POLI 8501
Models for Ordinal Responses II

Introduction

There are several possible ways of interpreting these models, all of which have similarities to binary logit/probit:

- Coefficient and standard error estimates . . .
- Marginal effects/first derivatives . . .
- Predicted probabilities, and changes in them.
- Cumulative predicted probabilities.
- Odds ratios in the ordered logit model.

We'll use the following running example: BEER!

- The data are from a rating of 69 domestic and imported beers conducted in June, 1996 by *Consumer Reports* magazine. Quality is coded on a four point scale:

  1. Fair
  2. Good
  3. Very Good
  4. Excellent

- Explanatory variables are:
  - *price* per six 12 ounce servings (in dollars, $\mathrm{E}[\beta > 0]$),
  - *calories* per 12 ounce serving ($\mathrm{E}[\beta > 0]$),
  - an indicator for whether ($= 1$) or not the beer was a "craft" beer (i.e, from a microbrewery) ($\mathrm{E}[\beta > 0]$), and
  - two indicators of flavor: a scale ($0-100$) of the beer's level of *bitterness* ($\mathrm{E}[\beta < 0]$), and a similar scale of its *maltiness* ($\mathrm{E}[\beta > 0]$).

  .

The data, and the basic model estimates, look like this:

```
. su quality price calories craftbeer bitter malty

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     quality |        69    2.536232    1.145062          1          4
       price |        69    4.963188    1.446516       2.36        7.8
    calories |        69    142.3478    29.90221         58        201
   craftbeer |        69    .4347826    .4993602          0          1
      bitter |        69    35.44203    17.99786          8       80.5
-------------+--------------------------------------------------------
       malty |        69    33.13043    26.17288          5         86

. ologit quality price calories craftbeer bitter malty

Ordered logistic regression                       Number of obs   =         69
                                                  LR chi2(5)      =      27.29
                                                  Prob > chi2     =     0.0000
Log likelihood = -81.871058                       Pseudo R2       =     0.1429

------------------------------------------------------------------------------
     quality |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       price |  -.4505038   .2933979    -1.54   0.125    -1.025553    .1245455
    calories |   .0467436   .0121975     3.83   0.000      .022837    .0706502
   craftbeer |  -1.704671    .941856    -1.81   0.070    -3.550675     .141333
      bitter |  -.0296914    .042417    -0.70   0.484    -.1128272    .0534444
       malty |   .0505279   .0245936     2.05   0.040     .0023254    .0987304
-------------+----------------------------------------------------------------
       /cut1 |   2.771035   1.673818                     -.5095873    6.051657
       /cut2 |   4.269676    1.72483                      .8890725     7.65028
       /cut3 |   5.577651   1.759538                      2.129021    9.026282
------------------------------------------------------------------------------
```

We’ll discuss, in turn, a number of different ways of analyzing/interpreting models like this.

Signs-N-Significance

- Signs-n-significance can tell you a little about the results, BUT
- There's also a problem with interpreting these models this way (over and above the usual reasons not to do so) – we'll talk about this below...

Odds Ratios

For the ordered logit, one can use an odds-ratio interpretation of the coefficients. For that model, the change in the odds of Y being greater than j (versus being less than or equal to j) associated with a δ-unit change in X_k is equal to $\exp(\delta\hat{\beta}_k)$.

Happily, Stata makes getting odds ratios (along with their associated standard errors, etc.) trivially easy:

```
. ologit, or

Ordered logistic regression                       Number of obs   =         69
                                                  LR chi2(5)      =      27.29
                                                  Prob > chi2     =     0.0000
Log likelihood = -81.871058                       Pseudo R2       =     0.1429
------------------------------------------------------------------------------
     quality | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       price |    .637307   .1869845    -1.54   0.125      .358598    1.132634
    calories |   1.047853   .0127812     3.83   0.000       1.0231    1.073206
   craftbeer |   .1818322   .1712598    -1.81   0.070     .0287053    1.151808
      bitter |    .970745   .0411761    -0.70   0.484      .893305    1.054898
       malty |   1.051826   .0258682     2.05   0.040     1.002328    1.103769
-------------+----------------------------------------------------------------
       /cut1 |   2.771035   1.673818                     -.5095873    6.051657
       /cut2 |   4.269676    1.72483                      .8890725     7.65028
       /cut3 |   5.577651   1.759538                      2.129021    9.026282
------------------------------------------------------------------------------
```

Interpretation of these odds ratios is straightforward: the change in odds associated with a one-unit change in X_k is $\exp(\hat{\beta}_k)$.

- So, in our example, for, e.g., the `craftbeer` variable:
  - $\exp(-1.705) = 0.18$
  - Thus, the odds of being rated "Good" or better (versus "Fair") are more than 80 percent lower for a microbrew than for a regular beer.

- And, the odds of being rated "Very Good" or "Excellent" (versus "Fair" or "Good") are about 80 percent lower for a microbrew than for a regular beer as well, etc.

- Similarly, for the (continuous) `calories` variable:
  - The odds ratio is $\exp(0.047) = 1.05$
  - This means that a one-calorie increase raises the probability of being in a higher set of categories (versus all lower ones) by about five percent.
  - We'll see this in terms of the predicted probabilities as well (below).

- The fact that the odds ratio is the same for all values of J is a function of the parallel regressions assumption, which we discussed last time (and will talk a bit more about later today...).

As was the case for binary logits, odds ratios are a nice, intuitive way of discussing the substantive impact of your covariates. They're probably best used "textually" – that is, as part of the written discussion of your research findings, rather than reported in a table or the like.

Partial Derivatives/Marginal Effects

Yet another way of looking at covariate effects is to consider the partial derivative of the probability of any particular outcome, with respect to the covariate of interest. In these models, this is:

$$\begin{aligned}\frac{\partial \Pr(Y=j)}{\partial X_k} &= \frac{\partial F(\hat{\tau}_j - \bar{\mathbf{X}}\hat{\beta})}{\partial X_k} - \frac{\partial F(\hat{\tau}_{j-1} - \bar{\mathbf{X}}\hat{\beta})}{\partial X_k} \\ &= \hat{\beta}_k[f(\hat{\tau}_{j-1} - \bar{\mathbf{X}}\hat{\beta}) - f(\hat{\tau}_j - \bar{\mathbf{X}}\hat{\beta})] \qquad (1)\end{aligned}$$

So, for (e.g.) an ordered probit, (1) is:

$$\hat{\beta}_k[\phi(\hat{\tau}_{j-1} - \bar{\mathbf{X}}\hat{\beta}) - \phi(\hat{\tau}_j - \bar{\mathbf{X}}\hat{\beta})]$$

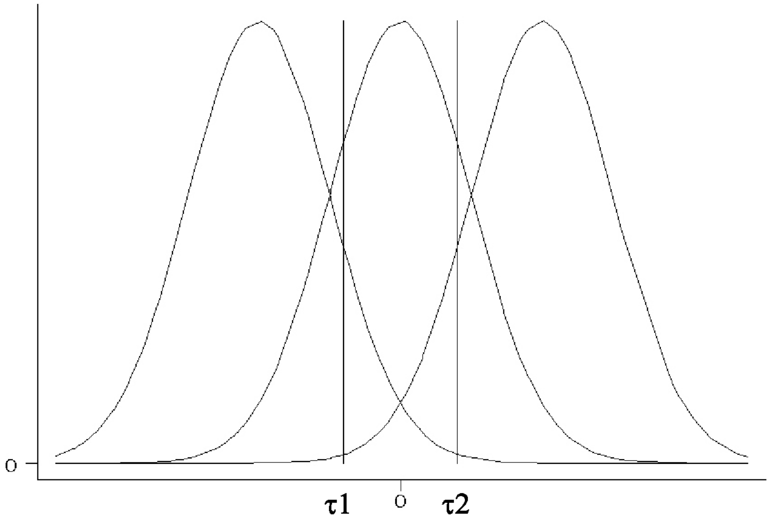
That is, the height of the density $f(\cdot)$ at $\hat{\tau}_{j-1} - \bar{\mathbf{X}}\hat{\beta}$ minus its height at $\hat{\tau}_j - \bar{\mathbf{X}}\hat{\beta}$ (see Figure 1). For an ordered logit, we'd substitute λ (the standard logistic density) for ϕ in (1).

Now, notice a few things about Equation (1):

- Remember that ϕ (or λ) is the density of the error terms around the regression line for the hypothetical Y^*.

- The effect of a change in X_k is to shift the bell-curve up or down the axis, while the positions of the "cut points" ($\tau_1 ... \tau_{j-1}$) stay the same...

Figure 1: Marginal Effects of X_k on $\Pr(Y = j)$



- Assume for the moment that we're holding $\bar{\mathbf{X}}\hat{\beta} = 0$. For the curve on the left, all the derivatives have the "correct" sign – that is, the same as the estimated coefficient. That is, since $f(\tau_1) - f(\tau_2) > 0$, then
  - Variables with positive coefficients will have positive derivatives...
  - The reverse will be true for X_k with $\hat{\beta}_k < 0$.
- The opposite is true for the curve on the right, where, because $\tau_1 < \tau_2$, the sign of the partial derivative will be "flipped" relative to that of $\hat{\beta}_k$.
- For the curve in the middle, the fact that the density is essentially equal at τ_1 and τ_2 means that the marginal will be very close to zero...

All this means that the size of the mass of probability in the middle category first gets larger, then smaller as we increase X_k (for a positive $\hat{\beta}_k$). Mathematically, this means that the change in the probability of getting a "middle category," as a function of X, is somewhat indeterminate...

- E.g., (assuming $\hat{\beta}_k > 0$), the probability of $Y = 1$ clearly decreases in X_k.
- Similarly, $\Pr(Y = 3)$ also increases in X_k.
- But $\Pr(Y = 2)$ may be going either up or down, relative to the other categories...

This last bit will be true for any "middle" ordinal category, in which the *PDF* "turns over" for that range of values of X_k. As you can imagine, this fact has implications for the predicted probabilities as well.

Getting Marginal Effects

While computing Eq. (1) isn't all that hard, Stata users can also use the `-mfx-` command to generate partial derivatives:

```
. mfx, predict(p outcome(1))

Marginal effects after ologit
      y  = Pr(quality==1) (predict, p outcome(1))
         =  .17837356
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
   price |   .0660242      .04343    1.52   0.128  -.019105  .151153   4.96319
calories |  -.0068506      .00209   -3.27   0.001  -.010954 -.002747   142.348
craftb~r*|   .2688865      .15549    1.73   0.084  -.035867   .57364   .434783
  bitter |   .0043515      .00629    0.69   0.489  -.007971  .016674    35.442
   malty |  -.0074052      .00368   -2.01   0.044  -.014626 -.000185   33.1304
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. mfx, predict(p outcome(2))

Marginal effects after ologit
      y  = Pr(quality==2) (predict, p outcome(2))
         =  .31443584
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
   price |   .0465784      .03534    1.32   0.188  -.022689  .115846   4.96319
calories |  -.0048329      .00215   -2.25   0.024   -.00904 -.000625   142.348
craftb~r*|   .1326553      .07074    1.88   0.061     -.006   .27131   .434783
  bitter |   .0030699      .00449    0.68   0.494  -.005737  .011876    35.442
   malty |  -.0052242      .00325   -1.61   0.108  -.011587  .001139   33.1304
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. mfx, predict(p outcome(3))
```

```
Marginal effects after ologit
      y  = Pr(quality==3) (predict, p outcome(3))
         =  .28950597
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
   price |  -.0358828       .0276    -1.30   0.194  -.089977  .018211   4.96319
calories |   .0037231      .00189     1.96   0.049   9.1e-06  .007437   142.348
craftb~r*|  -.1288796      .07264    -1.77   0.076  -.271256  .013497   .434783
  bitter |  -.0023649      .00357    -0.66   0.508  -.009361  .004631    35.442
   malty |   .0040246       .0026     1.55   0.121  -.001068  .009117   33.1304
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1

. mfx, predict(p outcome(4))

Marginal effects after ologit
      y  = Pr(quality==4) (predict, p outcome(4))
         =  .21768463
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
   price |  -.0767199      .05092    -1.51   0.132  -.176517  .023077   4.96319
calories |   .0079603      .00217     3.67   0.000   .003709  .012212   142.348
craftb~r*|  -.2726622      .14492    -1.88   0.060  -.556694  .011369   .434783
  bitter |  -.0050564       .0072    -0.70   0.482  -.019165  .009052    35.442
   malty |   .0086048      .00428     2.01   0.045   .000211  .016999   33.1304
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Note that here I've not specified `at(...)`, which means that, by default, Stata will set the values of the covariates to their means; one can change this relatively easily.

One can think of the marginal effects as the "slope" of the (nonlinear) regression line for that covariate on the probability of that outcome, assessed while holding constant the values of the variables in the model. So, consider the `malty` variable:

- At mean levels of **X**, the change in Pr(Fair) associated with an increase in `malty` is negative, and statistically differentiable from zero.
- The corresponding "slope" of `malty` on Pr(Good) is also negative, but not differentiable from zero.
- The effect of `malty` on Pr(Very Good) is positive, but "nonsignificant," and

- That effect vis-à-vis Pr(Excellent) is positive and "significant."

One can do a similar interpretation for each of the other variables. Note also that, by default, Stata calculates marginal effects for dummy variables as the discrete change in the probability for a one-unit change in that covariate; it notes this (with respect to the `craftbeer` variable) in the output of `mfx`.

Predicted Probabilities

Recall that

$$\widehat{\Pr(Y_i = j}|\mathbf{X}) = F(\hat{\tau}_j - \bar{\mathbf{X}}_i\hat{\beta}) - F(\hat{\tau}_{j-1} - \bar{\mathbf{X}}_i\hat{\beta}) \tag{2}$$

This is the basic formula from which we calculate predicted probabilities. As in the binary case, there are lots of things we can do with these, and lots of ways to do them...

- Calculate predicted probabilities that $Y = \{1, 2, 3, \text{etc.}\}$.
- Simply do it "by hand", or
- (better) let Stata do it for you, using `-predict-`, or
- use 𝔆larify.
- Consider changes for discrete changes/dummy covariates...
- ...or changes over ranges of continuous covariates.
- Can also plot graphs of predicted probabilities by continuous covariates...
  - Either simple predictions (that is, $\widehat{\Pr(Y = 1)}$), or
  - "Cumulative" probabilities (see below).

Let's reconsider our beer example...

Calculating Discrete Predicted Probabilities

To get our probabilities, we just set our variables to the values we want, and then "plug them in" to (2), above. Let's go with means for the four continuous variables, and the median for `craftbeer`.

1. Set `price` = 4.96, `calories` = 142, `craftbeer` = 0, `bitter` = 35.4, and `malty` = 33.1.

2. This gives us an “index value” of:

$$\begin{aligned}\sum_{k=1}^{K} \bar{\mathbf{X}}_k \hat{\beta}_k &= -0.45 \times 4.96 + 0.047 \times 142 - 1.70 \times 0 - 0.03 \times 35.4 + 0.05 \times 33.1 \\ &= -2.23 + 6.67 - 0 - 1.06 + 1.66 \\ &= \mathbf{5.04}.\end{aligned}$$

3. From this, we just calculate the probabilities of each ordered outcome, according to Equation (2):

   - $\Pr(Y=1) = \Lambda(2.77-5.04) - 0 = \Lambda(-2.27) = \frac{\exp(-2.27)}{1+\exp(-2.27)} = \mathbf{0.09}$.
   - $\Pr(Y=2) = \Lambda(4.27-5.04) - \Lambda(2.77-5.04) = \Lambda(-0.77) - \Lambda(-2.27) = \frac{\exp(-0.77)}{1+\exp(-0.77)} - \frac{\exp(-2.27)}{1+\exp(-2.27)} = 0.32 - 0.09 = \mathbf{0.23}$.
   - $\Pr(Y=3) = \Lambda(5.58-5.04) - \Lambda(4.27-5.04) = \Lambda(0.54) - \Lambda(-0.77) = \frac{\exp(0.54)}{1+\exp(0.54)} - \frac{\exp(-0.77)}{1+\exp(-0.77)} = 0.63 - 0.32 = \mathbf{0.31}$.
   - $\Pr(Y=4) = 1 - \Lambda(5.58-5.04) = 1 - \Lambda(0.54) = 1 - \frac{\exp(0.54)}{1+\exp(0.54)} = 1 - 0.63 = \mathbf{0.37}$.

Note that the probabilities (naturally) sum to one. Also, if we’d done an ordered probit, we’d use a standard normal CDF (that is, Φ) rather than Λ.

Changes in Predicted Probabilities

To get changes in probabilities associated with changes in binary Xs, we just calculate the predicted probability for each value $\{0, 1\}$ and take the difference. So, think about microbrews...

1. The new index value (holding all other variables at their means, but changing craftbeer to one) is just $-0.45 \times 4.96 + 0.047 \times 142 - 1.70 \times 1 - 0.03 \times 35.4 + 0.05 \times 33.1 = \mathbf{3.34}$.

2. The corresponding respective probabilities are:

   - $\Pr(Y=1) = \Lambda(2.77-3.34) - 0 = \mathbf{0.36}$.
   - $\Pr(Y=2) = \Lambda(4.27-3.34) - \Lambda(2.77-3.34) = 0.72 - 0.36 = \mathbf{0.36}$.
   - $\Pr(Y=3) = \Lambda(5.58-3.34) - \Lambda(4.27-3.34) = 0.90 - 0.72 = \mathbf{0.18}$.
   - $\Pr(Y=4) = 1 - 0.90 = \mathbf{0.10}$.

3. So, the changes in predicted probabilities as one goes from a “regular” beer to a microbrew are:

Outcome	Change in Probability
ΔPr(Fair)	0.27
ΔPr(Good)	0.13
ΔPr(Very Good)	-0.13
ΔPr(Excellent)	-0.27

We can do the same thing for a discrete-valued change in a continuous variable. In general, however, changes in probabilities associated with changes in continuous covariates are best illustrated graphically, rather than in table form.

Graphing Continuous Predicted Probabilities

Plots of Predicted Probabilities

Tables of changes are fine, but for continuous covariates, graphs are even better. Consider the effect of changes in `calories`.

1. Start by generating another "dummy" dataset, where `calories` varies across its range of actual values and all other variables are held constant at their means:

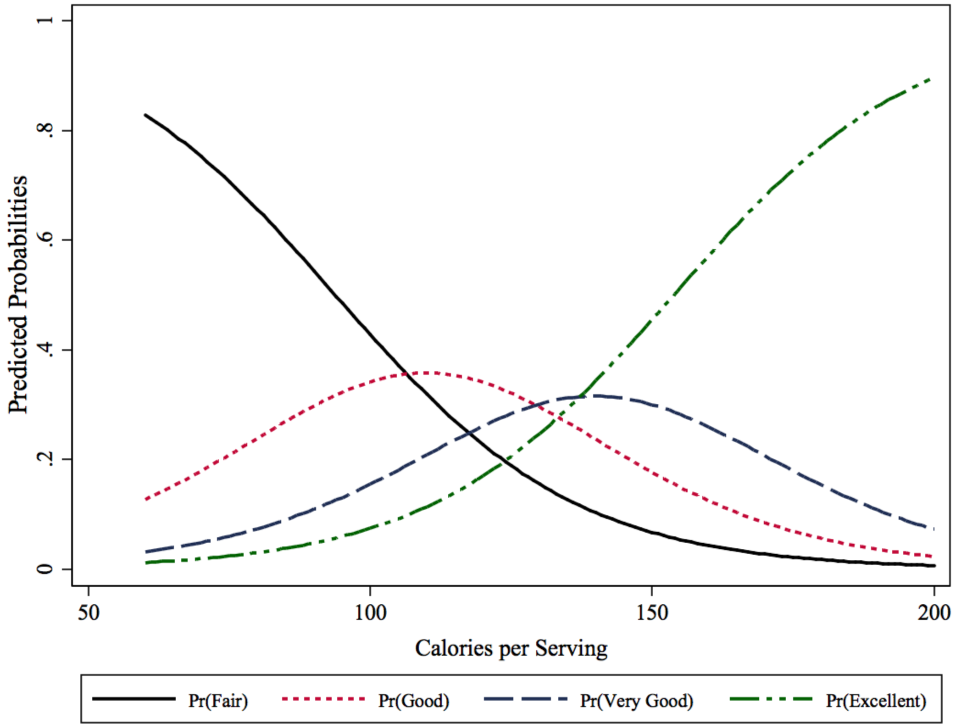
```
. clear
. set obs 141
obs was 0, now 141
. gen calories=_n+59
. gen price=4.96
. gen craftbeer=0
. gen bitter=35.44
. gen malty=33.13
```

2. Then, predict from the estimation data to the simulated data:

```
. save beersim
file beersim.dta saved

. use beerdata.dta

. ologit quality price calories craftbeer bitter malty

(output omitted)

. use beersim.dta

. predict ProbFair ProbGood ProbVG ProbExc
(option pr assumed; predicted probabilities)
```

3. Finally, we can graph these predictions:[2]

Figure 2: Changes in $\Pr(Y = j)$ as a Function of `calories`



Note several things about Figure 2...

- It looks like the "pieces" of a *PDF* (which, in fact, it is) that's been "cut up."
- `calories`' effects on each of the categories reflect the discussion above...
  - The probability of a "Fair" rating decreases with `calories` (from 0.83 to 0.01).
  - The probability of an "Excellent" rating increases from 0.01 to 0.90.
  - The probability of a "Good" rating first increases (from 0.13 to 0.36), reaching its maximum at about 110 calories, and then decreases to 0.02.
  - Similarly, the probability of a "Very Good" rating first increases (from 0.03 to 0.32), reaching its maximum at about 140 calories, and then decreases to 0.07.
    - These last two result reflects that ambiguous first derivative for the "middle" categories.
    - They also make substantive sense: at one level, you expect more caloric beers to be more likely to be "good;" but at much higher calorie levels, you'd be even more likely to think that they'd be "very good" or "excellent" (and so, correspondingly less likely to be "good").

[2] Note: Stata syntax for the plots shown in this section is available at the end of these notes.

Cumulative predicted probabilities

The cumulative probabilities are just the predicted probabilities “stacked” on top of one another. So, Pr(Fair) is just the predicted Pr(Fair); Pr(Fair or Good) is Pr(Fair) + Pr(Good); Pr(Fair or Good or Very Good) is Pr(Fair) + Pr(Good) + Pr(Very Good), and
Pr(Fair or Good or Very Good or Excellent) is 1.0. We can generate and plot these, as follows:

```
. gen CDzero=0

. gen CDFair=ProbFair

. gen CDGood=ProbFair + ProbGood

. gen CDVG=ProbFair + ProbGood + ProbVG

. gen CDExcellent=ProbFair + ProbGood + ProbVG + ProbExc
```

Figure 3: Changes in Cumulative Probabilities, as a Function of `calories`

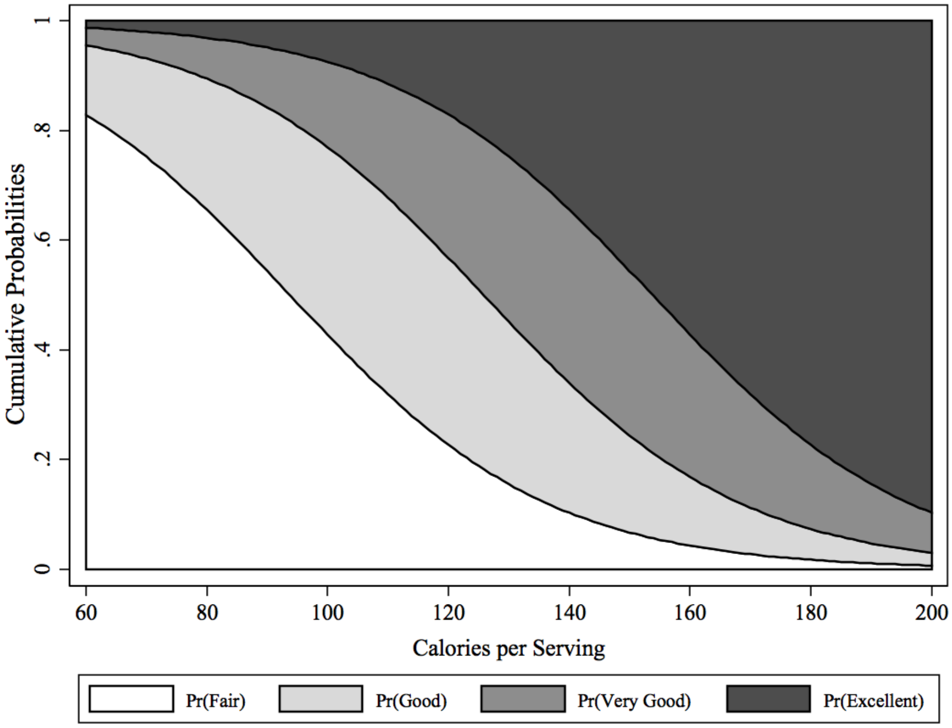


Figure 3 gives you the same information as the previous chart, but some people find the latter more intuitive. In particular, one can think of the area between the various curves as the probabilities of each of the outcomes. So, its very clear what happens as `calories` increases – first your chance of being rated "good" goes up, then (as `calories` gets even higher) your chances of being rated "very good" (and then "excellent") go up markedly.

Other Approaches

Beyond these, there are several other tools that can help you interpret the results of ordered-response models. All of these are for Stata users; the SAS, R, and other software users out there will have to do their own digging around.

- Both `ologit` and `oprobit` support a bunch of postestimation commands. We discussed `mfx` and `predict`, but other useful ones include `fitstat`, `lincom`, `nlcom`, and the various flavors of `test`. Check them out.

- As with binary-response models, King et al.'s Clarify program can provide useful post-estimation information, particularly in the form of changes in predicted probabilities and their associated measures of uncertainty. That software works for ordinal models more-or-less like it does for binary-response models.

- Scott Long and Jeremy Freese have written a suite of commands under the general heading `spost`. They automate several useful things, including predicted probabilities and the like. We'll be using one of these, `brant`, in a few minutes. Check them out.

Parallel Regressions

We said last time that the parallel regressions assumption can be an important one, both statistically and substantively. Let's revisit our model:

```
. ologit quality price calories craftbeer bitter malty

Ordered logistic regression                       Number of obs   =         69
                                                  LR chi2(5)      =      27.29
                                                  Prob > chi2     =     0.0000
Log likelihood = -81.871058                       Pseudo R2       =     0.1429
------------------------------------------------------------------------------
     quality |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       price |  -.4505038   .2933979    -1.54   0.125    -1.025553    .1245455
    calories |   .0467436   .0121975     3.83   0.000      .022837    .0706502
   craftbeer |  -1.704671    .941856    -1.81   0.070    -3.550675     .141333
      bitter |  -.0296914    .042417    -0.70   0.484    -.1128272    .0534444
       malty |   .0505279   .0245936     2.05   0.040     .0023254    .0987304
-------------+----------------------------------------------------------------
       /cut1 |   2.771035   1.673818                     -.5095873    6.051657
       /cut2 |   4.269676    1.72483                      .8890725     7.65028
       /cut3 |   5.577651   1.759538                      2.129021    9.026282
------------------------------------------------------------------------------
```

As we noted previously, the fact that we estimate only one parameter (β) for each covariate implies a restriction: that

$$\frac{\partial \Pr(Y_i = j)}{\partial X} = \frac{\partial \Pr(Y_i = j')}{\partial X} \; \forall \; j \neq j' \tag{3}$$

This is equivalent to a restriction on the βs; i.e., that all the β_ks for a given covariate X_k are equal across the $J-1$ categories. Recall that the basic probability statement for the standard ordered regression model we've been talking about is:

$$\Pr(Y_i = j|\mathbf{X}, \boldsymbol{\beta}) = F(\tau_j - \mathbf{X}_i\boldsymbol{\beta}) - F(\tau_{j-1} - \mathbf{X}_i\boldsymbol{\beta}) \tag{4}$$

We can imagine a generalization:

$$\Pr(Y_i = j|\mathbf{X}, \boldsymbol{\beta}) = F(\tau_j - \mathbf{X}_i\boldsymbol{\beta}_j) - F(\tau_{j-1} - \mathbf{X}_i\boldsymbol{\beta}_j) \tag{5}$$

This "generalized" ordered regression model estimates a separate vector of coefficients $\boldsymbol{\beta}_j$ for each of the $J-1$ categories of the response variable. One way of thinking about this is to imagine that we divide up the J categories into $J-1$ dummy variables, and estimate separate regressions of $\mathbf{X}$ on each. Consider our beer example. We can estimate a model of whether (=1) or not (=0) a beer was rated "Good" or better (with "Fair" being the zero category):

```
. gen goodplus=(quality>1)

. logit goodplus price calories craftbeer bitter malty

Logistic regression                               Number of obs   =         69
                                                  LR chi2(5)      =      15.40
                                                  Prob > chi2     =     0.0088
Log likelihood = -30.822655                       Pseudo R2       =     0.1999
------------------------------------------------------------------------------
    goodplus |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       price |  -.1216719   .4144182    -0.29   0.769    -.9339166    .6905727
    calories |   .0471238   .0148004     3.18   0.001     .0181156     .076132
   craftbeer |     .32331   1.375893     0.23   0.814    -2.373391    3.020011
      bitter |  -.0768229   .0562343    -1.37   0.172      -.18704    .0333943
       malty |   .0222671   .0332503     0.67   0.503    -.0429023    .0874365
       _cons |  -2.893688   1.997235    -1.45   0.147    -6.808197    1.020821
------------------------------------------------------------------------------
```

If we were to do this for each of the remaining categories – omitting one as the "baseline" – we'd have a set of three binary-response regression models: one estimating the effect of $\mathbf{X}$ on $\Pr(Y > 1)$, one for $\Pr(Y > 2)$, and a third for $\Pr(Y > 3)$. With these results, we could:

1. "Eyeball" the coefficients, and see if they are more-or-less the same across the various parts of the ordinal variable.

2. Calculate a likelihood ratio test, by simply summing the separate (log)-likelihoods of the $J - 1$ binary models, subtracting the log-likelihood of the ordered model from that value, and multiplying by -2; that statistic is $\sim \chi^2_{K(J-2)}$, and amounts to a test for whether, in the aggregate, the parallel regressions assumption holds or not.

3. Long also gives formulas for Lagrangian multiplier and Wald tests for the global parallel regression assumption (1997, pp. 142-44).

4. The "Brant test"[3] is an automated version of the Wald test mentioned in the Long book; it also reports the results for variable-specific tests for parallel regressions:[4]

[3] Brant, R. 1990. "Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression." *Biometrics* 46:1171-78.

[4] The `brant` command is part of the `spost` suite of commands; you'll have to install those to use `brant`. Type `findit spost` for help on this, if necessary.

```
. brant, detail

Estimated coefficients from j-1 binary regressions

                  y>1          y>2          y>3
    price  -.12167192  -.66679785  -.94164044
 calories   .04712378   .05587705    .0534708
craftbeer   .32330996  -3.1416146  -2.8559674
   bitter  -.07682288  -.02821661   .04355989
    malty   .02226708    .0686074   .05110993
    _cons  -2.8936879  -4.6327928  -6.4437228

Brant Test of Parallel Regression Assumption

    Variable |      chi2   p>chi2    df
-------------+--------------------------
         All |     13.61    0.191    10
-------------+--------------------------
       price |      2.53    0.283     2
    calories |      0.22    0.894     2
   craftbeer |      5.59    0.061     2
      bitter |      3.46    0.178     2
       malty |      1.99    0.369     2
----------------------------------------

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

The interpretation of this is straightforward:

- The initial table are the estimated $\hat{\boldsymbol{\beta}}$s from the series of $J-1$ implicit binary regressions; note, for example, that the $\hat{\boldsymbol{\beta}}$s for the `y>1` category are identical to those in the binary logit of `goodplus`, above.

- The test then calculates both variable-specific and global Wald tests for the restriction in Eq. (4), and reports the test statistics and P-values.

  - Significant values are evidence that the $\boldsymbol{\beta}$s are non-constant across different categories.
  - Non-significant test values suggest that the parallel regressions assumption is a reasonable one.

“Generalized” Ordered Logit

Of course, given the potential for violation of the parallel regressions assumption, one might ask “why not just estimate a model like (5)?” – i.e., one that allows each β to be different for each category of Y. In fact, such a “generalized” ordered logit model is easy to estimate:

```
. gologit2 quality price calories craftbeer bitter malty

Generalized Ordered Logit Estimates               Number of obs   =         69
                                                  LR chi2(15)     =      45.06
                                                  Prob > chi2     =     0.0001
Log likelihood = -72.986987                       Pseudo R2       =     0.2359

------------------------------------------------------------------------------
     quality |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1            |
       price |  -.0835826   .3804222    -0.22   0.826    -.8291963    .6620312
    calories |   .0467468   .0149238     3.13   0.002     .0174966    .0759969
   craftbeer |   .7262383   1.454381     0.50   0.618    -2.124295    3.576772
      bitter |  -.0928178   .0553309    -1.68   0.093    -.2012644    .0156288
       malty |     .02204   .0355225     0.62   0.535    -.0475829    .0916628
       _cons |  -2.679875   1.986657    -1.35   0.177    -6.573651    1.213901
-------------+----------------------------------------------------------------
2            |
       price |  -.2712559   .3807465    -0.71   0.476    -1.017505    .4749936
    calories |   .0602319   .0205639     2.93   0.003     .0199273    .1005364
   craftbeer |  -2.944466   1.415597    -2.08   0.038    -5.718986   -.1699462
      bitter |  -.0481728   .0627195    -0.77   0.442    -.1711006    .0747551
       malty |    .059636   .0427353     1.40   0.163    -.0241237    .1433957
       _cons |   -6.32298   2.977815    -2.12   0.034    -12.15939   -.4865707
-------------+----------------------------------------------------------------
3            |
       price |  -1.394406   .5978274    -2.33   0.020    -2.566126   -.2226854
    calories |   .0831619   .0377862     2.20   0.028     .0091024    .1572214
   craftbeer |  -3.363712    1.77395    -1.90   0.058     -6.84059    .1131653
      bitter |   .1537367   .0958194     1.60   0.109    -.0340658    .3415392
       malty |    .006209   .0496373     0.13   0.900    -.0910783    .1034962
       _cons |  -10.81515   5.361544    -2.02   0.044    -21.32359   -.3067199
------------------------------------------------------------------------------
```

This model simply relaxes the restriction in (4), and estimates the distinct $\boldsymbol{\beta}$s for each ordinal category. Interpretation, while somewhat more complicated, is identical to the standard ordered logit model, with the exception that odds ratios are now calculated separately for each category of Y.

A "Partially" Generalized Model

While the model in Equation (5) is certainly flexible, it also seems awfully inefficient. Consider the `calories` variable. By simply looking at the estimated βs, we can see that `calories`' effect seems to be more-or-less the same across the three regressions / categories. What might be useful, then, would be a model that was "partially" generalized. Suppose in $\mathbf{X}$ there are two kinds of variables:

1. For one subset of $\mathbf{X}$ – call it $\mathbf{X}'$ – the effects of $\mathbf{X}'$ on $\Pr(Y = j)$ are parallel; that is, those variables' effects meet the parallel regression assumption.

2. For another subset of $\mathbf{X}$ – call this one $\mathbf{X}''$ – the effects are not parallel; those variable violate the parallel regression assumption.

Our ideal model, then would look like:

$$\Pr(Y_i = j|\mathbf{X}, \boldsymbol{\beta}) = F(\tau_j - \mathbf{X}'_i\boldsymbol{\beta} - \mathbf{X}''_i\boldsymbol{\beta}_j) - F(\tau_{j-1} - \mathbf{X}'_i\boldsymbol{\beta} - \mathbf{X}''_i\boldsymbol{\beta}_j) \tag{6}$$

This model estimates separate, category-specific βs for those variables that "need" them (that is, for those whose effects are not parallel), and estimates a single β for variables whose effects meet the parallel regression assumption.

Of course, one challenge is knowing which variables are in $\mathbf{X}'$ and which are in $\mathbf{X}''$. We can use the Brant test to get an assessment of this; but it's also possible to "autofit" this model, using `gologit2`:

```
. gologit2 quality price calories craftbeer bitter malty, autofit

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

Step  1:  Constraints for parallel lines imposed for calories (P Value = 0.6283)
Step  2:  Constraints for parallel lines imposed for malty (P Value = 0.5639)
Step  3:  Constraints for parallel lines imposed for price (P Value = 0.1376)
Step  4:  Constraints for parallel lines imposed for craftbeer (P Value = 0.0880)
Step  5:  Constraints for parallel lines are not imposed for
          bitter (P Value = 0.04536)

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

 ( 1)  [1]calories - [2]calories = 0
 ( 2)  [1]malty - [2]malty = 0
 ( 3)  [1]price - [2]price = 0
 ( 4)  [1]craftbeer - [2]craftbeer = 0
 ( 5)  [1]calories - [3]calories = 0
 ( 6)  [1]malty - [3]malty = 0
 ( 7)  [1]price - [3]price = 0
 ( 8)  [1]craftbeer - [3]craftbeer = 0

           chi2(  8) =    9.88
         Prob > chi2 =    0.2736

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

------------------------------------------------------------------------------
Generalized Ordered Logit Estimates               Number of obs   =         69
                                                  Wald chi2(7)    =      24.73
                                                  Prob > chi2     =     0.0008
Log likelihood = -79.280813                       Pseudo R2       =     0.1700

 ( 1)  [1]calories - [2]calories = 0
 ( 2)  [1]malty - [2]malty = 0
 ( 3)  [1]price - [2]price = 0
 ( 4)  [1]craftbeer - [2]craftbeer = 0
 ( 5)  [2]calories - [3]calories = 0
 ( 6)  [2]malty - [3]malty = 0
 ( 7)  [2]price - [3]price = 0
 ( 8)  [2]craftbeer - [3]craftbeer = 0
```

```
------------------------------------------------------------------------------
     quality |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1            |
       price |  -.3974502   .2963908    -1.34   0.180    -.9783655    .1834652
    calories |   .0527095    .013121     4.02   0.000     .0269927    .0784263
   craftbeer |  -1.672272   .9492114    -1.76   0.078    -3.532692    .1881483
      bitter |  -.0499689   .0444101    -1.13   0.261    -.1370111    .0370733
       malty |   .0429082   .0245989     1.74   0.081    -.0053049    .0911212
       _cons |  -2.889296   1.680023    -1.72   0.085    -6.182081    .4034878
-------------+----------------------------------------------------------------
2            |
       price |  -.3974502   .2963908    -1.34   0.180    -.9783655    .1834652
    calories |   .0527095    .013121     4.02   0.000     .0269927    .0784263
   craftbeer |  -1.672272   .9492114    -1.76   0.078    -3.532692    .1881483
      bitter |  -.0388347   .0442736    -0.88   0.380    -.1256092    .0479399
       malty |   .0429082   .0245989     1.74   0.081    -.0053049    .0911212
       _cons |  -4.833708   1.889462    -2.56   0.011    -8.536985   -1.130431
-------------+----------------------------------------------------------------
3            |
       price |  -.3974502   .2963908    -1.34   0.180    -.9783655    .1834652
    calories |   .0527095    .013121     4.02   0.000     .0269927    .0784263
   craftbeer |  -1.672272   .9492114    -1.76   0.078    -3.532692    .1881483
      bitter |    -.00471   .0441368    -0.11   0.915    -.0912166    .0817966
       malty |   .0429082   .0245989     1.74   0.081    -.0053049    .0911212
       _cons |    -7.4055   2.038188    -3.63   0.000    -11.40027   -3.410725
------------------------------------------------------------------------------
```

The model starts by conducting a series of tests for whether or not each variable violates the parallel regression assumption. If it does, then it is left in the model unconstrained; if not, the equality constraint is imposed.

Here, the only variable that violates the proportional regression assumption is that for `bitter`ness. All other coefficients are constrained to be the same across the four categories, as in a standard ordered logit.

Other Variants on Ordered Response Models

We'll wrap up with a very quick discussion of two models that generalize the standard ordered logit/probit, in straightforward ways.

Heteroscedastic Ordered Probit

Alvarez and Brehm (1998), pulling from Greene (2002) and other work, extend the heteroscedastic probit model to the case of multiple ordered outcomes. The Appendix of their article lists the details; suffice it to say that the log-likelihood is:

$$\ln L = \sum_{i=1}^{N} \sum_{j=1}^{J} \delta_{ij} \ln \left[\Phi \left(\frac{\tau_j - \mathbf{X}_i \boldsymbol{\beta}}{\exp(\mathbf{Z}_i \gamma)} \right) - \Phi \left(\frac{\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta}}{\exp(\mathbf{Z}_i \gamma)} \right) \right] \tag{7}$$

where the δ_{ij}s are indicator variables as before, and $\mathbf{Z}_i\gamma$ are the variables believed to influence the variability of the latent variable Y^*. The intuition of this model is largely the same as in the binary case; we won't spend a bunch of time on it.

Variable "Cut-Points"

In a paper in *Political Analysis*, Mitch Sanders extends the ordered models even further, by allowing for both heteroscedasticity and for the "cut points" τ to vary across different individuals. The substantive issue is vote choice (or non-choice / abstention). Tinkering with his notation a bit,[5], we have a set of variables $\mathbf{X}_i\boldsymbol{\beta}$ that influence the mean of Y^*, another set $\mathbf{Z}_i\gamma$ that affect its variance, and a third set of variables (I'll call them $\mathbf{W}_i\eta$) that influence where the "cut-points" are (that is, each individuals' "decision threholds").

The model is a three-category ordinal response model; the basic probability statements are:

$$\Pr(Y_i = 1) = 1 - \Phi \left(\frac{\mathbf{W}_i\eta - \mathbf{X}_i \boldsymbol{\beta}}{\exp(\mathbf{Z}_i \gamma)} \right), \tag{8}$$

$$\Pr(Y_i = 2) = \Phi \left(\frac{\mathbf{W}_i\eta - \mathbf{X}_i \boldsymbol{\beta}}{\exp(\mathbf{Z}_i \gamma)} \right) - \Phi \left(\frac{-\mathbf{W}_i\eta - \mathbf{X}_i \boldsymbol{\beta}}{\exp(\mathbf{Z}_i \gamma)} \right), \text{and} \tag{9}$$

$$\Pr(Y_i = 3) = \Phi \left(\frac{-\mathbf{W}_i\eta - \mathbf{X}_i \boldsymbol{\beta}}{\exp(\mathbf{Z}_i \gamma)} \right). \tag{10}$$

[5] Mitch calls the variables that influence the "cut-points" $\mathbf{Z}_i\gamma$, but for consistency I'll leave those as the variance terms.

The log-likelihood then just becomes:

$$\ln L = \sum_{i=1}^{N} \sum_{j=1}^{J} \delta_{ij} \ln \left[\Phi \left(\frac{\mathbf{W}_i \eta - \mathbf{X}_i \boldsymbol{\beta}}{\exp(\mathbf{Z}_i \gamma)} \right) - \Phi \left(\frac{-\mathbf{W}_i \eta - \mathbf{X}_i \boldsymbol{\beta}}{\exp(\mathbf{Z}_i \gamma)} \right) \right] \tag{11}$$

where δ_{ij} is again defined as before. Note a few things about this model:

- The model has two "cut points," symmetrically located at $\mathbf{W}_i \eta$ and $-\mathbf{W}_i \eta$; scaling the parameters/variables thus changes the distance between them.
- This latter point is how the model is identified; changes in $\mathbf{W}_i \eta$ are constrained to affect the cut-points in a particular way, making them differentiable from the effects of $\mathbf{X}_i \boldsymbol{\beta}$.

Sanders does this in GAUSS, though in theory there's no reason one couldn't build a model like this in Stata/S-Plus/R/BUGS/whatever (and, in fact, there is now a way to estimate such a model in Stata ; drop me an e-mail if you think it might be useful for your work).

Appendix: **Stata** `graph` Commands for Figures

Note: These were – like all the figures you see in these notes – generated using the `Graphics` pull-down menu in Stata 9.2; the commands here are just cut-and-pasted from the `Results` window.

Figure 2

```
. twoway (line ProbFair calories, lcolor(black) lpattern(solid) lwidth(medthick))
(line ProbGood calories, lcolor(cranberry) lpattern(shortdash) lwidth(medthick))
(line ProbVG calories, lcolor(dknavy) lpattern(longdash) lwidth(medthick))
(line ProbExc calories, lcolor(dkgreen) lpattern(longdash_shortdash)
lwidth(medthick)), ytitle(Predicted Probabilities) xtitle(Calories per Serving,
size(medsmall)) legend(cols(4) order(1 "Pr(Fair)" 2 "Pr(Good)" 3 "Pr(Very
Good)" 4 "Pr(Excellent)") size(small)) graphregion(margin(vsmall))
```

Figure 3

```
. twoway (rarea CDFair CDzero calories, lcolor(black) lpattern(solid) fcolor(white))
(rarea CDGood CDFair calories, lcolor(black) lpattern(solid) fcolor(gs13)) (rarea
CDVG CDGood calories, lcolor(black) lpattern(solid) fcolor(gs7)) (rarea CDExcellent
CDVG calories, lcolor(black) lpattern(solid) fcolor(gs2)), ytitle(Cumulative
Probabilities) xtitle(Calories per Serving, size(medsmall)) xscale(range(60. 200.))
xlabel(60(20)200) legend(cols(4) order(1 "Pr(Fair)" 2 "Pr(Good)" 3 "Pr(Very
Good)" 4 "Pr(Excellent)") size(small)) graphregion(margin(vsmall))
```