ARE213 **Econometrics**

Spring 2006 UC Berkeley Department of Agricultural and Resource Economics

DISCRETE RESPONSE MODELS II:
ORDERED MULTINOMIAL RESPONSE MODELS

Now we consider discrete response models with more than two possible responses. In this lecture we limit ourselves to the case where the outcomes are ordered.

There are two important cases. In the first there is an underlying continuous variable but we only observe an indicator for a particular range. An example is earnings data that may come coded in intervals: 0, $[0, 10]$, $[10, 50]$, $[50, \infty)$. Another example is educational choices where outcomes may be coded as less than high school, high school, some college, more than college. Such data are referred to as interval-coded data. Typically we model the underlying continuous variable as linear and normal (or logistic) with a set of covariates:

$$Y_i^* = X_i'\beta + \varepsilon_i,$$

with the observed outcome an indicator for the interval for $j = 0, \ldots, J$:

$$Y_i = j \quad \text{if } \alpha_j \leq Y^* < \alpha_{j+1},$$

with $\alpha_{J+1} = \infty$, $\alpha_0 = -\infty$, and $\alpha_{j-1} < \alpha_j$. In this case the key assumption is that the boundaries α_j are known a priori. In this case we are typically interested in the conditional expectation of the latent outcome

$$\mathbb{E}[Y^*|X],$$

rather than in the distribution of the observed outcome,

$$\Pr(Y = j|X).$$

Another possibility arises when the responses are ordered but there is no clear mapping from the underlying continous response (which may itself be somewhat vague) to the discrete response. For example, respondents in a survey may be asked about their interest in a particular service and asked to respond in one of three categories: not interested, somewhat interested and very interested. In that case we may still wish to model the response through the same latent variable approach:

$$Y^* = X'\beta + \varepsilon,$$

$$Y_i = j \quad \text{if } \alpha_j \leq Y^* < \alpha_{j+1},$$

with $\alpha_{J+1} = \infty$, $\alpha_0 = -\infty$, and $\alpha_{j-1} < \alpha_j$. However, in this case we may not wish to impose the boundary values α_j a priori, and prefer to estimate them jointly with the regression parameters β. A second difference with the interval-coded case is that now we typically are interested in the distribution of the observed outcome.

In both cases the statistical model is the same, although the substantive interpretation differs. Let us assume for the moment that the ε_i are normal with mean zero and variance σ^2. The conditional probability that $Y = 0$ is

$$\Pr(Y = 0|X) = \Pr(X'\beta + \varepsilon < \alpha_1) = \Pr(\varepsilon/\sigma < (\alpha_1 - X'\beta)/\sigma) = \Phi((\alpha_1 - X'\beta)/\sigma).$$

For $0 < j < J$ the probability is

$$\Pr(Y = j|X) = \Pr(\alpha_j < X'\beta + \varepsilon < \alpha_{j+1}) = \Pr((\alpha_j - X'\beta)/\sigma < \varepsilon/\sigma < (\alpha_{j+1} - X'\beta)/\sigma)$$

$$= \Phi((\alpha_{j+1} - X'\beta)/\sigma) - \Phi((\alpha_j - X'\beta)/\sigma),$$

and for the last probability we have

$$\Pr(Y = J|X) = \Pr(\alpha_J < X'\beta + \varepsilon) = \Pr((\alpha_J - X'\beta)/\sigma < \varepsilon/\sigma) = 1 - \Phi((\alpha_J - X'\beta)/\sigma).$$

This way we can build up the log likelihood function as a function of β, σ^2, and α. Denote this by $L(\beta, \alpha, \sigma^2)$:

$$L(\beta, \alpha, \sigma^2) = \sum_{i=1}^{N} (1\{Y_i = 0\} \cdot \ln \Phi((\alpha_1 - X_i'\beta)/\sigma)$$

$$+ \sum_{j=1}^{J-1} 1\{Y_i = j\} \cdot \ln \left(\Phi((\alpha_{j+1} - X'\beta)/\sigma) - \Phi((\alpha_j - X'\beta)/\sigma) \right)$$

$$+1\{Y_i = J\} \cdot (1 - \Phi((\alpha_J - X'\beta)/\sigma)))\,.$$

There are a couple of caveats in maximizing the log likelihood function.

First consider the case where the α are known. If $J = 1$, so there are two choices, we cannot identify σ^2. We are back in the binary response world where the error variance is not identified, and we typically set it equal to unity. With $J > 1$, we can identify the variance separately and we do not need to normalize the parameters.

With J large we get approximately back to the case where we observe the actual value of the outcome. In fact, in all cases we only observe Y_i discretely, so you could argue that the ordered discrete response model is always appropriate. In practice if J is more than five or so people rarely bother using a discrete reponse model and just go ahead with the standard linear model.

Second, consider the case with α unknown. This is a much more difficult case, both in terms of computation and in terms of interpretation. Now there are a couple of normalizations. We cannot identify the intercept in β separately from the location of the boundary points α, so typically we normalize the intercept to zero. Second, we cannot identify the scale of the boundary points from the error variance so typically we normalize σ^2 to one.

Another issue for the case with unknow α is what we should report. The parameters β are not very useful in their own right, considerably less even than in the binary response model. They tell us whether the corresponding covariate is positively or negatively associated with the latent outcome, but that does not have a real meaning. A positive value for β also tells us that an increase in X is associated with a lower probability of $Y = 0$, and an increase in the probability that $Y = J$. It does not even tell us what the sign is on the probability that $Y = j$ for interior j. We can calculate the derivative of the probability with respect to the covariate, but this is different for all choices, and so there is potentially a lot to report. In practice you should look for meaningful things to report. Consider a specific policy, and the effect on the predicted probabilities for all the choices, and possibly do this for a couple of policy experiments.