.222	0.593
Std.dev. of x	14.6	1.58	0.500	0.497	0.303	0.416	0.491

We can use this formula to calculate the probability of voting SV and Ap for tax supporters, based on the estimates in column B of table 3.2. The variable 'tax attitude' has the value $X = 1$ for tax supporters. The part of the table labeled 'SV' contains the estimates for β_0^S and β_1^S. We can enter the estimates in the linear expression for the SV equation for an observation where $X = 1$:

$$\beta_0^S + \beta_1^S X = -1.836 + 0.725 = -1.111$$

In the same way, we can enter the estimates from the AP equation:

$$\beta_0^S + \beta_1^S X = -1.158 + 0.490 = -0.668$$

We find the probability of voting SV by inserting these two elements in the top line of expression (3.2):

$$Pr(Y = S|X) = \frac{exp(\beta_0^S + \beta_1^S X)}{1 + exp(\beta_0^S + \beta_1^S X) + exp(\beta_0^A + \beta_1^A X)}$$

$$= \frac{exp(-1.111)}{1 + exp(-1.111) + exp(-.668)} = \frac{0.329}{1 + 0.329 + 0.513} = 0.179$$

We find the probability of voting Ap in the same way:

$$Pr(Y = A|X) = \frac{exp(\beta_0^A + \beta_1^A X)}{1 + exp(\beta_0^S + \beta_1^S X) + exp(\beta_0^A + \beta_1^A X)}$$

$$= \frac{exp(-0.668)}{1 + exp(-1.111) + exp(-.668)} = \frac{0.0.513}{1 + 0.329 + 0.513} = 0.278$$

18

Table 3.8: Change in predicted probability

Probability of voting	Tax attitude		EU vote		Reduce foreign aid	
	reduce	maintain	'no'	'yes'	no	yes
FrP	0.068	0.072	0.078	0.052	0.048	0.222
(95% C.I.)	(0.038, 0.098)	(0.040, 0.103)	(0.042, 0.113)	(0.020, 0.084)	(0.023, 0.072)	(0.133, 0.311)
Høyre	0.323	0.170	0.155	0.339	0.196	0.378
(95% C.I.)	(0.256, 0.390)	(0.116, 0.223)	(0.117, 0.193)	(0.280, 0.397)	(0.152, 0.241)	(0.271, 0.484)
Center	0.280	0.280	0.428	0.150	0.320	0.169
(95% C.I.)	(0.212, 0.348)	(0.211, 0.349)	(0.362, 0.495)	(0.098, 0.202)	(0.266, 0.373)	(0.120, 0.219)
Ap	0.252	0.265	0.172	0.374	0.289	0.153
(95% C.I.)	(0.196, 0.308)	(0.206, 0.324)	(0.132, 0.211)	(0.315, 0.434)	(0.241, 0.337)	(0.108, 0.198)
SV/RV	0.077	0.213	0.168	0.085	0.148	0.078
(95% C.I.)	(0.041, 0.114)	(0.150, 0.277)	(0.116, 0.219)	(0.047, 0.124)	(0.107, 0.188)	(0.049, 0.107)

These are the same probabilities that we saw in table 3.1.

We report descriptive statistics in table 3.7. The top row shows the predicted probabilities of voting for each of the five party groupings when all explanatory variables are at their mean. The two bottom rows report means and standard deviations for the seven variables in table 3.6.

The two first columns with numbers in table 3.8 show the predicted probabilities of voting for FrP, H, or any of the other party groupings for respondents who wish to reduce or maintain the tax rates on high incomes. The numbers in brackets are 95% confidence intervals around the predictions. All other variables are held at their mean as shown in table 3.7.[6] The change in the probability of voting FrP is negligible for a change in tax attitudes. In contrast, the probability of voting H is reduced from 0.323 to 0.170—it is almost halved. The probability of voting SV/RV is almost three times as high for those who want to maintain tax rates as opposed to those who want to reduce taxes.

Attitudes towards Norwegian EU membership have a sizeable impact on voting. Respondents who voted 'yes' to joining the EU in 1994 have a somewhat lower probability of voting FrP than those who voted 'no'; they have more than twice as high a probability of voting for H or Ap, but less than half the probability of voting for the centrist parties or SV/RV.

---

[6] Since logistic regression models are non-linear, changes in predicted probabilities will depend on the values of other variables in the model. Setting remaining variables to their mean yields a reasonable comparison. An alternative could have been to set all remaining variables at their median or mode.

Table 3.9: Change in predicted probability

Variable	FrP	H	Center	Ap	SV/RV
Maintain taxes (0→1)	0.0036	-0.153	0.0003	0.013	0.136
'yes' to EU (0→1)	-0.023	0.189	-0.285	0.208	-0.090
Environment matters (0→1)	-0.044	-0.034	-0.044	-0.037	0.159
Reduce foreign aid (0→1)	0.169	0.177	-0.183	-0.115	-0.047
Pro choice on abortion (0→1)	-0.037	0.084	-0.211	0.093	0.071
Age (±0.5 std.dev.)	0.021	0.0070	0.014	0.0077	-0.049
Education (±0.5 std.dev.)	-0.014	0.030	-0.0032	-0.052	0.039

Table 3.8 also shows that attitudes to foreign aid explain how people vote. Those who want to reduce foreign aid have a considerably higher probability of voting for FrP or H, and lower probability of voting for the centrist parties, Ap, or SV/RV.

Table 3.9 demonstrates a last way in which we can present the results from table 3.6. We have, for each variable, reported changes in the probability of each of the outcomes for given changes in the explanatory variables. All remaining variables are set to their mean. For example, if we compare two respondents with different attitudes towards taxes, the individual who wishes to maintain taxes has a 0.0036 higher probability of voting FrP than the individual who wishes to reduce them. This is the same difference as in the two cells for FrP in the top row of table 3.8.

Table 3.9 reports fewer details and therefore has room for all the variables in the model. For example, we can see that the environment variable increases the probability of voting for SV/RV by 15.9%, the abortion variable reduces the probability of voting for the centrist parties by 21.1%, and education reduces the probability of voting for Ap or the centrist parties.

Table 3.9 also shows another aspect of the multinomial model that is worth remembering. When we simplified the model to the version in table 3.6, we assumed that the level of education does not affect the odds of voting Ap vs. the reference outcome (FrP). Even so, table 3.9 shows that education reduces the probability of voting for Ap, and to a greater extent than is the case for FrP. Why is that so?

This derives from the fact that the odds of voting H and SV/RV increase strongly with education. It means that when education increases, a greater share of respondents vote for these two party groupings. In other words, there are fewer left who can vote for FrP or Ap. So the odds of voting for these two party groupings

decrease even if the relative odds of voting FrP or Ap — the relative distribution between the parties — is not affected by education. The probability of voting Ap decreases more than the probability of voting FrP because Ap in 2001 was a much larger party than FrP.

Chapter 4

Ordinal Logistic Regression

Ordinal logistic regression is the third model based on odds and odds ratios. It could be considered either as a simplification of multinomial logistic regression (as mentioned in the previous chapter) or as the analysis of a grouped continuous variable.

4.1 Cumulative Odds and Odds Ratios

Multinomial logistic regression is often inefficient. Let us again cross-tabulate party choices and tax attitudes to derive a new form of simplification (table 4.1). We see that those who want to maintain the tax rates on high incomes have a higher probability of voting Ap than those who want to reduce taxes, and that supporting taxes increases the probability of voting SV to an even greater extent. It is clear that the tax variable is such that it increases the probability of voting for parties to the left. The multinomial logistic regression does not take advantage of this.

Table 4.1: Voting for Ap vs. SV vs. other parties as a function of attitudes towards current taxes

Attitude towards current ta xes	Other		Ap		SV		Total	
	No.	%	No.	%	No.	%	No.	%
Reduce taxes	207	67.9	65	21.3	33	10.8	305	100.0
Maintain taxes	158	54.3	81	27.8	52	17.9	291	100.0
Total	365	61.2	146	24.5	85	14.3	596	100.0

Source: Valgundersøkelsen 2001 (Aardal et al., 2003)

In chapter 3 we calculated reference-outcome odds and built the model around these. The reference-outcome odds for Ap was

$$o_{Ap} = 0.245/0.612 = 0.40$$

Table 4.2: Cumulative probabilities, odds, and log odds

Attitude towards current ta xes	Cumulative probabilities			Cumulative odds		Log cumulative odds	
	$j = 1$: Bourgeois	$j = 2$: Ap	$j = 3$: SV	Odds(Y>1) (Ap/SV vs. B)	Odds(Y>2) (SV vs. B/Ap)	L_{k1} α_1	L_{k2} α_2
Reduce taxes	0.679	0.892	1.000	0.473	0.121	−0.748	−2.109
Maintain taxes	0.543	0.821	1.000	0.842	0.218	−0.172	−1.525
All	0.612	0.857	1.000	0.633	0.166	−0.457	−1.794

and the equivalent reference-outcome odds for SV was

$$o_{SV} = 0.143/0.612 = 0.23$$

But now it is natural to think of the three party groupings as being placed on a right–left dimension. It can therefore be helpful to calculate the odds of being to the left of some point relative to being to the right of that point. The odds of voting Ap or something to the left of Ap are then

$$o_{left_1} = \frac{(0.245 + 0.143)}{0.612} = \frac{0.388}{0.612} = 0.633$$

The odds of voting SV or something to the left of SV are

$$o_{left_2} = \frac{0.143}{(0.612 + 0.245)} = \frac{0.143}{0.857} = 0.166$$

Odds such as these are called *cumulative odds*.

Let us introduce some notation to define cumulative odds more precisely. Let the party categories be called j. There are $J = 3$ categories: Bourgeois: $j = 1$, Ap: $j = 2$, and SV: $j = 3$.

We define the *cumulative probability* as the probability of being in category j or lower: $P(Y \leq j) = p_1 + \cdots + p_j$. It is called cumulative because it is the sum of the probabilities of being in the categories 1, 2, and so on, up to j. In the row for those who want to reduce taxes (table 4.1) it means that:

- For $j = 1$ (Bourgeois): $P(Y \leq 1) = p_1 = 0.679$
- For $j = 2$ (Ap): $P(Y \leq 2) = p_1 + p_2 = 0.892$

- For $j = 3$ (SV): $P(Y \leq 3) = p_1 + p_2 + p_3 = 1.000$

The cumulative probabilities are arranged like this to reflect the ordering of the party variable, and it makes sense only if the ordering is natural. In table 4.2 we have calculated all the relevant cumulative probabilities as well as other figures that we need. In the two top rows we have the cumulative probabilities and odds for the two tax attitude categories. In the last row we have reported the same for the entire sample.

Then, the *cumulative odds* for the first $J - 1$ categories are:

$$Odds(Y > j) = \frac{P(Y > j)}{P(Y \leq j)} = \frac{1 - P(Y \leq j)}{P(Y \leq j)} = \frac{p_{j+1} + \ldots + p_J}{p_1 + \ldots + p_j}$$

We have arranged the odds to reflect the likelihood of being in a high category versus a lower category.[7] The cumulative odds of voting Ap or to the left of Ap for those who want to reduce taxes are then

$$Odds(Y > 1) = \frac{1 - P(Y \leq 1)}{P(Y \leq 1)} = \frac{1 - 0.679}{0.679} = 0.473$$

The cumulative odds of voting SV or to the left of SV are

$$Odds(Y > 2) = \frac{1 - 0.892}{0.892} = 0.121$$

The four other cumulative odds are summarized in table 4.2.

As before, we have to transform the odds to log odds or logits in order to use them in a generalized linear model. Cumulative log odds for the first $J - 1$ categories is

$$logit(Y > j) = ln\left(\frac{P(Y > j)}{P(Y \leq j)}\right) = ln\left(\frac{p_{j+1} + \ldots + p_J}{p_1 + \ldots + p_j}\right)$$

---

[7] Other textbooks arrange this the opposite way, that is *logit*[$P(Y \leq j)$] as opposed to *logit*[$P(Y > j)$] as here (see, for example, Agresti 2002). The advantage of arranging the cumulative odds the way we do here is that it yields the same parameter signs as those reported in Stata.

The cumulative log odds among tax opponents is the logarithm of the cumulative odds we calculated above. Let us call it L_{01}. The '0' refers to category 0 on the independent variable, and the '1' refers to the odds of $(Y > 1)$. In a calculation this is:

$$L_{01} = logit_0(Y > 1) = ln(0.473) = -0.748$$

The remaining log odds are also reported in table 4.2.

4.2 The Proportional Odds Model

We can use the cumulative log odds to make a model with few parameters if we are willing to assume that an explanatory variable changes the different cumulative odds equally much. Such an assumption of proportional odds is the foundation of the ordinal logistic regression model. This model is also called the 'proportional odds model'.

Let us consider what the proportional odds assumption means here. We first have to consider what the tax variable does to the odds of the outcomes as they appear in table 4.1.

We see from table 4.2 that the odds of voting Ap or SV relative to Bourgeois ($odds(Y > 1)$) is 0.842 among those who wish to maintain taxes. The odds among tax opponents are $odds(Y > 1) = 0.473$. In other words, the odds of voting Ap or SV is 1.78 times higher among tax supporters than among tax opponents. The odds of voting SV ($odds(Y > 2$ relative to $Y \leq 2$)) is 0.218 among those who want to maintain taxes, or 1.79 times higher than 0.121, which is the equivalent odds for the other group. These cumulative odds are almost identical.

We can see the same relationship by looking at log odds. The difference between L_{01} and L_{11} tells us how much the cumulative log odds of voting Ap or higher increases when we compare tax opponents to tax supporters:

$$LOR_1 = L_{11} - L_{01} = -0.172 - (-0.748) = 0.576$$

26

Table 4.3: Ordinal logistic regression of party choice: SV vs. Ap. vs other. Log odds ratios.

	A	B
cut1	0.457***	0.748***
α	(0.0841)	(0.121)
cut2	1.794***	2.105***
α	(0.117)	(0.151)
Maintain taxes		0.577***
		(0.165)
Observations	596	596
Log likelihood	-549.9	-543.7
χ^2	-2.27e-12	12.31

Standard errors in parantheses
* p < 0.05, ** p < 0.01, *** p < 0.001

The equivalent difference for cumulative log odds of SV or higher is $L_{12} - L_{02} = 0.584$. These are log odds ratios for cumulative odds as we see in the table. They are quite similar, so assuming that they are equal will probably not yield a much poorer fit to the data.

We have coded the data such that $X_1 = 0$ for respondents who want to reduce taxes and $X_1 = 1$ for those who want to maintain them. $L_{1k} - L_{0k}$ is therefore the change in cumulative log odds when we increase X_1 by one unit. We can use this to specify a proportional odds model:

$$logit_i(Y > j) = \alpha_j + \beta_1 X_i \qquad j = 1, ..., J - 1$$

As in all generalized linear models, the parameter β_1 reflects the change in the dependent variable when we increase X_1 by one unit. Here, the dependent variable is cumulative log odds ($logit[P(Y > j)]$). The parameters α_j are log cumulative odds of each category on the outcome variable when all X variables are 0. In other words, they are a kind of intercept. In table 4.2 all X variables are 0 in the group that wants to reduce taxes. α_1 is log cumulative odds of voting Ap or SV vs. Bourgeois as reference outcome — what we called L_{01} above. α_2 is the log cumulative odds of voting SV vs. Bourgeois or Ap — what we called L_{02} above.

In table 4.3 we have estimated this ordinal logistic regression model on the data from table 4.1. Model A only has an intercept whereas model B includes the tax

variable. The estimates labeled 'cut1' and 'cut2' are the same as $-\alpha_1$ and $-\alpha_2$. Many statistical software packages report $-\alpha_j$ rather than α_j. In column A we have omitted any explanatory variables. The 'cut points' that the statistics software reports are the same log cumulative odds as for the whole sample, as shown in table 4.2, but with opposite sign.

We have added the tax variable in column B. The estimated $\beta_1 = 0.577$ suggests that cumulative log odds of voting for parties to the left increases by 0.577 when the tax variable increases by one unit. This is very similar to what we calculated based on table 4.1. We found, in that table, that $LOR_1 = 0.576$ and $LOR_2 = 0.584$ (see table 4.2). The estimate from the ordinal logistic model is close to a weighted average of these two odds ratios.[8]

The estimated cut points $-\alpha_j$ are also quite similar to those we calculated from table 4.1. The cut points are a kind of intercept, and the estimated cumulative log odds for the case where all X variables are 0. In this case, this would be when the tax variable is in the reference category, corresponding to respondents who want to reduce taxes. The first cut point is the same as L_{01}, or log cumulative odds of voting for Ap or further left, but with opposite sign. The other cut point represents the log cumulative odds of voting for SV vs. parties to the right of SV.

In table 4.4 we report the same results with exponentiated coefficients, that is, in odds ratio form. As in table 3.2 the intercepts are omitted. The table shows that the odds of voting a notch further to the left increases with 78% when we compare respondents who want to maintain taxes to those who want to reduce them.

It is helpful to compare the results in table 4.4 with the results from our multinomial logistic regression (table 3.3 (or the equivalent results in log odds ratio form)). The estimates for the tax variable in table 3.3 were 1.633 in the Ap equation and 2.064 in the SV equation. Changes in tax attitudes therefore change the odds of voting Ap vs. Bourgeois by 63%, and the odds of voting SV vs. Bourgeois by 106%. We find in the ordinal model that the odds of voting Ap *or* SV vs. Bourgeois increase by 78%. This result is pretty similar to what we found for Ap and SV in the multinomial model. The difference is that we have assumed that the odds of voting SV vs. Ap *or* Bourgeois also increases by 78%. Column B in table 4.3

---

[8] In contrast to model B in table 3.2, this model is not saturated, and so the odds ratios are not identical to what we see in table 4.1.

therefore only has three estimated parameters, whereas the same column in 3.2 has four.

Table 4.4: Ordinal logistic regression of party choice: SV vs. Ap vs. other. Odds ratios.

	A	B
Maintain taxes		1.781***
		(0.295)
Observations	596	596
Log likelihood	-549.9	-543.7
χ^2	-2.27e-12	12.31

Standard errors in parantheses
* p < 0.05, ** p < 0.01, *** p < 0.001

The estimated log likelihood of the multinomial logistic regression model in table 3.3 was −543.7, almost identical to the log likelihood in table 4.3. The ordinal logistic regression model is nested in the multinomial model, so we can perform a formal likelihood ratio test. The fact that the log likelihood is almost the same in the simpler model is in itself an indication that the ordinal model fits as well to the data. In this instance, the ordinal model is preferable to the multinomial model because it is simpler.

How can we know whether the assumption of proportional odds holds or not? Statistical software offer several tests of this assumption. Roughly speaking, they work as follows: the program estimates an ordinal logistic regression model that allows β_k to differ for each cut point (similarly to multinomial logistic regression, but with cumulative log odds as the metric) and yields the log likelihood of this model. The program then re-estimates the model under the condition that β_k is to be the same for each cut point. It then compares the log likelihood of the reduced model to the log likelihood of the more general alternative model.

In this case, such a test yields a χ_2 statistic of less than 0.01 with one degree of freedom. It is very far from statistically significant, and so we have no foundation for rejecting the reduced model. There is no doubt that we can assume that parties are ordered along a right–left axis with respect to the tax variable. The ordinal logistic regression model is therefore preferable to the multinomial model because it is more parsimonious — it has fewer parameters.

Table 4.5: Ordinal logistic regression of party choice: SV vs. Ap. vs other. Log odds ratios.

	ologit			mlogit	
	triko01	cut1	cut2	Ap	SV
Maintain taxes	0.746***			0.367	1.224***
	(0.202)			(0.242)	(0.318)
Age (at	-0.0153*			-0.00473	-0.0293**
01.12.97)	(0.00712)			(0.00850)	(0.0111)
Education	-0.00792			-0.127	0.125
	(0.0697)			(0.0842)	(0.107)
Female	0.127			0.143	0.183
	(0.191)			(0.229)	(0.299)
'yes' to EU	0.110			0.849***	-0.605
	(0.201)			(0.242)	(0.329)
Environmental	1.187***			0.540	1.378***
issues matter	(0.321)			(0.438)	(0.406)
Increase	-0.160			-0.557	-0.0263
foreign aid	(0.285)			(0.382)	(0.392)
Reduce	-0.747**			-0.849**	-0.687
foreign aid	(0.259)			(0.304)	(0.419)
Pro choice on	0.726***			0.716**	0.820*
abortion	(0.213)			(0.252)	(0.328)
$-\alpha_j$		0.632	2.119***	-1.121	-1.908*
		(0.528)	(0.539)	(0.638)	(0.794)
Observations	493			493	
Log likelihood	-417.9			-397.7	
χ^2	66.81			107.2	

It looks like the three party groupings are arranged along a right–left axis with respect to the tax variable, but such an ordering may not make sense with respect to other variables. For example, in the previous chapter we saw that positive attitudes to EU increase the probability of voting Ap relative to any Bourgeois party, but reduce the probability of voting SV. In table 4.5 we have added more independent variables. The three columns labeled 'ologit' show the results of estimating an ordinal logistic regression model. The two columns labeled 'mlogit' show the results of estimating a multinomial model with the same variables.

The test of the proportional odds assumption yields $\chi_2 = 30.43$. Compared to the chi squared distribution with nine degrees of freedom it shows that the general

model fits the data significantly better than the reduced proportional odds model. It does not make sense to say that the three party groupings are placed along a single dimension for all or most of the variables in this model. The Brant test of proportional odds suggests that the strongest violation of the assumption is for the education and EU variables. We can see why if we compare the estimates for these two variables in the two models in table 4.5. The education variable has a negative estimate in the Ap equation in the multinomial model, but a positive estimate in the SV equation. Similarly, the EU variable has a positive estimate for Ap and a negative estimate for SV in the multinomial model. In the ordinal model we assume that a positive attitude increases the probability of voting Ap and SV relative to Bourgeois, and that it increases the probability of voting SV vs. Ap or Bourgeois even more. It is evident that the proportional odds assumption does not sit well with our data.

A final indication that the ordinal logit regression fits poorly is the estimated log likelihood, which is significantly lower in the ordinal than in the multinomial model.

4.3 Predicted Probabilities

We can also write the proportional odds model as the probability p_{ij} that observation i is in outcome-category j:

$$p_{ij} = Pr(-\alpha_{j-1} < \beta X_i + u_i <= -\alpha_j)$$

$$= \frac{1}{1 + exp(\alpha_j + \beta X_i)} - \frac{1}{1 + exp(\alpha_{j-1} + \beta X_i)}$$

The first part of this expression shows why many statistics packages present $-\alpha_j$, that is with the opposite sign than the notation we have used in this chapter. With the opposite sign the parameter is clearly associated with the predicted probability of an outcome.

We can use this formula to calculate the estimated probability of voting Ap ($j = 2$) when $X_1 = 0$, based on the estimates in column B of table 4.3:

$$Pr(Y = 2) = \frac{1}{1 + exp(\alpha_j + \beta X_i)} - \frac{1}{1 + exp(\alpha_{j-1} + \beta X_i)}$$

$$= \frac{1}{1 + exp(-2.105 + 0)} - \frac{1}{1 + exp(-.784 + 0)} = 0.891 - 0.687 = 0.205$$

We find the probability of voting Ap ($j = 2$) when $X_1 = 1$ in the same way:

$$Pr(Y = 2) = \frac{1}{1 + exp(-2.105 + 0.577)} - \frac{1}{1 + exp(-0.784 + 0.577)}$$

$$= 0.822 - 0.552 = 0.270$$

These predicted probabilities are fairly similar to those we observed in table 4.1.

Martin Austvoll Nome

17 April 2011

Predicted probabilities in ordinal logit regression: the Hegre/STATA expression and the Stock & Watson expression.

Following Hegre/STATA:

Consider a *J*-category ordinal dependent variable.

All expressions below hold for a regression with only one independent variable, using STATA output.

In ordinal logistic regression, using STATA results, the predicted probability that observation *i* is in category *j* is:

$$p_{ij} = \frac{1}{1 + exp(\alpha_j + \beta X_i)} - \frac{1}{1 + exp(\alpha_{j-1} + \beta X_i)}$$

Following this notation, the probability that observation *i* is in category 1 is then:

$$p_{i1} = \frac{1}{1 + exp(\alpha_1 + \beta X_i)}$$

And the probability that observation *i* is in category *J* is:

$$p_{iJ} = 1 - \frac{1}{1 + exp(\alpha_{J-1} + \beta X_i)}$$

Following Stock & Watson's expression for turning logit into probability:

The above relationships hold, following the notation in Hegre (2011), and — I presume — the notation in the STATA ologit manual entry.

However, this way of setting up the predicted probabilities might not be the most accessible way of teaching it to students.

Students will already have learnt to transform a logit to a probability in binary logistic regression. Following Stock & Watson:

$$\hat{p}_i = \frac{1}{1 + e^{-(\hat{\beta}_0 + \hat{\beta}X_i)}} \equiv \frac{1}{1 + exp\left(-(\hat{\beta}_0 + \hat{\beta}X_i)\right)}$$

When teaching ordinal logistic regression, it might be more accessible to use this expression as the starting point for expressing the probability of the jth outcome, and to teach it in terms of the difference between cumulative probabilities.

Then, the probability that observation i is in category 1 would be 1 minus the cumulative probability of category 2:

$$p_{i1} = 1 - \frac{1}{1 + exp\left(-(\alpha_1 + \beta X_i)\right)}$$

The predicted probability that observation i is in category j would be the difference between the cumulative probability of category j and the cumulative probability of category j +1:

$$p_{ij} = \frac{1}{1 + exp\left(-(\alpha_{j-1} + \beta X_i)\right)} - \frac{1}{1 + exp\left(-(\alpha_j + \beta X_i)\right)}$$

And the probability that observation i is in category J would be equivalent to the cumulative probability of category J:

$$p_{iJ} = \frac{1}{1 + exp\left(-\left(\alpha_{J-1} + \beta X_i\right)\right)}$$