Ordered Probit

Econ 674

Purdue University

March 9, 2009

In some cases, the variable to be modeled has a natural *ordinal* interpretation.

Some examples include:

1. Education, measured categorically, (*e.g.* $1 =<$ HS, $2 = HS$, $3 =$ Some college, etc.).
2. Income, also measured categorically.
3. Survey responses, coded as a degree of opinion (*e.g.* $1 =$ Strongly Disagree, $2=$ Disagree, $3 =$ Agree, $4 =$ Strongly Agree.
4. These are in contrast to other choices such as type of insurance or selected mode of transportation, for example, that are not ordered.

- In this lecture we discuss *ordinal choice* models, and focus on the *ordered probit* in particular.

The Ordered Probit Model

Suppose that the variable to be modeled, y takes on J different values, which are naturally ordered:

$$y_i = \begin{cases} 1 \\ 2 \\ \vdots \\ J \end{cases}, \qquad i = 1, 2, \ldots, n.$$

As with the probit model, we assume that the observed y is generated by a latent variable y^*, where

-

The link between the latent and observed data is given as follows:

-

The Ordered Probit Model

The α_j are called *cutpoints* or *threshold* parameters. They are estimated by the data and help to match the probabilities associated with each discrete outcome.

Without any additional structure, the model is not identified. In particular, there are too many cutpoints and some restrictions are required. The most common way to achieve identification is to set:

-

and retain an intercept parameter in the model.

What happens, for example when $J = 2$ and these restrictions are imposed?

The Ordered Probit Model

The likelihood for the ordered probit is simply the product of the probabilities associated with each discrete outcome:

$$\overline{L}(\beta, \alpha) = \prod_{i=1}^{n} \Pr(y_i = j | x_i),$$

where

$$\alpha = [\alpha_3 \quad \alpha_4 \quad \cdots \quad \alpha_J].$$

The i^{th} observation's contribution to the likelihood is

-

The Ordered Probit Model

Therefore,

$$\overline{L}(\beta, \alpha) = \prod_{i=1}^{n} \Phi(\alpha_{y_i+1} - x_i\beta) - \Phi(\alpha_{y_i} - x_i\beta)$$

and

-

For purposes of computing the MLE, it can be useful to define

$$Z_{ij} = I(y_i = j).$$

Thus, we can write:

$$L(\beta, \alpha) = \sum_{i=1}^{n} \sum_{j=1}^{J} z_{ij} \left(\log \left[\Phi(\alpha_{j+1} - x_i\beta) - \Phi(\alpha_j - x_i\beta) \right] \right).$$

(Some textbooks present the material this way, though we will not make use of this here).

The Ordered Probit Model

This yields the score for the parameter vector β:

-

and likewise, we obtain the FOC for $\alpha_k, \; k = 3, 4, \cdots, J$:

$$L_{\alpha_k}(\beta, \alpha) = \sum_{i: y_i = k} - \frac{\phi(\alpha_k - x_i\beta)}{\Phi(\alpha_{k+1} - x_i\beta) - \Phi(\alpha_k - x_i\beta)} + \sum_{i: y_i = k-1} \frac{\phi(\alpha_k - x_i\beta)}{\Phi(\alpha_k - x_i\beta) - \Phi(\alpha_{k-1} - x_i\beta)}$$

We do not report the Hessian here, as the expressions are rather lengthy. Nonetheless, standard MLE can be applied.

Marginal Effects

To fix ideas, consider the case of an ordered probit model with $J = 3$, in which case we have:

-

From these, we obtain the category-specific *marginal effects*:

-

Marginal Effects

What do we learn from this simple model?

1. Like the probit, the marginal effects depend on x. We can evaluate these at sample means, or take a sample average of the marginal effects.

2. Unlike the probit, *the signs of the "interior" marginal effects are unknown* and not completely determined by the sign of β_k.

3. We can, however, sign the effects of the lowest and highest categories based on β_k. The others, however, can not be known by the reader simply by looking at a table of point estimates.

Interpretation

Continue to consider the case with $J = 3$ and suppose there are no covariates and only an intercept parameter is included.
In this case we have

$$\begin{aligned}\Pr(y_i = 1) &= 1 - \Phi(\beta) \\ \Pr(y_i = 2) &= \Phi(\alpha - \beta) - [1 - \Phi(\beta)] = \Phi(\alpha - \beta) - \Phi(-\beta) \\ \Pr(y_i = 3) &= 1 - \Phi(\alpha - \beta)\end{aligned}$$

What do you think will happen in terms of the MLE's?
The likelihood is:

$$\overline{L}(\alpha, \beta) = \prod_{i:y_i=1} [1 - \Phi(\beta)] \prod_{i:y_i=2} [\Phi(\alpha - \beta) - \Phi(-\beta)] \prod_{i:y_i=3} [1 - \Phi(\alpha - \beta)]$$

which reduces to

-

Interpretation

In the last slide, we have defined n_j as the number of observations for which $y_i = j$ and also note that $n_1 + n_2 + n_3 = n$.
Thus we obtain the log-likelihood

$$L(\alpha, \beta) = n_1 \log[1 - \Phi(\beta)] + n_2 \log[\Phi(\alpha - \beta) - \Phi(-\beta)] + n_3 \log[1 - \Phi(\alpha - \beta)].$$

The α FOC gives:

-

with $\hat{P}_j$ denoting the fitted probability for category j.

Interpretation

Likewise, we get an FOC for β:

$$-n_1 \frac{\phi(\hat{\beta})}{1-\Phi(\hat{\beta})} + n_2 \frac{\phi(\hat{\beta}) - \phi(\hat{\alpha}-\hat{\beta})}{\Phi(\hat{\alpha}-\hat{\beta}) - \Phi(-\hat{\beta})} + n_3 \frac{\phi(\hat{\alpha}-\hat{\beta})}{1-\Phi(\hat{\alpha}-\hat{\beta})} = 0.$$

Grouping terms, and using our α FOC, the β FOC can be shown to imply:

$$\frac{n_2}{\hat{P}_2} = \frac{n_1}{\hat{P}_1}.$$

interpretation

Noting that

$$\hat{P}_1 + \hat{P}_2 + \hat{P}_3 = 1$$

and

$$n_1 + n_2 + n_3 = n,$$

these two FOC's can be manipulated to yield:

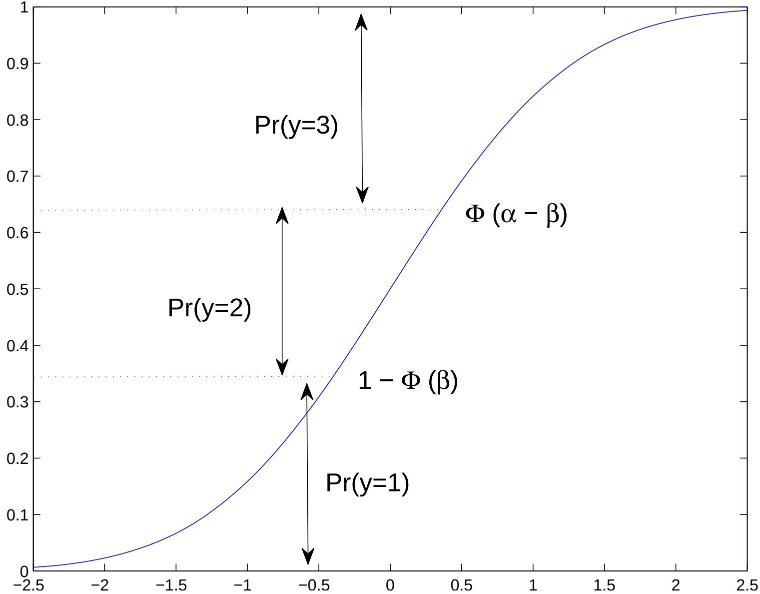
-

That is, the parameters will be selected so that the fitted values of each category exactly match the observed frequencies of outcomes in that category. Note that this generalizes to any J *and any link function!*
For $\hat{\beta}$, for example, we obtain:

$$1 - \Phi(\hat{\beta}) = n_1/n$$

or

$$\hat{\beta} = \Phi^{-1}\left(\frac{n - n_1}{n}\right).$$


Pr(y=3)
Φ (α − β)
Pr(y=2)
1 − Φ (β)
Pr(y=1)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5

Ordered Probit and the EM Algorithm

Suppose that $J = 3$ and consider the following model:

$$y_i^* = x_i\beta + \epsilon_i, \quad \epsilon_i | x_i \overset{iid}{\sim} \mathcal{N}(0,1)$$

and

$$y_i = \begin{cases} 1 & \text{if } y_i^* \leq 0 \\ 2 & \text{if } 0 < y_i^* \leq \alpha \\ 3 & \text{if } y_i^* > \alpha. \end{cases}$$

Let

-

This reparameterization defines an equivalent model:

-

where

-

Note that, in this representation, there are *no unknown cutpoints*.

Ordered Probit and the EM Algorithm

Step 1: E-Step
Note that

-

and thus

$$L(\delta, \sigma^2; z^*) = constant - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(z^* - X\delta)'(z^* - X\delta).$$

The E-step is completed by taking expectations over $z^*|\theta = \theta_t, y$:

$$E[L(\delta, \sigma^2; z^*)] \equiv constant - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}E_{z^*|\theta=\theta_t,y}(z^* - X\delta)'(z^* - X\delta).$$

Ordered Probit and the EM Algorithm

Step 2: M-Step:
To implement the $M-$ step, we must evaluate this expectation and then maximize over δ and σ^2.

You will probably recognize the δ-part of this exercise. It will follow similarly to the probit, where:

-

with

-

To evaluate this mean, suppose

-

and we seek

-

That is,

-

Ordered Probit and the EM Algorithm

Applying this result, we can evaluate $\mu(\delta_t, \sigma_t^2, y)$

$$\mu(\delta_t, \sigma_t^2, y_i) = \begin{cases} x_i\delta_t - \sigma \frac{\phi(x_i\delta_t/\sigma_t)}{\Phi(-x_i\delta_t/\sigma_t)} & \text{if } y_i = 1 \\ x_i\delta_t + \sigma \frac{\phi(x_i\delta_t/\sigma_t) - \phi([1-x_i\delta_t]/\sigma_t)}{\Phi([1-x_i\delta_t]/\sigma_t) - \Phi(-x_i\delta_t/\sigma_t)} & \text{if } y_i = 2 \\ x_i\delta_t + \sigma \frac{\phi([1-x_i\delta_t]/\sigma_t)}{1-\Phi([1-x_i\delta_t]/\sigma_t)} & \text{if } y_i = 3 \end{cases}$$

Thus, the parameters δ are easily updated.

Ordered Probit and the EM Algorithm

As for the updating of σ^2, note that it will be obtained as:

$$\sigma^2_{t+1} = \frac{1}{n} \sum_i E_{z_i^* | \delta = \delta_t, \sigma^2 = \sigma^2_t, y_i} \left(z_i^* - x_i \delta_{t+1} \right)^2 .$$

Expanding this out, we obtain (dropping the subscript on the expectation for simplicity):

$$\sigma^2_{t+1} = \frac{1}{n} \sum_i \left[E\left([z_i^*]^2 \right) - 2\mu(\delta_t, \sigma^2_t, y_i) x_i \delta_{t+1} + (x_i \delta_{t+1})^2 \right] .$$

Only the first term in the summation above requires further evaluation. We first note that

$$E\left([z_i^*]^2 \right) = (x_i \delta_t)^2 + 2 x_i \delta_t E(v_i | \delta_t, \sigma^2_t, y_i) + E(v_i^2 | \delta_t, \sigma^2_t, y_i).$$

We first recognize that

$$E(v_i | \theta_t, y_i) = \mu(\delta_t, \sigma^2_t, y_i) - x_i \delta_t .$$

As for the $E(v_i^2|\delta_t, \sigma_t^2, y_i)$ term, a little work gives:

$$E(v_i^2|\delta_t, \sigma_t^2, y_i) = \sigma^2 \left[1 + \frac{-x_i\delta_t/\sigma_t\phi(x_i\delta_t/\sigma_t) - [1 - x_i\delta_t]/\sigma_t\phi([1 - x_i\delta_t]/\sigma_t)}{\Phi([1 - x_i\delta_t]/\sigma_t) - \Phi(-x_i\delta_t/\sigma_t)} \right].$$

Ordered Probit and the EM Algorithm

So, we have everything we need to implement the EM algorithm. It would proceed as follows:

1. Pick some starting values.
2. Calculate $\mu(\delta_t, \sigma_t^2, y_i)$ using the formula provided. Use this to update δ_t to δ_{t+1}.
3. Calculate $E(v_i^2|\delta_t, \sigma_t^2, y_i)$ $\forall i$ using the formula provided, and use it [together with δ_{t+1} and $\mu(\delta_t, \sigma_t^2, y_i)$] to update σ_t^2 to σ_{t+1}^2.
4. Iterate to convergence.
5. Transform back by setting

$$\hat{\alpha} = \hat{\sigma}^{-1}, \quad \hat{\beta} = \hat{\delta}\hat{\alpha}.$$

Ordered Probit and the EM Algorithm

We use $n = 139$ law school applications from 1985.

The dependent variable is the rank of the law school with $y = 1$ if the rank is less than or equal to 25, $y = 2$ if the rank is between 25 and 50 and $y = 3$ if the rank exceeds 50.

The independent variables include the applicant's LSAT score, GPA and student/faculty ratio (the latter is rather questionable).

In the following slides, we present the EM ordered probit estimates (which matched STATA's EXACTLY and were obtained faster!) We report some statistics evaluated at the sample mean of the x's and also setting *LSAT* and *GPA* to their maximum sample values.

Ordered Probit and the EM Algorithm

$$\hat{\beta} = [49.1 \quad -.24 \quad -2.73 \quad -.01], \qquad \hat{\alpha} = 1.00.$$

Category	Fitted Probability		Marginal Effect		
	xbar	xmax	LSAT/xbar	LSAT/ xmax	GPA/ xbar
$y = 1$	.04	.99	.02	.003	.25
$y = 2$	.20	.01	.05	-.003	.59
$y = 3$	.76	.00	-.08	.000	-.85