

5 Models for ordinal outcomes

Although the categories for an ordinal variable can be ordered, the distances between the categories are unknown. For example, in survey research, questions often provide the response categories of strongly agree, agree, disagree, and strongly disagree, but an analyst would probably not assume that the distance between strongly agreeing and agreeing is the same as the distance from agree to disagree. Educational attainments can be ordered as elementary education, high school diploma, college diploma, and graduate or professional degree. Ordinal variables also commonly result from limitations of data availability that require a coarse categorization of a variable that could, in principle, have been measured on an interval scale. For example, we might have a measure of income that is simply low, medium, or high.

Ordinal variables are often coded as consecutive integers from 1 to the number of categories. Perhaps because of this coding, it is tempting to analyze ordinal outcomes with the linear regression model. However, an ordinal dependent variable violates the assumptions of the LRM, which can lead to incorrect conclusions, as demonstrated strikingly by McKelvey and Zavoina (1975, 117) and Winship and Mare (1984, 521–523). Accordingly, with ordinal outcomes it is much better to use models that avoid the assumption that the distances between categories are equal. Although many different models have been designed for ordinal outcomes, in this chapter we focus on the logit and probit versions of the *ordinal regression model* (ORM), introduced by McKelvey and Zavoina (1975) in terms of an underlying latent variable and in biostatistics by McCullagh (1980), who referred to the logit version as the *proportional odds model*.

As with the binary regression model, the ORM is nonlinear, and the magnitude of the change in the outcome probability for a given change in one of the independent variables depends on the levels of all the independent variables. As with the BRM, the challenge is to summarize the effects of the independent variables to fully reflect key substantive processes without overwhelming and distracting detail. For ordinal outcomes, as well as for the models for nominal outcomes in chapter 6, the difficulty of this task is increased by having more than two outcomes to explain.

Before proceeding, we caution that researchers should think carefully before concluding that their outcome is indeed ordinal. Simply because the values of a variable *can* be ordered, do not assume that the variable *should* be analyzed as ordinal. A variable that can be ordered when considered for one purpose could be unordered or ordered differently when used for another purpose. Miller and Volker (1985) show how different assumptions about the ordering of occupations resulted in different conclusions. A variable might also reflect ordering on more than one dimension, such as attitude

scales that reflect both the intensity and the direction of opinion. Moreover, surveys commonly include the category “don’t know”, which probably does not correspond to the middle category in a scale, even though analysts might be tempted to treat it this way. Overall, when the proper ordering is ambiguous, the models for nominal outcomes discussed in chapter 6 should be considered.

We begin by reviewing the statistical model, followed by an examination of testing, fit, and methods of interpretation. These discussions are intended as a review for those who are familiar with the models. For a complete discussion, see Long (1997). We end the chapter by considering several less common models for ordinal outcomes, which can be fitted using ado-files that others have developed. We also introduce the stereotype logit model, added in Stata 9 with the `slogit` command, but we postpone a full discussion until chapter 6. As always, you can obtain sample do-files and data files by downloading the `spost9.do` and `spost9.ado` packages (see chapter 1 for details).

5.1 The statistical model

The ORM can be developed in different ways, each of which leads to the same form of the model. These approaches to the model parallel those for the BRM. Indeed, the BRM can be viewed as a special case of the ordinal model in which the ordinal outcome has only two categories.

5.1.1 A latent-variable model

The ordinal regression model is commonly presented as a latent-variable model. Defining y^* as a latent variable ranging from $-\infty$ to ∞ , the *structural model* is

$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i$$

Or, for the case of one independent variable,

$$y_i^* = \alpha + \beta x_i + \varepsilon_i$$

where i is the observation and ε is a random error, as discussed further below.

The *measurement model* for binary outcomes is expanded to divide y^* into J ordinal categories,

$$y_i = m \quad \text{if } \tau_{m-1} \leq y_i^* < \tau_m \quad \text{for } m = 1 \text{ to } J$$

where the *cutpoints* τ_1 through τ_{J-1} are estimated. (Some authors refer to these as thresholds.) We assume $\tau_0 = -\infty$ and $\tau_J = \infty$ for reasons that will be clear shortly.

To illustrate the measurement model, consider the example used in this chapter. People are asked to respond to the following statement:

A working mother can establish just as warm and secure of a relationship with her child as a mother who does not work.

Possible responses are 1 = Strongly Disagree (SD), 2 = Disagree (D), 3 = Agree (A), and 4 = Strongly Agree (SA). The continuous latent variable can be thought of as the *propensity* to agree that working mothers can be warm and secure mothers. The observed response categories are tied to the latent variable by the measurement model:

$$y_i = \begin{cases} 1 \Rightarrow \text{SD} & \text{if } \tau_0 = -\infty \leq y_i^* < \tau_1 \\ 2 \Rightarrow \text{D} & \text{if } \tau_1 \leq y_i^* < \tau_2 \\ 3 \Rightarrow \text{A} & \text{if } \tau_2 \leq y_i^* < \tau_3 \\ 4 \Rightarrow \text{SA} & \text{if } \tau_3 \leq y_i^* < \tau_4 = \infty \end{cases}$$

Thus when the latent y^* crosses a cutpoint, the observed category changes. Anderson (1984) referred to ordinal variables created in this fashion as “grouped continuous” variables, and referred to what we call the ordinal regression model as the “grouped continuous model”.

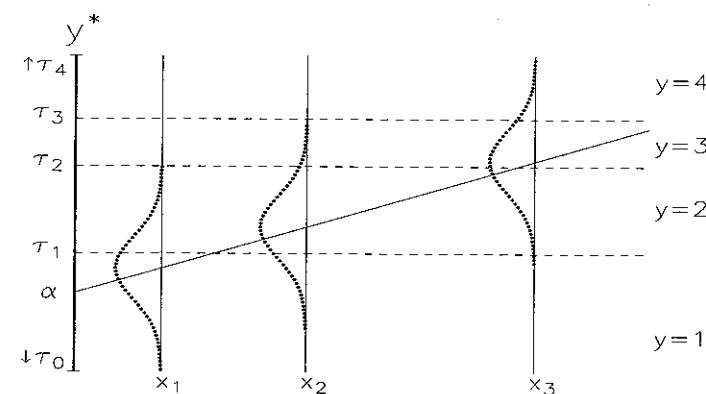


Figure 5.1: Relationship between observed y and latent y^* in ordinal regression model with one independent variable.

For one independent variable, the structural model is $y^* = \alpha + \beta x + \varepsilon$, which is plotted in figure 5.1 along with the cutpoints for the measurement model. This figure is similar to that for the binary regression model, except that there are now three horizontal lines representing the cutpoints τ_1 , τ_2 , and τ_3 . The three cutpoints lead to four levels of y that are labeled on the right-hand side of the graph.

The probability of an observed outcome for a given value of x is the area under the curve between a pair of cutpoints. For example, the probability of observing $y = m$ for given values of the x s corresponds to the region of the distribution where y^* falls between τ_{m-1} and τ_m :

$$\Pr(y = m | \mathbf{x}) = \Pr(\tau_{m-1} \leq y^* < \tau_m | \mathbf{x})$$

Substituting $\mathbf{x}\boldsymbol{\beta} + \varepsilon$ for y^* and using some algebra leads to the standard formula for the predicted probability in the ORM,

$$\Pr(y = m | \mathbf{x}) = F(\tau_m - \mathbf{x}\boldsymbol{\beta}) - F(\tau_{m-1} - \mathbf{x}\boldsymbol{\beta}) \quad (5.1)$$

where F is the cdf for ε . In ordinal probit, F is normal with $\text{Var}(\varepsilon) = 1$; in ordinal logit, F is logistic with $\text{Var}(\varepsilon) = \pi^2/3$. For $y = 1$, the second term on the right drops out because $F(-\infty - \mathbf{x}\beta) = 0$, and for $y = J$, the first term equals $F(\infty - \mathbf{x}\beta) = 1$.

Comparing these equations with those for the BRM shows that the ORM is identical to the binary regression model, with one exception. To show this, we fit chapter 4's binary model for labor-force participation using both logit and ologit (the command for ordinal logit):

```
. use http://www.stata-press.com/data/lf2/binlfp2, clear
(Data from 1976 PSID-T Mroz)
. logit lfp k5 k618 age wc hc lwg inc, nolog
(output omitted)
. estimates store logit
. ologit lfp k5 k618 age wc hc lwg inc, nolog
(output omitted)
. estimates store ologit
```

To compare the coefficients, we combine them using `estimates table`, which leads to the following table:¹

```
. estimates table logit ologit, b(%9.3f) t label varwidth(30) equations(1:1)
```

	Variable	logit	ologit
#1	# kids < 6	-1.463	-1.463
		-7.43	-7.43
	# kids 6-18	-0.065	-0.065
		-0.95	-0.95
	Wife's age in years	-0.063	-0.063
		-4.92	-4.92
	Wife College: 1=yes 0=no	0.807	0.807
		3.51	3.51
	Husband College: 1=yes 0=no	0.112	0.112
		0.54	0.54
	Log of wife's estimated wages	0.605	0.605
		4.01	4.01
	Family income excluding wife's	-0.034	-0.034
	-4.20	-4.20	
	Constant	3.182	
		4.94	
cut1	Constant		-3.182
			-4.94

legend: b/t

1. Because `logit` has a constant and `ologit` has a cutpoint, by default `estimates table` will not line up the coefficients from the two models. Rather, each of the independent variables will be listed twice. `equations(1:1)` tells `estimates table` to line up the coefficients. This is easiest to understand if you try our command without the `equations(1:1)` option.

The slope coefficients and their standard errors are identical, but for logit an intercept is reported (i.e., the coefficient associated with `_cons`), whereas for `ologit` the constant is replaced by the cutpoint labeled `/cut1`, which is equal but of opposite sign.

This difference is due to how the two models are identified. As the ORM has been presented, there are "too many" free parameters; that is, you cannot estimate $J - 1$ thresholds and the constant, too. For a unique set of ML estimates to exist, an identifying assumption needs to be made about either the intercept or one of the cutpoints. In Stata, the ORM is identified by assuming that the intercept is 0, and the values of all cutpoints are estimated. Some statistics packages for the ORM instead fix one of the cutpoints to 0 and estimate the intercept. In presenting the BRM, we immediately assumed that the value that divided y^* into observed 0s and 1s was 0. In effect, we identified the model by assuming a threshold of 0. Although different parameterizations can be confusing, keep in mind that the slope coefficients and predicted probabilities are the same under either parameterization (see Long 1997, 122-23 for more details).

5.1.2 A nonlinear probability model

The ordinal regression model can also be developed as a nonlinear probability model without appealing to the idea of a latent variable. Here we show how this can be done for the ordinal logit model. First, we define the odds that an outcome is less than or equal to m versus greater than m given \mathbf{x} :

$$\Omega_{\leq m | > m}(\mathbf{x}) \equiv \frac{\Pr(y \leq m | \mathbf{x})}{\Pr(y > m | \mathbf{x})} \quad \text{for } m = 1, J - 1$$

For example, we could compute the odds of disagreeing or strongly disagreeing (i.e., $m \leq 2$) versus agreeing or strongly agreeing ($m > 2$). The log of the odds is assumed to equal

$$\ln \Omega_{\leq m | > m}(\mathbf{x}) = \tau_m - \mathbf{x}\beta \quad (5.2)$$

For one independent variable and three categories (where we are fixing the intercept to equal 0),

$$\ln \frac{\Pr(y \leq 1 | \mathbf{x})}{\Pr(y > 1 | \mathbf{x})} = \tau_1 - \beta_1 x_1$$

$$\ln \frac{\Pr(y \leq 2 | \mathbf{x})}{\Pr(y > 2 | \mathbf{x})} = \tau_2 - \beta_1 x_1$$

Although it may seem confusing that the model subtracts βx rather than adding it, this is a consequence of computing the logit of $y \leq m$ versus $y > m$. While we agree that it would be simpler to stick with $\tau_m + \beta x$, this is not the way the model is normally presented.

5.2 Estimation using ologit and oprobit

The ordered logit and probit models can be fitted with the following commands and their basic options:

```
ologit devar [indepvars] [if] [in] [weight] [, robust cluster(varname)
      level(#) nolog]
```

```
oprobit devar [indepvars] [if] [in] [weight] [, robust cluster(varname)
      level(#) nolog]
```

In our experience, these models take more steps to converge than the models for either binary or nominal outcomes.

Variable lists

devar is the dependent variable. The specific values assigned to the outcome categories are irrelevant, except that larger values are assumed to correspond to “higher” outcomes. For example, if you had three outcomes, you could use the values 1, 2, and 3, or -1.23, 2.3, and 999. Up to 50 outcomes are allowed in Intercooled Stata and Stata/SE; 20 outcomes are allowed in Small Stata.

indepvars is a list of independent variables. If *indepvars* is not included, Stata fits a model with only cutpoints.

Specifying the estimation sample

if and *in* qualifiers can be used to restrict the estimation sample. For example, if you want to fit an ordered logit model for only those in the 1989 sample, you could specify `ologit warm age ed prst male white if yr89==1`.

Listwise deletion Stata excludes cases in which there are missing values for any of the variables in the model. Accordingly, if two models are fitted using the same dataset but have different sets of independent variables, it is possible to have different samples. We recommend that you use `mark` and `markout` (discussed in chapter 3) to explicitly remove cases with missing data.

Weights

Both `ologit` and `oprobit` can be used with `fweights`, `pweights`, and `iweights`. See chapter 3 for more details.

Options

`robust` indicates that robust variance estimates are to be used. When `cluster()` is specified, robust standard errors are automatically used. See chapter 3 for more details.

`cluster(varname)` specifies that the observations are independent across the groups specified by unique values of *varname* but not necessarily within the groups. See chapter 3 for further details.

`level(#)` specifies the level of the confidence interval for estimated parameters. By default, Stata uses a 95% interval. You can also change the default level, say, to a 90% interval, with the command `set level 90`.

`nolog` suppresses the iteration history.

5.2.1 Example of attitudes toward working mothers

Our example is based on a question from the 1977 and 1989 General Social Survey. As we have already described, respondents were asked to evaluate the following statement: “A working mother can establish just as warm and secure of a relationship with her child as a mother who does not work”. Responses were coded as: 1 = Strongly Disagree (SD), 2 = Disagree (D), 3 = Agree (A), and 4 = Strongly Agree (SA). A complete description of the data can be obtained by using `describe`, `summarize`, and `tabulate`:

```
. use http://www.stata-press.com/data/lf2/ordwarm2
(77 & 89 General Social Survey)
. describe warm yr89 male white age ed prst
```

variable name	storage type	display format	value label	variable label
warm	byte	%10.0g	SD2SA	Mom can have warm relations with child
yr89	byte	%10.0g	yr1bl	Survey year: 1=1989 0=1977
male	byte	%10.0g	sex1bl	Gender: 1=male 0=female
white	byte	%10.0g	race21bl	Race: 1=white 0=not white
age	byte	%10.0g		Age in years
ed	byte	%10.0g		Years of education
prst	byte	%10.0g		Occupational prestige

```
. summarize warm yr89 male white age ed prst
```

Variable	Obs	Mean	Std. Dev.	Min	Max
warm	2293	2.607501	.9282156	1	4
yr89	2293	.3986044	.4897178	0	1
male	2293	.4648932	.4988748	0	1
white	2293	.8765809	.3289894	0	1
age	2293	44.93546	16.77903	18	89
ed	2293	12.21805	3.160827	0	20
prst	2293	39.58526	14.49226	12	82

. tabulate warm

Mom can have warm relations with child	Freq.	Percent	Cum.
SD	297	12.95	12.95
D	723	31.53	44.48
A	856	37.33	81.81
SA	417	18.19	100.00
Total	2,293	100.00	

Using these data, we fitted the model

$$\Pr(\text{warm} = m | x_i) = F(\tau_m - x\beta) - F(\tau_{m-1} - x\beta)$$

where

$$x\beta = \beta_{\text{yr89}}\text{yr89} + \beta_{\text{male}}\text{male} + \beta_{\text{white}}\text{white} + \beta_{\text{age}}\text{age} + \beta_{\text{ed}}\text{ed} + \beta_{\text{prst}}\text{prst}$$

Here are the results from ologit and oprobit. We store each model with estimates store so that we can later make a table combining the results:

. ologit warm yr89 male white age ed prst, nolog

Ordered logistic regression Number of obs = 2293
 LR chi2(6) = 301.72
 Prob > chi2 = 0.0000
 Pseudo R2 = 0.0504
 Log likelihood = -2844.9123

warm	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
yr89	.5239025	.0798988	6.56	0.000	.3673037 .6805013
male	-.7332997	.0784827	-9.34	0.000	-.8871229 -.5794766
white	-.3911595	.1183808	-3.30	0.001	-.6231815 -.1591374
age	-.0216655	.0024683	-8.78	0.000	-.0265032 -.0168278
ed	.0671728	.015975	4.20	0.000	.0358624 .0984831
prst	.0060727	.0032929	1.84	0.065	-.0003813 .0125267
/cut1	-2.465362	.2389126			-2.933622 -1.997102
/cut2	-.630904	.2333155			-1.088194 -.173614
/cut3	1.261854	.2340179			.8031873 1.720521

. estimates store ologit

Constant

. oprobit warm yr89 male white age ed prst, nolog

Ordered probit regression Number of obs = 2293
 LR chi2(6) = 294.32
 Prob > chi2 = 0.0000
 Pseudo R2 = 0.0491
 Log likelihood = -2848.611

warm	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
yr89	.3188147	.0468519	6.80	0.000	.2269867 .4106427
male	-.4170287	.0455459	-9.16	0.000	-.5062971 -.3277603
white	-.2265002	.0694773	-3.26	0.001	-.3626733 -.0903272
age	-.0122213	.0014427	-8.47	0.000	-.0150489 -.0093937
ed	.0387234	.0093241	4.15	0.000	.0204485 .0569982
prst	.003283	.001925	1.71	0.088	-.0004899 .0070559
/cut1	-1.428578	.1387742			-1.700571 -1.156586
/cut2	-.3605589	.1369219			-.6289209 -.092197
/cut3	.7681637	.1370564			.4995381 1.036789

. estimates store oprobit

The information in the header and the table of coefficients is in the same form as discussed in chapter 3. Since we stored the results for both models, we can compare the results using estimates table:

. estimates table ologit oprobit, b(%9.3f) t label varwidth(30)

Variable	ologit	oprobit
warm		
Survey year: 1=1989 0=1977	0.524	0.319
	6.56	6.80
Gender: 1=male 0=female	-0.733	-0.417
	-9.34	-9.16
Race: 1=white 0=not white	-0.391	-0.227
	-3.30	-3.26
Age in years	-0.022	-0.012
	-8.78	-8.47
Years of education	0.067	0.039
	4.20	4.15
Occupational prestige	0.006	0.003
	1.84	1.71
cut1		
Constant	-2.465	-1.429
	-10.32	-10.29
cut2		
Constant	-0.631	-0.361
	-2.70	-2.63
cut3		
Constant	1.262	0.768
	5.39	5.60

legend: b/t

As with the BRM, the estimated coefficients differ from logit to probit by a factor of about 1.7, reflecting the differing scaling of the ordered logit and ordered probit models. Values of the z -tests are very similar because they are not affected by the scaling, but they are not identical because of slight differences in the shape of the assumed distribution of the errors.

5.2.2 Predicting perfectly

If the dependent variable does not vary within one of the categories of an independent variable, there will be a problem with estimation. To see what happens, let's transform the prestige variable `prst` into a dummy variable:

```
. gen dumprst = (prst<20 & warm==1)
. tab dumprst warm, miss
```

dumprst	Mom can have warm relations with child				Total
	SD	D	A	SA	
0	257	723	856	417	2,253
1	40	0	0	0	40
Total	297	723	856	417	2,293

In all cases where `dumprst` is 1, respondents have values of `SD` for `warm`. That is, if you know `dumprst` is 1, you can predict perfectly that `warm` is 1 (i.e., `SD`). Although we purposely constructed `dumprst` so this would happen, perfect prediction can also occur in real data. If we fit the ORM using `dumprst` rather than `prst`, the perfectly predicted observations are dropped from the estimation sample.

```
. ologit warm yr89 male white age ed dumprst, nolog
Ordered logistic regression      Number of obs =      2293
                                LR chi2(6) =      447.02
                                Prob > chi2 =      0.0000
Log likelihood = -2772.2621      Pseudo R2 =      0.0746
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
warm					
yr89	.5268578	.0805997	6.54	0.000	.3688853 .6848303
male	-.7251825	.0792896	-9.15	0.000	-.8805872 -.5697778
white	-.4240687	.1197416	-3.54	0.000	-.658758 -.1893795
age	-.0210592	.0024462	-8.61	0.000	-.0258536 -.0162648
ed	.072143	.0133133	5.42	0.000	.0460494 .0982366
dumprst	-34.58373	1934739	-0.00	1.000	-3792053 3791983
/cut1	-2.776233	.243582			-3.253645 -2.298822
/cut2	-.8422903	.2363736			-1.305574 -.3790065
/cut3	1.06148	.236561			.5978287 1.525131

Note: 40 observations completely determined. Standard errors questionable.

The note at the bottom of the output above indicates the problem. In practice, the next step would be to delete the 40 cases in which `dumprst` equals 1 (you could use

the command `drop if dumprst==1` to do this) and refit the model without `dumprst`. This corresponds to what is done automatically for binary models fitted by `logit` and `probit`.

5.3 Hypothesis testing with `test` and `lrtest`

Hypothesis tests of regression coefficients can be evaluated with the z -statistics in the estimation output, with `test` for Wald tests of simple and complex hypotheses, and with `lrtest` for the corresponding likelihood-ratio tests. We will briefly review each.

5.3.1 Testing individual coefficients

If the assumptions of the model hold, the ML estimators from `ologit` and `oprobit` are distributed asymptotically normally. The hypothesis $H_0: \beta_k = \beta^*$ can be tested with $z = (\hat{\beta}_k - \beta^*) / \hat{\sigma}_{\hat{\beta}_k}$. Under the assumptions justifying ML, if H_0 is true, then z is distributed approximately normally with a mean of 0 and a variance of 1 for large samples. For example, consider the results for the variable `male` from the `ologit` output above:

```
. ologit warm male yr89 white age ed prst, nolog
(output omitted)
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
warm					
male	-.7332997	.0784827	-9.34	0.000	-.8871229 -.5794766

(output omitted)

We conclude that sex significantly affects attitudes toward working mothers ($z = -9.34$, $p < 0.01$ for a two-tailed test).

Either a one-tailed or a two-tailed test can be used as discussed in chapter 4.

The z -test in the output of estimation commands is a Wald test, which can also be computed using `test`. For example, to test $H_0: \beta_{\text{male}} = 0$,

```
. test male
(1) [warm]male = 0
      chi2( 1) = 87.30
      Prob > chi2 = 0.0000
```

We conclude that sex significantly affects attitudes toward working mothers ($X^2 = 87.30$, $df = 1$, $p < 0.01$).

The value of a chi-squared test with 1 degree of freedom is identical to the square of the corresponding z -test, which can be demonstrated with the `display` command:

```
. display "z*z=" -9.343*-9.343
z*z=87.291649
```

Another way to verify this is to use the information that `test` leaves in memory. After `test`, we type

```
. return list
scalars:
  r(drop) = 0
  r(chi2) = 87.30028866314404
  r(df) = 1
  r(p) = 9.32342950928e-21
```

The return `r(chi2)` contains the value of the chi-squared test statistics. We can take the square root of this quantity to confirm the relationship between the z -test and the chi-squared test.

```
. di "chi2=" r(chi2) "; sqrt(chi2)= " sqrt(r(chi2))
chi2=87.300289; sqrt(chi2)= 9.3434623
```

An LR test is computed by comparing the log likelihood from a full model with that of a restricted model. To test one coefficient, we begin by fitting the full model:

```
. ologit warm yr89 male white age ed prst, nolog
Ordered logistic regression      Number of obs   =      2293
                                LR chi2(6)          =      301.72
                                Prob > chi2         =      0.0000
                                Pseudo R2          =      0.0504
Log likelihood = -2844.9123
(output omitted)
. estimates store fmodel
```

Then we fit the model, excluding male:

```
. ologit warm yr89 white age ed prst, nolog
Ordered logistic regression      Number of obs   =      2293
                                LR chi2(5)          =      212.98
                                Prob > chi2         =      0.0000
                                Pseudo R2          =      0.0355
Log likelihood = -2889.278
(output omitted)
. estimates store nmodel
. lrtest fmodel nmodel
Likelihood-ratio test          LR chi2(1)      =      88.73
(Assumption: nmodel nested in fmodel) Prob > chi2    =      0.0000
```

The resulting LR test can be interpreted to mean that the effect of being male is significant at the .01 level ($LRX^2 = 88.73$, $df = 1$, $p < .01$).

5.3.2 Testing multiple coefficients

We can also test a complex hypothesis that involves more than one coefficient. For example, our model has three demographic variables: `age`, `white`, and `male`. To test that all the demographic factors are simultaneously equal to zero, $H_0: \beta_{age} = \beta_{white} = \beta_{male} = 0$, we can use either a Wald or an LR test. For the Wald test, we fit the full model as before and then type

```
. test age white male
( 1) [warm]age = 0
( 2) [warm]white = 0
( 3) [warm]male = 0
      chi2( 3) = 166.62
      Prob > chi2 = 0.0000
```

We conclude that the hypothesis that the demographic effects of age, race, and sex are simultaneously equal to zero can be rejected at the .01 level ($X^2 = 166.62$, $df = 3$, $p < .01$).

`test` can also be used to test the equality of effects as shown in chapter 4.

To compute an LR test of multiple coefficients, we first fit the full model and save the results with `lrtest`, `saving(0)`. Then to test $H_0: \beta_{age} = \beta_{white} = \beta_{male} = 0$, we fit the model that excludes these three variables and run `lrtest`:

```
. ologit warm yr89 male white age ed prst, nolog
(output omitted)
. estimates store fmodel
. ologit warm yr89 ed prst, nolog
(output omitted)
. estimates store nmodel
. lrtest fmodel nmodel
Likelihood-ratio test          LR chi2(3)      =      171.58
(Assumption: nmodel nested in fmodel) Prob > chi2    =      0.0000
```

We conclude that the hypothesis that the demographic effects of age, race, and sex are simultaneously equal to zero can be rejected at the .01 level ($X^2 = 171.58$, $df = 3$, $p < .01$).

We find that the Wald and LR tests usually lead to the same decisions. When there are differences, they generally occur when the tests are near the cutoff for statistical significance. Because the LR test is invariant to reparameterization, we prefer the LR test.

5.4 Scalar measures of fit using `fitstat`

As we discuss more in chapter 3, scalar measures of fit can be useful in comparing competing models (see also Long 1997, 85–113). Several different measures can be computed after either `ologit` or `oprobit` with the `SPost` command `fitstat`:

```

. ologit warm yr89 male white age ed prst, nolog
(output omitted)

. fitstat

Measures of Fit for ologit of warm
Log-Lik Intercept Only:    -2995.770   Log-Lik Full Model:    -2844.912
D(2284):                  5689.825   LR(6):                 301.716
                          Prob > LR:         0.000
McFadden's R2:           0.050   McFadden's Adj R2:     0.047
ML (Cox-Snell) R2:       0.123   Cragg-Uhler(Nagelkerke) R2: 0.133
McKelvey & Zavoina's R2: 0.127
Variance of y*:          3.768   Variance of error:     3.290
Count R2:                 0.432   Adj Count R2:          0.093
AIC:                      2.489   AIC*n:                 5707.825
BIC:                      -11982.891  BIC':                  -255.291
BIC used by Stata:        5759.463   AIC used by Stata:     5707.825

```

Using simulations, both Hagle and Mitchell (1992) and Windmeijer (1995) find that, for ordinal outcomes, McKelvey and Zavoina's R^2 most closely approximates the R^2 obtained by fitting the linear regression model on the underlying latent variable.

5.5 Converting to a different parameterization*

Earlier, we noted that different software packages use different parameterizations to identify the model. Stata sets $\beta_0 = 0$ and estimates τ_1 , whereas some programs fix $\tau_1 = 0$ and estimate β_0 . Although all quantities of interest for the purpose of interpretation (e.g., predicted probabilities) are the same under both parameterizations, it is useful to see how Stata can fit the model under either parameterization. The key to understanding how this is done is the equation

$$\Pr(y = m | \mathbf{x}) = F([\tau_m - \delta] - [\beta_0 - \delta] - \mathbf{x}\beta) - F([\tau_{m-1} - \delta] - [\beta_0 - \delta] - \mathbf{x}\beta)$$

Without further constraints, it is possible to estimate only the differences $\tau_m - \delta$ and $\beta_0 - \delta$. Stata assumes $\delta = \beta_0$, which forces the estimate of β_0 to be 0, whereas some other programs assume $\delta = \tau_1$, which forces the estimate of τ_1 to be 0. For example,

Model parameter	Stata's estimate	Alternative parameterization
β_0	$\beta_0 - \beta_0 = 0$	$\beta_0 - \tau_1$
τ_1	$\tau_1 - \beta_0$	$\tau_1 - \tau_1 = 0$
τ_2	$\tau_2 - \beta_0$	$\tau_2 - \tau_1$
τ_3	$\tau_3 - \beta_0$	$\tau_3 - \tau_1$

Although you would only need to compute the alternative parameterization if you wanted to compare your results with those produced by another statistics package, seeing how this is done illustrates why the intercept and thresholds are arbitrary. To

estimate the alternative parameterization, we use `lincom` to estimate the difference between Stata's estimates (see page 190) and the estimated value of the first cutpoint:²

```

. ologit warm yr89 male white age ed prst, nolog
(output omitted)

. * intercept
. lincom 0 - _b[/cut1]
( 1) - [cut1]_cons = 0

```

warm	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
(1)	2.465362	.2389126	10.32	0.000	1.997102 2.933622

Here we are computing the alternative parameterization of the intercept. `ologit` assumes that $\beta_0 = 0$, so we simply estimate $0 - \tau_1$; that is, $0 - \text{b}[/\text{cut1}]$. The trick is that the cutpoints are contained in the vector `_b[]`; using the labels `/cut1`, `/cut2`, and `/cut3`. For the thresholds, we are estimating $\tau_2 - \tau_1$ and $\tau_3 - \tau_1$, which correspond to `_b[/cut2] - _b[/cut1]` and `_b[/cut3] - _b[/cut1]`:

```

. * cutpoint 2
. lincom _b[/cut2] - _b[/cut1]
( 1) - [cut1]_cons + [cut2]_cons = 0

```

warm	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
(1)	1.834458	.0630432	29.10	0.000	1.710895 1.95802

```

. * cutpoint 3
. lincom _b[/cut3] - _b[/cut1]
( 1) - [cut1]_cons + [cut3]_cons = 0

```

warm	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
(1)	3.727216	.0826215	45.11	0.000	3.565281 3.889151

The estimate of $\tau_1 - \tau_1$ is, of course, 0.

5.6 The parallel regression assumption

Before discussing interpretation, it is important to understand an assumption that is implicit in the ORM, known as both the *parallel regression assumption* and, for the ordinal logit model, the *proportional odds assumption*. Using (5.1), the ORM can be written as

2. Prior to version 9, Stata labeled the cutpoints as, for example, `_cut1` instead of the newer `/cut1`. If you are using a version of Stata prior to version 9, simply replace the `/s` with `_s` in the commands in this chapter.

$$\begin{aligned}\Pr(y = 1 | \mathbf{x}) &= F(\tau_m - \mathbf{x}\beta) \\ \Pr(y = m | \mathbf{x}) &= F(\tau_m - \mathbf{x}\beta) - F(\tau_{m-1} - \mathbf{x}\beta) \quad \text{for } m = 2 \text{ to } J - 1 \\ \Pr(y = J | \mathbf{x}) &= 1 - F(\tau_{m-1} - \mathbf{x}\beta)\end{aligned}$$

These equations can be used to compute the cumulative probabilities, which have the simple form

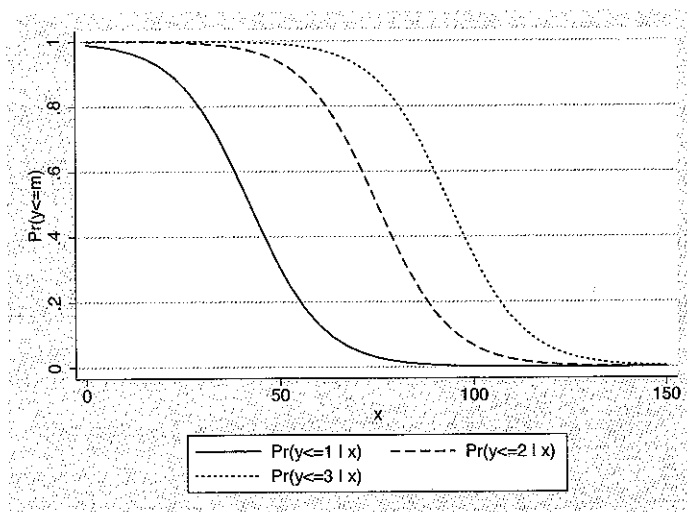
$$\Pr(y \leq m | \mathbf{x}) = F(\tau_m - \mathbf{x}\beta) \quad \text{for } m = 1 \text{ to } J - 1$$

This equation shows that the ORM is equivalent to $J - 1$ binary regressions with the critical assumption that the slope coefficients are identical across each regression.

For example, with four outcomes and one independent variable, the equations are

$$\begin{aligned}\Pr(y \leq 1 | \mathbf{x}) &= F(\tau_1 - \beta x) \\ \Pr(y \leq 2 | \mathbf{x}) &= F(\tau_2 - \beta x) \\ \Pr(y \leq 3 | \mathbf{x}) &= F(\tau_3 - \beta x)\end{aligned}$$

The intercept α is not in the equation since it has been assumed to equal 0 to identify the model. These equations lead to the following figure:³



Each probability curve differs *only* in being shifted to the left or right. That is, they are parallel as a consequence of the assumption that the β s are equal for each equation.

3. This plot illustrates how `graph` can be used to construct graphs that are not based on real data. The commands for this graph are contained in `st9ch5ordinal.do`, which is part of the package `spost9.do`. See chapter 1 for details.

This figure suggests that the parallel regression assumption can be tested by comparing the estimate from the $J - 1$ binary regressions,

$$\Pr(y \leq m | \mathbf{x}) = F(\tau_m - \mathbf{x}\beta_m) \quad \text{for } m = 1, J - 1$$

where the β s are allowed to differ across the equations. The parallel regression assumption implies that $\beta_1 = \beta_2 = \dots = \beta_{J-1}$. To the degree that the parallel regression assumption holds, the coefficients $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_{J-1}$ should be close. There are two commands in Stata that perform this test.

An approximate LR test

The command `omodel` (Wolfe and Gould 1998) is not part of official Stata but can be obtained by typing `net search omodel` and following the prompts. `omodel` computes an approximate LR test. Essentially, this method compares the log likelihood from `ologit` (or `oprobit`) with that obtained from pooling $J - 1$ binary models fitted with `logit` (or `probit`), making an adjustment for the correlation between the binary outcomes defined by $y \leq m$. The syntax is

```
omodel [logit|probit] depvar [indepvars] [if] [in] [weight]
```

where the subcommand `logit` or `probit` indicates whether ordered logit or ordered probit is to be used. For example,

```
. omodel logit warm yr89 male white age ed prst
  (same output as for ologit warm yr89 male white age ed prst)
Approximate likelihood-ratio test of proportionality of odds
across response categories:
      chi2(12) =    48.91
      Prob > chi2 =    0.0000
```

Here the parallel regression assumption can be rejected at the .01 level.

A Wald test

The LR test is an omnibus test that the coefficients for all variables are simultaneously equal. Accordingly, you cannot determine whether the coefficients for some variables are identical across the binary equations while coefficients for other variables differ. To this end, a Wald test by Brant (1990) is useful since it tests the parallel regression assumption for each variable. The messy details of computing this test are found in Brant (1990) or Long (1997, 143–144). In Stata, the test is computed quickly with `brant`, which is part of `SPost`. After running `ologit` (`brant` does not work with `oprobit`), you run `brant` with the syntax:

```
brant [, detail]
```

The `detail` option provides a table of coefficients from each of the binary models. For example,

```
. brant, detail
Estimated coefficients from j-1 binary regressions
      yr89      male      white      age      ed      prst      _cons
      y>1      y>2      y>3
      .9647422  -.56540626  .31907316
      -.30536425  -.69054232  -1.0837888
      -.55265759  -.31427081  -.39299842
      -.0164704  -.02533448  -.01859051
      .10479624  .05285265  .05755466
      -.00141118  .00953216  .00553043
      1.8584045  .73032873  -1.0245168
```

Brant Test of Parallel Regression Assumption

Variable	chi2	p>chi2	df
All	49.18	0.000	12
yr89	13.01	0.001	2
male	22.24	0.000	2
white	1.27	0.531	2
age	7.38	0.025	2
ed	4.31	0.116	2
prst	4.33	0.115	2

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

The chi-squared of 49.18 for the Brant test is close to the value of 48.91 from the LR test. However, the Brant test shows that the largest violations are for `yr89` and `male`, which indicates that there may be problems related to these variables.

Caveat regarding the parallel regression assumption We find that the parallel regression assumption is frequently violated. When the assumption of parallel regressions is rejected, alternative models that do not impose the constraint of parallel regressions should be considered. Violation of the parallel regression assumption is not a rationale for using ordinary least squares regression since the assumptions implied by the application of the LRM to ordinal data are even stronger. Alternative models that can be considered include models for nominal outcomes discussed in chapter 6 or other models for ordinal outcomes discussed in section 5.9.

5.7 Residuals and outliers using `predict`

Although no methods for detecting influential observations and outliers have been developed specifically for the ORM, Hosmer and Lemeshow (2000, 305) suggest applying the methods for binary models to the $J - 1$ cumulative probabilities that were discussed in

the last section. As noted by Hosmer and Lemeshow, the disadvantage of this approach is that you are evaluating only an approximation to the model you have fitted, because the coefficients of the binary models differ from those fitted in the ordinal model. But, if the parallel regression assumption is *not* rejected, you can be more confident in the results of your residual analysis.

To illustrate this approach, we start by generating three binary variables corresponding to `warm < 2`, `warm < 3`, and `warm < 4`:

```
. gen warmlt2 = (warm<2) if warm <.
. gen warmlt3 = (warm<3) if warm <.
. gen warmlt4 = (warm<4) if warm <.
```

For example, `warmlt3` is 1 if `warm` equals 1 or 2, else 0. Next we estimate binary logits for `warmlt2`, `warmlt3`, and `warmlt4` using the same independent variables as in our original `ologit` model. After estimating each logit, we generate standardized residuals using `predict` (for a detailed discussion of generating and inspecting these residuals, see chapter 4):

```
* warm < 2
. logit warmlt2 yr89 male white age ed prst
(output omitted)
. predict rstd_lt2, rs

* warm < 3
. logit warmlt3 yr89 male white age ed prst
(output omitted)
. predict rstd_lt3, rs

* warm < 4
. logit warmlt4 yr89 male white age ed prst
(output omitted)
. predict rstd_lt4, rs
```

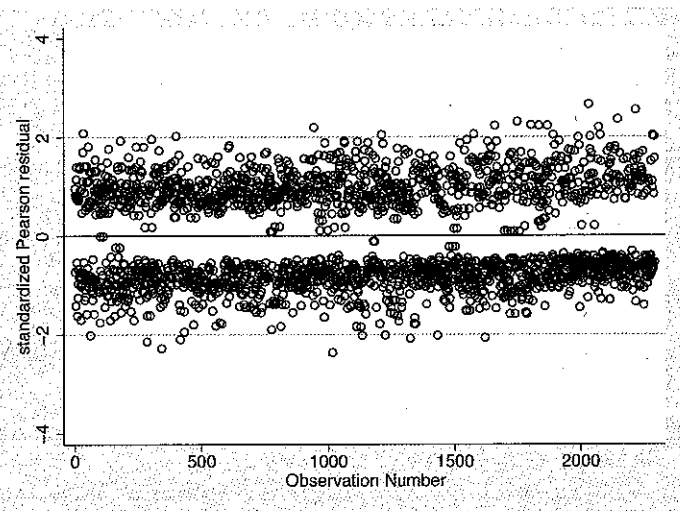
Next we create an index plot for each of the three binary equations. Using the results from the logit of `warmlt3` yields the following graph:

(Continued on next page)

```

. sort prst
. gen index = _n
. graph twoway scatter rstd_lt3 index, yline(0) ylabel(-4(2)4) ///
> xtitle("Observation Number") xlabel(0(500)2293) ///
> msymbol(Oh)

```



Given the size of the dataset, no residual stands out as being especially large. See section 4.4 for other ways in which you can examine outliers and influential cases.

5.8 Interpretation

If the idea of a latent variable makes substantive sense, simple interpretations are possible by rescaling y^* to compute standardized coefficients that can be used just like coefficients for the linear regression model. If the focus is on the categories of the ordinal variable (e.g., what affects the likelihood of strongly agreeing), the methods illustrated for the BRM can be extended to multiple outcomes. Because the ORM is nonlinear in the outcome probabilities, no approach can fully describe the relationship between a variable and the outcome probabilities. Consequently, you should consider each of these methods before deciding which approach is most effective in your application. For purposes of illustration, we continue to use the example of attitudes toward working mothers. Remember that the test of the parallel regression assumption suggests that this model is not appropriate for these data.

5.8.1 Marginal change in y^*

In the ORM, $y^* = \mathbf{x}\beta + \varepsilon$, and the marginal change in y^* with respect to x_k is

$$\frac{\partial y^*}{\partial x_k} = \beta_k$$

Because y^* is latent (and hence its metric is unknown), the marginal change cannot be interpreted without standardizing by the estimated standard deviation of y^* ,

$$\widehat{\sigma}_{y^*}^2 = \widehat{\beta}' \widehat{\text{Var}}(\mathbf{x}) \widehat{\beta} + \text{Var}(\varepsilon)$$

where $\widehat{\text{Var}}(\mathbf{x})$ is the covariance matrix for the observed x s, $\widehat{\beta}$ contains ML estimates, and $\text{Var}(\varepsilon) = 1$ for ordered probit and $\pi^2/3$ for ordered logit. Then the y^* -standardized coefficient for x_k is

$$\beta_k^{S_{y^*}} = \frac{\beta_k}{\sigma_{y^*}}$$

which can be interpreted that for a unit increase in x_k , y^* is expected to increase by $\beta_k^{S_{y^*}}$ standard deviations, holding all other variables constant.

The fully standardized coefficient is

$$\beta_k^S = \frac{\sigma_k \beta_k}{\sigma_{y^*}} = \sigma_k \beta_k^{S_{y^*}}$$

which can be interpreted that for a standard deviation increase in x_k , y^* is expected to increase by β_k^S standard deviations, holding all other variables constant.

These coefficients can be computed with `listcoef` using the `std` option. For example, after fitting the ordered logit model,

```

. listcoef, std help
ologit (N=2293): Unstandardized and Standardized Estimates
Observed SD: .9282156
Latent SD: 1.9410634

```

	warm	b	z	P> z	bStdX	bStdY	bStdXY	SDofX
yr89		0.52390	6.557	0.000	0.2566	0.2699	0.1322	0.4897
male		-0.73330	-9.343	0.000	-0.3658	-0.3778	-0.1885	0.4989
white		-0.39116	-3.304	0.001	-0.1287	-0.2015	-0.0663	0.3290
age		-0.02167	-8.778	0.000	-0.3635	-0.0112	-0.1873	16.7790
ed		0.06717	4.205	0.000	0.2123	0.0346	0.1094	3.1608
prst		0.00607	1.844	0.065	0.0880	0.0031	0.0453	14.4923

```

b = raw coefficient
z = z-score for test of b=0
P>|z| = p-value for z-test
bStdX = x-standardized coefficient
bStdY = y-standardized coefficient
bStdXY = fully standardized coefficient
SDofX = standard deviation of X

```

If we think of the dependent variable as measuring support for mothers in the workplace, then the effect of the year of the interview can be interpreted as indicating that in 1989 support was .27 standard deviations higher than in 1977, holding all other variables constant.

To consider the effect of education, each standard deviation increase in education increases support by .11 standard deviations, holding all other variables constant.

5.8.2 Predicted probabilities

We usually prefer interpretations based somehow on predicted probabilities. These probabilities can be estimated with the formula

$$\widehat{\Pr}(y = m | \mathbf{x}) = F(\hat{\tau}_m - \mathbf{x}\hat{\beta}) - F(\hat{\tau}_{m-1} - \mathbf{x}\hat{\beta})$$

with cumulative probabilities computed as

$$\widehat{\Pr}(y \leq m | \mathbf{x}) = F(\tau_m - \mathbf{x}\hat{\beta})$$

The values of \mathbf{x} can be based on observations in the sample or can be hypothetical values of interest. The most basic command for computing probabilities is `predict`, but our `SPost` commands can be used to compute predicted probabilities in particularly useful ways.

5.8.3 Predicted probabilities with predict

After fitting a model with `ologit` or `oprobit`, a useful first step is to compute the in-sample predictions with the command

```
predict newvar1 [newvar2[newvar3...]] [if] [in]
```

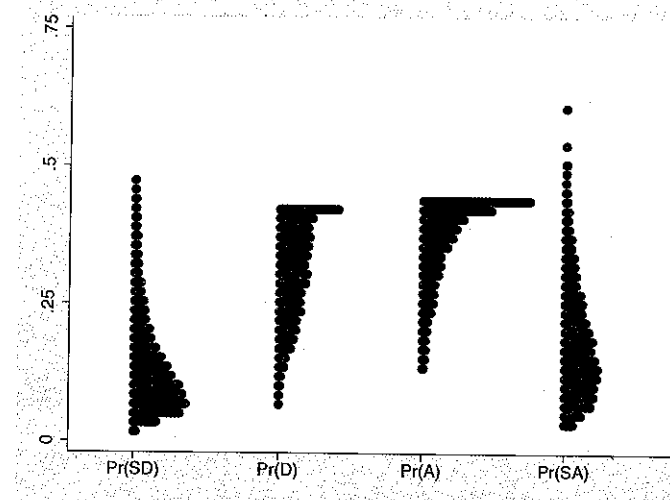
where you indicate one new variable name for each category of the dependent variable. For instance, in the following example `predict` specifies that the variables `SDwarm`, `Dwarm`, `Awarm`, and `SAwarm` should be created with predicted values for the four outcome categories:

```
. ologit warm yr89 male white age ed prst, nolog
  (output omitted)
. predict SDlogit Dlogit Alogit SAllogit
  (option p assumed; predicted probabilities)
```

The message (option p assumed; predicted probabilities) reflects that `predict` can compute many different quantities. Because we did not specify an option indicating which quantity to predict, option `p` for predicted probabilities was assumed.

An easy way to see the range of the predictions is with `dotplot`, one of our favorite commands for quickly checking data:

```
. label var SDwarm "Pr(SD)"
. label var Dwarm "Pr(D)"
. label var Awarm "Pr(A)"
. label var SAwarm "Pr(SA)"
. dotplot SDwarm Dwarm Awarm SAwarm, ylabel(0(.25).75)
```



The predicted probabilities for the extreme categories tend to be less than .25, with most predictions for the middle categories falling between .25 and .5. In only a few cases is the probability of any outcome greater than .5.

Examining predicted probabilities within the sample provides a first, quick check of the model. To understand and present the substantive findings, however, it is usually more effective to compute predictions at specific, substantively informative values. Our commands `prvalue`, `prtab`, `prgen`, and `prchange` are designed to make this simple.

5.8.4 Individual predicted probabilities with prvalue

Predicted probabilities for individuals with a particular set of characteristics can be computed with `prvalue`. For example, we might want to examine the predicted probabilities for individuals with the following characteristics:

- Working-class men in 1977 who are near retirement.
- Young, highly educated women with prestigious jobs.
- An “average” individual in 1977.
- An “average” individual in 1989.

Each of these can be easily computed with `prvalue` (see chapter 3 for a discussion of options for this command). The predicted probabilities for older, working-class men are

```
. prvalue, x(yr89=0 male=1 prst=20 age=64 ed=16) rest(mean)
ologit: Predictions for warm
Confidence intervals by delta method

          95% Conf. Interval
Pr(y=SD|x):  0.2317 [ 0.1776,  0.2857]
Pr(y=D|x):   0.4221 [ 0.3942,  0.4500]
Pr(y=A|x):   0.2723 [ 0.2249,  0.3198]
Pr(y=SA|x):  0.0739 [ 0.0523,  0.0954]

      yr89   male   white   age   ed   prst
x=      0       1   .8765809   64   16   20
```

For young, highly educated women with prestigious jobs, they are

```
. prvalue, x(yr89=1 male=0 prst=80 age=30 ed=24) rest(mean) brief
ologit: Predictions for warm

          95% Conf. Interval
Pr(y=SD|x):  0.0164 [ 0.0106,  0.0222]
Pr(y=D|x):   0.0781 [ 0.0554,  0.1008]
Pr(y=A|x):   0.3147 [ 0.2636,  0.3658]
Pr(y=SA|x):  0.5908 [ 0.5143,  0.6673]
```

and so on, for other sets of values.

We have set the values of the independent variables that define our hypothetical person using the `x()` and `rest()` options. The output from the first call of `prvalue` lists the values that have been set for all independent variables. This allows you to verify that `x()` and `rest()` did what you intended. For the second call, we added the `brief` option. This suppresses the output showing the levels of the independent variables. If you use this option, be certain that you have correctly specified the levels of all variables. Remember that the output of `prvalue` labels the categories according to the value labels assigned to the dependent variable. For example, $\text{Pr}(y=SD | x) = 0.2317$. As it is easy to be confused about the outcome categories when using these models, it is prudent to assign clear value labels to your dependent variable (see chapter 2).

We can summarize the results in a table that lists the ideal types and provides a clear indication of which variables are important:

Ideal type	Probability for outcome category:			
	SD	D	A	SA
Working-class men in 1977 who are near retirement	.23	.42	.27	.07
Young, highly educated women in 1989 with prestigious jobs	.02	.08	.32	.59
An "average individual" in 1977	.13	.36	.37	.14
An "average individual" in 1989	.08	.28	.43	.21

5.8.5 Tables of predicted probabilities with `prtab`

In other cases, it can be useful to compute predicted probabilities for all combinations of a set of categorical independent variables. For example, the ideal types illustrate the importance of sex and the year when the question was asked. Using `prtab`, we can easily show the degree to which these variables affect opinions for those average on other characteristics.

```
. prtab yr89 male, novarlbl
ologit: Predicted probabilities for warm
Predicted probability of outcome 1 (SD)
```

yr89	male	
	Women	Men
1977	0.0989	0.1859
1989	0.0610	0.1191

Predicted probability of outcome 2 (D)

yr89	male	
	Women	Men
1977	0.3083	0.4026
1989	0.2282	0.3394

Predicted probability of outcome 3 (A)

yr89	male	
	Women	Men
1977	0.4129	0.3162
1989	0.4406	0.3904

Predicted probability of outcome 4 (SA)

yr89	male	
	Women	Men
1977	0.1799	0.0953
1989	0.2703	0.1510

```
x=      yr89      male      white      age      ed      prst
      .39860445 .46489315 .8765809 44.935456 12.218055 39.585259
```

(tables for other outcomes omitted)

Tip Sometimes the output of `prtab` is clearer without the variable labels. These can be suppressed with the `novarlbl` option.

The output from `prtab` can be rearranged into a table that clearly shows that men are more likely than women to strongly disagree or disagree with the proposition that working mothers can have relationships with their children that are as warm as those of mothers who do not work. The table also shows that between 1977 and 1989 there was a movement for both men and women toward more positive attitudes about working mothers.

1977	SD	D	A	SA
Men	.19	.40	.32	.10
Women	.10	.31	.41	.18
Difference	.09	.09	-.09	-.08
1989	SD	D	A	SA
Men	.12	.34	.39	.15
Women	.06	.23	.44	.27
Difference	.06	.11	-.05	-.12
Change from 1977 to 1989				
	SD	D	A	SA
Men	-.07	-.06	.07	.05
Women	-.04	-.08	.03	.09

`prtab` does not include confidence intervals. If we want confidence intervals for the probabilities in the table, we need to make separate calls to `prvalue`. For the first row in the above table, we could compute the confidence intervals as follows:

```
. prvalue, x(yr89=0 male=1)
ologit: Predictions for warm
Confidence intervals by delta method

          Pr(y=SD|x):      0.1859 [ 0.1631, 0.2088]
          Pr(y=D|x):      0.4026 [ 0.3777, 0.4275]
          Pr(y=A|x):      0.3162 [ 0.2925, 0.3400]
          Pr(y=SA|x):     0.0953 [ 0.0814, 0.1092]

          yr89      male      white      age      ed      prst
x=          0          1      .8765809 44.935456 12.218055 39.585259
```

The confidence intervals for the other cells can be generated by the following commands:

```
. prvalue, x(yr89=0 male=0)
. prvalue, x(yr89=1 male=1)
. prvalue, x(yr89=1 male=0)
```

5.8.6 Graphing predicted probabilities with `prgen`

Graphing predicted probabilities for each outcome can also be useful for the ORM. Here we consider women in 1989 and show how predicted probabilities are affected by age. Of course, the plot could also be constructed for other sets of characteristics. The predicted probabilities as age ranges from 20 to 80 are generated by `prgen`:

```
. prgen age, from(20) to(80) generate(w89) x(male=0 yr89=1) ncases(13)
ologit: Predicted values as age varies from 20 to 80.

          yr89      male      white      age      ed      prst
x=          1          0      .8765809 44.935456 12.218055 39.585259
```

You should be familiar with how `x()` operates, but it is useful to review the other options:

`from(20)` and `to(80)` specify the minimum and maximum values over which `inc` is to vary. The default is the variable's minimum and maximum values.

`generate(w89)` is the root name for the new variables.

`ncases(13)` indicates that 13 evenly spaced values of `age` between 20 and 80 should be generated.

Here `w89x` contains values of `age` ranging from 20 to 80. The `p#` variables contain the predicted probability for outcome # (e.g., `w89p2` is the predicted probability of outcome 2). With ordinal outcomes, `prgen` also computes cumulative probabilities (i.e., summed) that are indicated by `s` (e.g., `w89s2` is the sum of the predicted probability of outcomes 1 and 2).

```
. desc w89*
variable name      storage  display  value  variable label
                  type      format   label
-----
w89x               float    %9.0g   Age in years
w89p1              float    %9.0g   pr(SD)=Pr(1)
w89p2              float    %9.0g   pr(D)=Pr(2)
w89p3              float    %9.0g   pr(A)=Pr(3)
w89p4              float    %9.0g   pr(SA)=Pr(4)
w89s1              float    %9.0g   pr(y<=1)
w89s2              float    %9.0g   pr(y<=2)
w89s3              float    %9.0g   pr(y<=3)
w89s4              float    %9.0g   pr(y<=4)
```

Although `prgen` assigns variable labels to the variables it creates, we can change these to improve the look of the plot that we are creating. Specifically,

```
. label var w89p1 "SD"
. label var w89p2 "D"
. label var w89p3 "A"
. label var w89p4 "SA"
. label var w89s1 "SD"
. label var w89s2 "SD or D"
. label var w89s3 "SD, D or A"
```

First, we plot the probabilities of individual outcomes using `graph`. Because the `graph` command is long, we use `///` to allow the commands to be longer than one line in our do-file.

```
. // step 1: graph predicted probabilities
. graph twoway connected w89p1 w89p2 w89p3 w89p4 w89x, ///
> title("Panel A: Predicted Probabilities") ///
> xtitle("Age") xlabel(20(10)80) ylabel(0(.25).50) ///
> yscale(noline) ylabel("") xline(44.93) ///
> ytitle("") name(graph1, replace)
```

This `graph` command plots the four predicted probabilities against generated values for age contained in `w89x`. Standard options for `graph` are used to specify the axes and labels. The vertical line specified by `xline(44.93)` marks the average age in the sample. This line is used to illustrate the marginal effect discussed in section 5.8.7. Option `name(graph1, replace)` saves the graph in memory under the name `graph1` so that we can combine it with the next graph, which plots the cumulative probabilities:

```
. // step 2: graph cumulative probabilities
. graph twoway connected w89s1 w89s2 w89s3 w89x, ///
> title("Panel B: Cumulative Probabilities") ///
> xtitle("Age") xlabel(20(10)80) ylabel(0(.25)1) ///
> yscale(noline) ylabel("") name(graph2, replace) ///
> ytitle("")
```

Next we combine these two graphs (see chapter 2 for details on combining graphs):

```
. // step 3: combine graphs
. graph combine graph1 graph2, col(1) iscale(*.9) imargin(small) ///
> ysize(4.31) xsize(3.287)
```

This leads to figure 5.2. Panel A plots the predicted probabilities and shows that with age the probability of SA decreases rapidly, whereas the probability of D (and to a lesser degree, SD) increases. Panel B plots the cumulative probabilities. Both panels present the same information; which method you use is up to you.

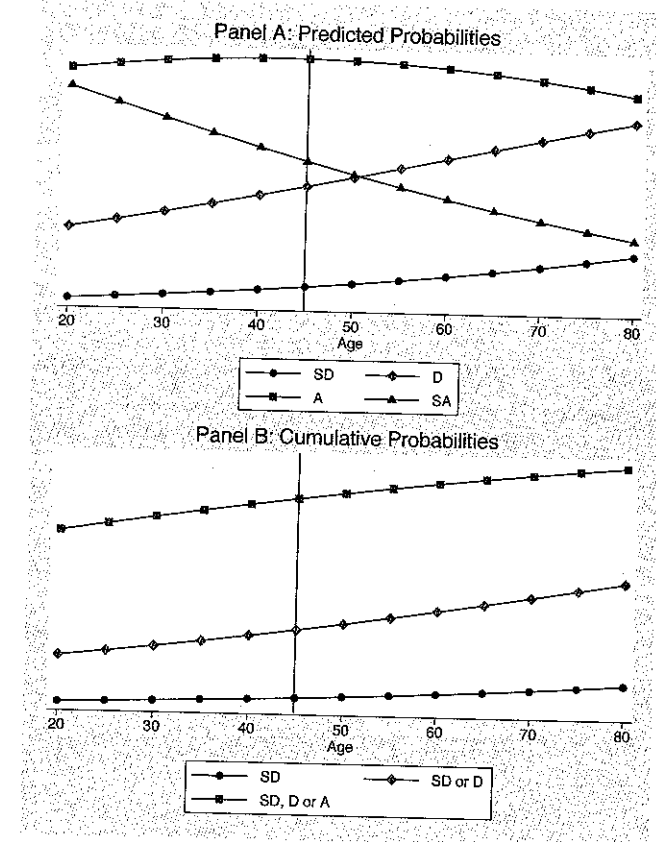


Figure 5.2: Plot of predicted probabilities for the ordered logit model.

5.8.7 Changes in predicted probabilities

When there are many variables in the model, it is impractical to plot them all. In such cases, measures of change in the outcome probabilities are useful for summarizing the effects of each variable. Before proceeding, however, we hasten to note that values of

both discrete and marginal change depend on the levels of all variables in the model. We return to this point shortly.

Marginal change with prchange

The marginal change in the probability is computed as

$$\frac{\partial \Pr(y = m | \mathbf{x})}{\partial x_k} = \frac{\partial F(\tau_m - \mathbf{x}\beta)}{\partial x_k} - \frac{\partial F(\tau_{m-1} - \mathbf{x}\beta)}{\partial x_k}$$

which is the slope of the curve relating x_k to $\Pr(y = m | \mathbf{x})$, holding all other variables constant. Here we consider the marginal effect of age $\{\partial \Pr(y = m | \mathbf{x}) / \partial \text{age}\}$ for women in 1989 who are average on all other variables. This corresponds to the slope of the curves in panel A of figure 5.2 evaluated at the vertical line (recall that this line is drawn at the average age in the sample). The marginal is computed with `prchange`, where we specify that only the coefficients for age should be computed:

```
. prchange age, x(male=0 yr89=1) rest(mean)
ologit: Changes in Probabilities for warm
age
      Avg|Chg|      SD      D      A      SA
Min->Max .16441458 .10941909 .21941006 -.05462247 -.27420671
  +1/2 .00222661 .00124099 .00321223 -.0001803 -.00427291
  +sd/2 .0373125 .0208976 .05372739 -.00300205 -.07162295
MargEfct .00222662 .00124098 .00321226 -.00018032 -.00427292

      SD      D      A      SA
Pr(y|x) .06099996 .22815652 .44057754 .27026597

      yr89      male      white      age      ed      prst
x=      1      0      .876581      44.9355      12.2181      39.5853
sd(x)= .489718 .498875 .328989 16.779 3.16083 14.4923
```

The first thing to notice is the row labeled `Pr(y|x)`, which is the predicted probabilities at the values set by `x()` and `rest()`. In panel A, these probabilities correspond to the intersection of the vertical line and the probability curves. The row `MargEfct` lists the slopes of the probability curves at the point of intersection with the vertical line in the figure. For example, the slope for `SD` (shown with circles) is .00124, whereas the slope for `A` (shown with squares) is negative and small. As with the BRM, the size of the slope indicates the instantaneous rate of change but does not correspond exactly to the amount of change in the probability for a change of one unit in the independent variable. However, when the probability curve is approximately linear, the marginal effect can be used to summarize the effect of a unit change in the variable on the probability of an outcome.

Marginal change with mfx

Marginal change can also be computed using `mfx`, where `at()` is used to set values of the independent variables. And it estimates the marginal effects for only one outcome category at a time, where the category is specified with the option `predict(outcome(#))`.

Using the same values for the independent variables as in the example above, we obtain the following results:

```
. mfx, at(male=0 yr89=1) predict(outcome(1))
warning: no value assigned in at() for variables white age ed prst;
means used for white age ed prst
Marginal effects after ologit
      y = Pr(warm==1) (predict, outcome(1))
      = .06099996

variable |      dy/dx      Std. Err.      z      P>|z|      [      95% C.I.      ]      x
-----+-----+-----+-----+-----+-----+-----
yr89*   | -.0378526      .00601      -6.30      0.000      -.049633      -.026072      1
male*   | .0581355      .00731      7.95      0.000      .043803      .072468      0
white*  | .0197511      .0055      3.59      0.000      .008972      .03053      .876581
age     | .001241      .00016      7.69      0.000      .000925      .001557      44.9355
ed      | -.0038476      .00097      -3.96      0.000      -.005754      -.001941      12.2181
prst    | -.0003478      .00019      -1.83      0.068      -.000721      .000025      39.5853
```

(*) `dy/dx` is for discrete change of dummy variable from 0 to 1

The marginal for age is .001241, which matches the result obtained from `prchange`. The advantage of `mfx` is that it computes standard errors and confidence intervals.

Discrete change with prchange

As the marginal change can be misleading when the probability curve is changing rapidly or when an independent variable is a dummy variable, we prefer using discrete change (`mfx` computes discrete change for independent variables that are binary but not for other independent variables). The discrete change is the change in the predicted probability for a change in x_k from the start value x_S to the end value x_E (e.g., a change from $x_k = 0$ to $x_k = 1$). Formally,

$$\frac{\Delta \Pr(y = m | \mathbf{x})}{\Delta x_k} = \Pr(y = m | \mathbf{x}, x_k = x_E) - \Pr(y = m | \mathbf{x}, x_k = x_S)$$

where $\Pr(y = m | \mathbf{x}, x_k)$ is the probability that $y = m$ given \mathbf{x} , noting a specific value for x_k . The change is interpreted as indicating that when x_k changes from x_S to x_E , the predicted probability of outcome m changes by $\Delta \Pr(y = m | \mathbf{x}) / \Delta x_k$, holding all other variables at \mathbf{x} .

The value of the discrete change depends on (1) the value at which x_k starts, (2) the amount of change in x_k , and (3) the values of all other variables. Most often, each continuous variable except x_k is held at its mean. For dummy independent variables, the change could be computed for both values of the variable. For example, we could compute the discrete change for age separately for men and women.

Here the discrete-change coefficients for `male`, `age`, and `prst` for women in 1989, with other variables at their mean, are computed as follows:


```

. prchange male age prst, x(male=0 yr89=1) rest(mean)
ologit: Changes in Probabilities for warm
male
      Avg|Chg|      SD      D      A      SA
0->1  .08469636  .05813552  .11125721  -.05015317  -.11923955
age
      Avg|Chg|      SD      D      A      SA
Min->Max .16441458  .10941909  .21941006  -.05462247  -.27420671
--1/2  .00222661  .00124099  .00321223  -.0001803  -.00427291
--sd/2  .0373125  .0208976  .05372739  -.00300205  -.07162295
MargEfct .00222662  .00124098  .00321226  -.00018032  -.00427292
prst
      Avg|Chg|      SD      D      A      SA
Min->Max .04278038  -.02352008  -.06204067  .00013945  .08542132
--1/2  .00062411  -.00034784  -.00090037  .00005054  .00119767
--sd/2  .00904405  -.00504204  -.01304607  .00073212  .01735598
MargEfct .00062411  -.00034784  -.00090038  .00005054  .00119767
      SD      D      A      SA
Pr(y|x) .06099996  .22815652  .44057754  .27026597
      yr89  male  white  age  ed  prst
x=      1      0  .876581  44.9355  12.2181  39.5853
sd(x)=  .489718  .498875  .328989  16.779  3.16083  14.4923

```

For variables that are not binary, the discrete change can be interpreted for a unit change centered on the mean, for a standard deviation change centered on the mean, or as the variable changes from its minimum to its maximum value. The following are two examples:

For a standard deviation increase in age, the probability of disagreeing increases by .05, holding other variables constant at their means.

Moving from the minimum prestige to the maximum prestige changes the predicted probability of strongly agreeing by .08, holding all other variables constant at their means.

The J discrete-change coefficients for a variable can be summarized by computing the average of the *absolute values* of the changes across all the outcome categories:

$$\bar{\Delta} = \frac{1}{J} \sum_{j=1}^J \left| \frac{\Delta \Pr(y = j | \bar{x})}{\Delta x_k} \right|$$

The absolute value must be used because the sum of the changes without taking the absolute value is necessarily zero. These are labeled as Avg|Chg|. For example, the effect of being a male is on average 0.08, which is larger than the average effect of a standard deviation change in either age or occupational prestige.

Confidence intervals for discrete changes

Although prchange computes discrete changes, it does not compute confidence intervals for these changes. If you want confidence intervals, you need to use prvalue with the save and diff options. In the first call to prvalue, you provide the starting values for the explanatory variables and specify save. In the second call, you provide the ending values and specify diff.

Consider the change in probabilities associated with a change from 0 to 1 in a dummy variable. With prchange, we calculated the difference in probabilities between men and women in 1989, holding the other variables at their means. We can compute this same change using prvalue, which will give us the confidence intervals as well.

```

. qui prvalue, x(male=0 yr89=1) rest(mean) save
. prvalue, x(male=1 yr89=1) rest(mean) diff
ologit: Change in Predictions for warm
Confidence intervals by delta method

```

	Current	Saved	Change	95% CI for Change
Pr(y=SD x):	0.1191	0.0610	0.0581	[0.0438, 0.0725]
Pr(y=D x):	0.3394	0.2282	0.1113	[0.0876, 0.1350]
Pr(y=A x):	0.3904	0.4406	-0.0502	[-0.0679, -0.0324]
Pr(y=SA x):	0.1510	0.2703	-0.1192	[-0.1448, -0.0937]

	yr89	male	white	age	ed	prst
Current=	1	1	.8765809	44.935456	12.218055	39.585259
Saved=	1	0	.8765809	44.935456	12.218055	39.585259
Diff=	0	1	0	0	0	0

If you compare the results listed in the Change column with the earlier results from prchange, you will see that they are the same. For example, the probability of strongly agreeing with the item about working mothers is .119 lower for men than for women. The estimated bounds of the 95% confidence interval for differences in probabilities between men and women in 1989 are $-.145$ and $-.094$, holding other variables at their means. Or, we might say, the results suggest that the difference between men and women in the predicted probability of strongly agreeing is .12 and the difference could be as small as .09 or as large as .14 with 95% confidence.

Now let's consider discrete change with a continuous variable. Earlier, prchange computed the changes associated with a standard deviation increase in age centered around the mean. The mean age is 44.9, with a standard deviation of 16.8. Accordingly, we want to compute the change in probabilities when age increases from 36.5 to 53.3. The start and end values for age are computed using the display command:

```

. di 44.935 - (.5*16.779)
36.5455
. di 44.935 + (.5*16.779)
53.3245

```

Using these values in a pair of prvalue commands, we find

```
. qui prvalue, x(male=0 yr89=1 age=36.5455) rest(mean) save
. prvalue, x(male=0 yr89=1 age=53.3245) rest(mean) diff
ologit: Change in Predictions for warm
Confidence intervals by delta method
```

	Current	Saved	Change	95% CI for Change	
Pr(y=SD x):	0.0723	0.0514	0.0209	[0.0155,	0.0262]
Pr(y=D x):	0.2556	0.2019	0.0537	[0.0414,	0.0661]
Pr(y=A x):	0.4362	0.4392	-0.0030	[-0.0118,	0.0058]
Pr(y=SA x):	0.2359	0.3076	-0.0716	[-0.0883,	-0.0549]

	yr89	male	white	age	ed	prst
Current=	1	0	.8765809	53.3245	12.218055	39.585259
Saved=	1	0	.8765809	36.5455	12.218055	39.585259
Diff=	0	0	0	16.779	0	0

The amount of change matches the results from prchange, but now we have confidence intervals. For otherwise average women in 1989, a standard deviation increase in age, about 17 years, increases the probability of disagreeing by .054 with estimated bounds for the 95% confidence interval at .041 and .066.

Computing discrete change for a 10-year increase in age

In the example above, age was measured in years. Not surprisingly, the change in the predicted probability for a 1-year increase in age is trivially small. But, to characterize the effect of age, we could report the effect of a 10-year change in age.

Warning It is tempting to compute the discrete change for a 10-year change in age by simply multiplying the 1-year discrete change by 10. This will give you approximately the right answer if the probability curve is nearly linear over the range of change. But, when the curve is not linear, simply multiplying can give misleading results and even the wrong sign. To be safe, do not do it!

The delta(#) option for prchange computes the discrete change as an independent value changes from #/2 units below the base value to #/2 above. Here we use delta(10) and set the base value of age to its mean:

```
. prchange age, x(male=0 yr89=1) rest(mean) delta(10)
ologit: Changes in Probabilities for warm
(Note: d = 10)
age
```

	Avg Chg	SD	D	A	SA
Min->Max	.16441458	.10941909	.21941006	-.05462247	-.27420671
--d/2	.02225603	.01242571	.03208634	-.00179818	-.04271388
--sd/2	.0373125	.0208976	.05372739	-.00300205	-.07162295
MargEfct	.00222662	.00124098	.00321226	-.00018032	-.00427292

	SD	D	A	SA		
Pr(y x)	.06099996	.22815652	.44057754	.27026597		
	yr89	male	white	age	ed	prst
x=	1	0	.876581	44.9355	12.2181	39.5853
sd(x)=	.489718	.498875	.328989	16.779	3.16083	14.4923

For females interviewed in 1989, the results in the --d/2 row show the changes in the predicted probabilities associated with a 10-year increase in age centered on the mean.

5.8.8 Odds ratios using listcoef

For ologit, but not oprobit, we can interpret the results using odds ratios. Earlier, (5.2) defined the ordered logit model as

$$\Omega_{\leq m | > m}(\mathbf{x}) = \exp(\tau_m - \mathbf{x}\beta)$$

For example, with four outcomes we would simultaneously estimate three equations:

$$\begin{aligned} \Omega_{\leq 1 | > 1}(\mathbf{x}) &= \exp(\tau_1 - \mathbf{x}\beta) && 1 \text{ vs } 2 \\ \Omega_{\leq 2 | > 2}(\mathbf{x}) &= \exp(\tau_2 - \mathbf{x}\beta) && 2 \text{ vs } 3 \\ \Omega_{\leq 3 | > 3}(\mathbf{x}) &= \exp(\tau_3 - \mathbf{x}\beta) && 3 \text{ vs } 4 \end{aligned}$$

Using the same approach as shown for binary logit, the effect of a change in x_k of 1 equals

$$\frac{\Omega_{\leq m | > m}(\mathbf{x}, x_k + 1)}{\Omega_{\leq m | > m}(\mathbf{x}, x_k)} = e^{-\beta_k} = \frac{1}{e^{\beta_k}}$$

which can be interpreted as indicating that for a unit increase in x_k , the odds of an outcome being less than or equal to m is changed by the factor $\exp(-\beta_k)$, holding all other variables constant.

The value of the odds ratio does *not* depend on the value of m , which is why the parallel regression assumption is also known as the proportional odds assumption. We could interpret the odds ratio as follows:

For a unit increase in x_k , the odds of a lower outcome compared with a higher outcome are changed by the factor $\exp(-\beta_k)$, holding all other variables constant.

or, for a change in x_k of δ ,

$$\frac{\Omega_{\leq m | > m}(\mathbf{x}, x_k + \delta)}{\Omega_{\leq m | > m}(\mathbf{x}, x_k)} = \exp(-\delta \times \beta_k) = \frac{1}{\exp(\delta \times \beta_k)}$$

so that

for an increase of δ in x_k , the odds of a lower outcome compared with a higher outcome change by the factor $\exp(-\delta \times \beta_k)$, holding all other variables constant.

In these results, we are discussing factor changes in the odds of lower outcomes compared with higher outcomes. This is done because the model is traditionally written as $\ln \Omega_{\leq m | > m}(\mathbf{x}) = \tau_m - \mathbf{x}\beta$, which leads to the factor change coefficient of $\exp(-\beta_k)$. For purposes of interpretation, we could just as well consider the factor change in the odds of higher versus lower values; that is, changes in the odds $\Omega_{> m | \leq m}(\mathbf{x})$. This would equal $\exp(\beta_k)$.

The odds ratios for both a unit and a standard deviation change of the independent variables can be computed with `listcoef`, which lists the factor changes in the odds of higher versus lower outcomes. Here we request coefficients for only male and age:

```
. ologit warm yr89 male white age ed prst, nolog
(output omitted)
```

```
. listcoef male age, help
```

```
ologit (N=2293): Factor Change in Odds
```

```
Odds of: >m vs <=m
```

warm	b	z	P> z	e ^{-b}	e ^{-b} StdX	SDofX
male	-0.73330	-9.343	0.000	0.4803	0.6936	0.4989
age	-0.02167	-8.778	0.000	0.9786	0.6952	16.7790

b = raw coefficient

z = z-score for test of b=0

P>|z| = p-value for z-test

e^{-b} = exp(b) = factor change in odds for unit increase in X

e^{-b}StdX = exp(b*SD of X) = change in odds for SD increase in X

SDofX = standard deviation of X

or to compute percent changes in the odds,

```
. listcoef male age, help percent
ologit (N=2293): Percentage Change in Odds
Odds of: >m vs <=m
```

warm	b	z	P> z	%	%StdX	SDofX
male	-0.73330	-9.343	0.000	-52.0	-30.6	0.4989
age	-0.02167	-8.778	0.000	-2.1	-30.5	16.7790

b = raw coefficient

z = z-score for test of b=0

P>|z| = p-value for z-test

% = percent change in odds for unit increase in X

%StdX = percent change in odds for SD increase in X

SDofX = standard deviation of X

These results can be interpreted as follows:

The odds of having more positive attitudes toward working mothers are .48 times smaller for men than women, holding all other variables constant. Equivalently, the odds of having more positive values are 52% smaller for men than women, holding other variables constant.

For a standard deviation increase in age, the odds of having more positive attitudes decrease by a factor of .69, holding all other variables constant.

When presenting odds ratios, we find that people find it easier to understand the results if you talk about *increases* in the odds rather than *decreases*. That is, it is clearer to say, "The odds increased by a factor of 2" than to say, "The odds decreased by a factor of .5". If you agree, then you can reverse the order when presenting odds. For example, we could say that

The odds of having more *negative* attitudes toward working mothers are 2.08 times larger for men than women, holding all other variables constant.

This new factor change, 2.08, is just the inverse of the old value .48 (that is, 1/.48). `listcoef` computes the odds of a lower category versus a higher category if you specify the reverse option:

```
. listcoef male, reverse
```

```
ologit (N=2293): Factor Change in Odds
```

```
Odds of: <=m vs >m
```

warm	b	z	P> z	e ^{-b}	e ^{-b} StdX	SDofX
male	-0.73330	-9.343	0.000	2.0819	1.4417	0.4989

The output now says Odds of: $\leq m$ vs $> m$ instead of Odds of: $> m$ vs $\leq m$, as it did earlier.

When interpreting the odds ratios, remember two points that are discussed in detail in chapter 4. First, because odds ratios are multiplicative coefficients, *positive and negative effects should be compared by taking the inverse of the negative effect* (or vice versa). For example, a negative factor change of .5 has the same magnitude as a positive factor change of $2 = 1/.5$. Second, the interpretation assumes only that the other variables have been held constant, not held at any specific values (as was required for discrete change). But, a *constant factor change in the odds does not correspond to a constant change or constant factor change in the probability*.

5.9 Less common models for ordinal outcomes

Stata can also be used to fit several less commonly used models for ordinal outcomes. In concluding this chapter, we describe these models briefly and note their commands for estimation. Our SPost commands do not work with these models. For `ocratio`, this is mainly because these commands do not fully incorporate the new methods of returning information that were introduced with an earlier version of Stata.

5.9.1 The stereotype model

The *stereotype logistic model* (SLM), also referred to as the *stereotype ordered regression model* (SORM), was proposed by Anderson (1984) in response to the restrictive assumption of parallel regressions in the ordered regression model. This SLM, which can be fitted with in Stata using the `slogit` command, is a compromise between allowing the coefficients for each independent variable to vary by outcome category (as is the case with multinomial logit model considered in the next chapter) and restricting the coefficients to be identical across all outcomes as was the case with the OLM. The SLM is defined as

$$\ln \frac{\Pr(y = q | \mathbf{x})}{\Pr(y = r | \mathbf{x})} = (\theta_q - \theta_r) - (\phi_q - \phi_r) (\mathbf{x}\beta) \quad (5.3)$$

where β is a vector of coefficients associated with the independent variables, the θ s are intercepts, and the ϕ s are scale factors that mediate the effects of the x 's. Because this model is so closely related to the multinomial logit model, and is sometimes used for nominal data, we will postpone discussion until the next chapter.

5.9.2 The generalized ordered logit model

The parallel regression assumption results from assuming the same coefficient vector β for all comparisons in the $J - 1$ equations

$$\ln \Omega_{\leq m | > m}(\mathbf{x}) = \tau_m - \mathbf{x}\beta$$

5.9.3 The continuation ratio model

where $\Omega_{\leq m | > m}(\mathbf{x}) = \{\Pr(y \leq m | \mathbf{x})\} / \{\Pr(y > m | \mathbf{x})\}$. The generalized ordered logit model (GOLM) allows β to differ for each of the $J - 1$ comparisons. That is,

$$\ln \Omega_{\leq m | > m}(\mathbf{x}) = \tau_m - \mathbf{x}\beta_m \quad \text{for } j = 1 \text{ to } J - 1$$

where predicted probabilities are computed as

$$\begin{aligned} \Pr(y = 1 | \mathbf{x}) &= \frac{\exp(\tau_1 - \mathbf{x}\beta_1)}{1 + \exp(\tau_1 - \mathbf{x}\beta_1)} \\ \Pr(y = j | \mathbf{x}) &= \frac{\exp(\tau_j - \mathbf{x}\beta_j)}{1 + \exp(\tau_j - \mathbf{x}\beta_j)} - \frac{\exp(\tau_{j-1} - \mathbf{x}\beta_{j-1})}{1 + \exp(\tau_{j-1} - \mathbf{x}\beta_{j-1})} \quad \text{for } j = 2 \text{ to } J - 1 \\ \Pr(y = J | \mathbf{x}) &= 1 - \frac{\exp(\tau_{J-1} - \mathbf{x}\beta_{J-1})}{1 + \exp(\tau_{J-1} - \mathbf{x}\beta_{J-1})} \end{aligned}$$

To ensure that the $\Pr(y = j | \mathbf{x})$ is between 0 and 1, the condition

$$(\tau_j - \mathbf{x}\beta_j) \geq (\tau_{j-1} - \mathbf{x}\beta_{j-1})$$

must hold. Once predicted probabilities are computed, all the approaches used to interpret the ORM results can be readily applied. This model has been discussed by Clogg and Shihadeh (1994, 146–147), Fahrmeir and Tutz (1994, 91), and McCullagh and Nelder (1989, 155). It can be fitted in Stata with the add-on command `gologit` (Fu 1998). This command has not been recently updated and thus is no longer supported in SPost. More recently, Williams (2005) has written `gologit2`, which extends the original command to fit two special cases of the general models: the proportional odds model and the partial proportional odds model (see Lall et al. 2002; Peterson and Harrell 1990). These models are less restrictive than the ordinal logit model fitted by `ologit` but more parsimonious than the multinomial logit model fitted by `mlogit`. We plan to add support for `gologit2` to SPost.

5.9.3 The continuation ratio model

The *continuation ratio model* was proposed by Fienberg (1980, 110) and was designed for ordinal outcomes in which the categories represent the progression of events or stages in some process through which an individual can advance. For example, the outcome could be faculty rank, where the stages are assistant professor, associate professor, and full professor. A key characteristic of the process is that an individual must pass through each stage. For example, to become an associate professor you must be an assistant professor; to be a full professor, an associate professor. Although there are versions of this model based on other binary models (e.g., probit), here we consider the logit version.

If $\Pr(y = m | \mathbf{x})$ is the probability of being in stage m given \mathbf{x} and $\Pr(y > m | \mathbf{x})$ is the probability of being in a stage later than m , the continuation ratio model for the log odds is

$$\ln \left\{ \frac{\Pr(y = m | \mathbf{x})}{\Pr(y > m | \mathbf{x})} \right\} = \tau_m - \mathbf{x}\beta \quad \text{for } m = 1 \text{ to } J - 1$$

where the β s are constrained to be equal across outcome categories, whereas the constant term τ_m differs by stage. As with other logit models, we can also express the model in terms of the odds:

$$\frac{\Pr(y = m | \mathbf{x})}{\Pr(y > m | \mathbf{x})} = \exp(\tau_m - \mathbf{x}\beta)$$

Accordingly, $\exp(-\beta_k)$ can be interpreted as the effect of a unit increase in x_k on the odds of being in m compared with being in a higher category given that an individual is in category m or higher, holding all other variables constant. From this equation, the predicted probabilities can be computed as

$$\Pr(y = m | \mathbf{x}) = \frac{\exp(\tau_m - \mathbf{x}\beta)}{\prod_{j=1}^m \{1 + \exp(\tau_j - \mathbf{x}\beta)\}} \quad \text{for } m = 1 \text{ to } J - 1$$

$$\Pr(y = J | \mathbf{x}) = 1 - \sum_{j=1}^{J-1} \Pr(y = j | \mathbf{x})$$

These predicted probabilities can be used for interpreting the model. In Stata, this model can be fitted using `ocratio` by Wolfe (1998); type `net search ocratio` and follow the prompts to download.

6 Models for nominal outcomes with case-specific data

An outcome is nominal when the categories are assumed to be unordered. For example, marital status can be grouped nominally into the categories of divorced, never married, married, or widowed. Occupations might be organized as professional, white collar, blue collar, craft, and menial, which is the example we use in this chapter. Other examples include reasons for leaving the parents' home, the organizational context of scientific work (e.g., industry, government, and academia), and the choice of language in a multilingual society. Further, in some cases a researcher might prefer to treat an outcome as nominal, even though it is ordered or partially ordered. For example, if the response categories are strongly agree, agree, disagree, strongly disagree, and don't know, the category "don't know" invalidates models for ordinal outcomes. Or, you might decide to use a nominal regression model when the assumption of parallel regressions is rejected. In general, if you have concerns about the ordinality of the dependent variable, the potential loss of efficiency in using models for nominal outcomes is outweighed by avoiding potential bias.

This chapter focuses on three closely related models for nominal (and sometimes ordinal) outcomes with case-specific data. The multinomial logit model (MNL) is the most frequently used nominal regression model. In this model, you are essentially estimating a separate binary logit for each pair of outcome categories. Next we consider the multinomial probit model with uncorrelated errors, which is the normal counterpart to the MNL. We then discuss the stereotype logistic regression model (SLM). Although this model is often used for ordinal outcomes, it is closely related to the MNL. All these models assume that the data are case specific, meaning that each independent variable has one value for each individual. Examples of such variables are an individual's race or education. In the next chapter, we consider models that include alternative-specific data.

Models for nominal outcomes, both in this chapter and the next, require us to be more exacting about some basic terminology. Until now we have used "individual", "observation", and "case" interchangeably to refer to observational units, where each observational unit corresponds to a single row or record in the dataset. In the next two chapters, we will use only the term "case" for this purpose. Most of the time, we use the word "alternative" to refer to a possible outcome. Sometimes we refer to an alternative as an outcome category or a comparison group in order to be consistent with the usual terminology for a model or the output generated by Stata. The term "choice"