# ECON 312

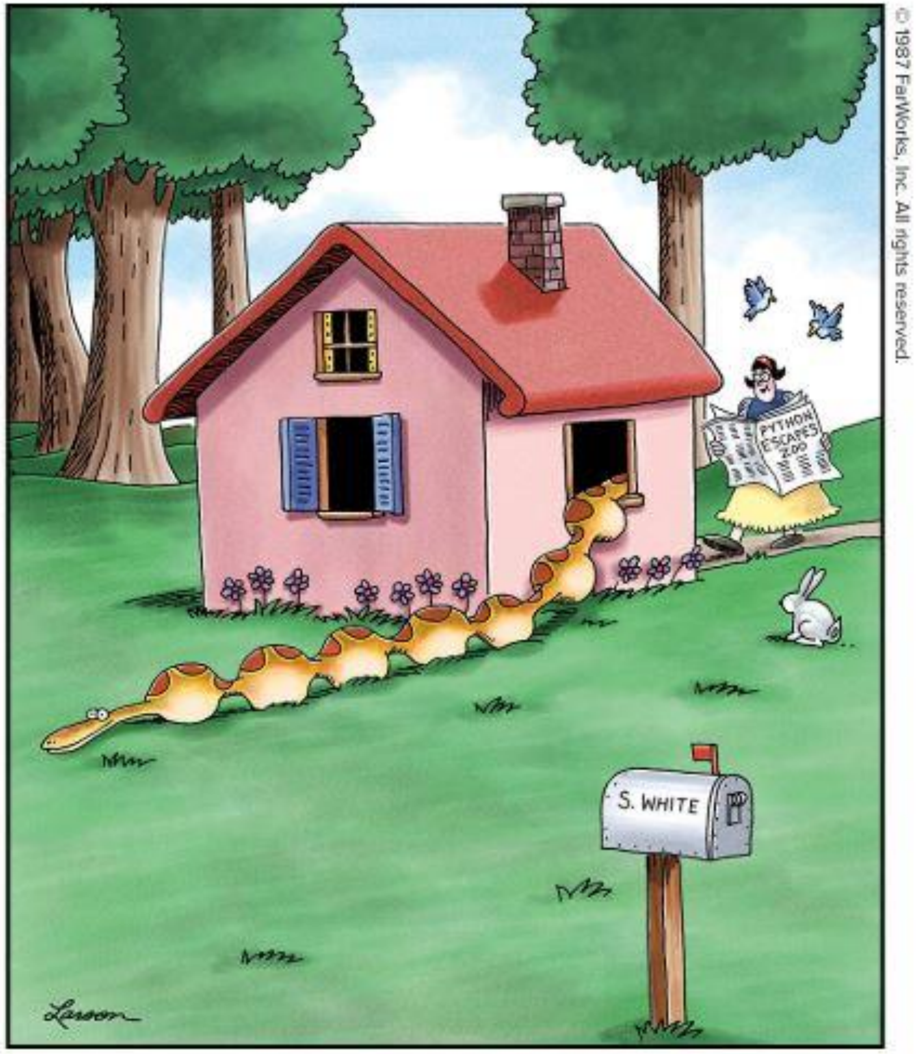
## Friday, April 24

## Discrete Choice, Ordered, and Count Variables

Readings: Wooldridge, Section 17.3

Class notes: 160 - 164

# Today's Far Side offering




Yup … seven humps!

# Context and overview

- In this class, we build on the probit and logit models to consider three other kinds of limited dependent variables
- All of the models are maximum-likelihood estimators

- More than two (unordered) discrete choices: multinomial (polytomous) logit
- Ordinal dependent variable: ordered probit or logit
- Count dependent variable: Poisson or negative binomial regression

# Multiple discrete choices

- Probit and logit analyze dependent variable with two choices (0/1)
- What if there are more than two?
  - Attend college A, B, or C
  - Major in sciences, humanities, or social sciences
- These choices are, in principle, unordered
  - No unidimensional rank-ordering is possible
- We model in similar way to probit/logit, but the model is more complex
  - **Multinomial logit** is the standard model
  - **Multinomial probit** also works, but is computationally very hard

# Multinomial logit

- Not in Wooldridge, but discussed in section 23.11 of Greene's 6th edition
- There are $M$ **discrete choices**
  - Each choice has its own, individual equation
  - We model probability of choice $j$ as $\Pr\left[y_i = j \mid x_i\right] = \dfrac{e^{x_i \beta_j}}{\sum_{m=1}^{M} e^{x_i \beta_m}}$
  - Note that the denominator is the sum of all the numerators for the $M$ values of $j$
  - This assures that the probabilities all sum to one
  - It means that we can only estimate equations for $M - 1$ of the choices

# Interpretation of coefficients

- There are $M(k + 1)$ parameters, but only $(M - 1)(k + 1)$ are independent
  - If a change in $x$ increases probability of choosing $y = 2, 3, \ldots, M$, then it must lower probability of choosing $y = 1$
  - We can determine all the elements of the $\beta_1$ vector from $\beta_2$ through $\beta_M$
- If $m = 1$ is omitted category, then

$$\ln\left(\frac{\Pr[y_i = m \mid x_i]}{\Pr[y_i = 1 \mid x_i]}\right) = x_i\beta_m$$

- $\beta_m$ is effect of $x$ on log-odds ratio of choice $m$ relative to choice 1

# Issues and related estimators

- One weakness of multinomial logit is that we must assume **independence of irrelevant alternatives**
  - Splitting one alternative into two must not affect the probability of choosing any of the others
  - Is that reasonable?
- **Conditional logit** allows the *x* variables to vary by choice as well as by person
- **Nested logit** models nested decisions, where making one choice opens up other sub-choices
- In Stata: mlogit, NOT clogit, nlogit, mprobit

# Ordinal dependent variables

- For **ordinal variables**, we can rank the outcomes on some scale, but don't know how far apart they should be
  - Bond ratings
  - Opinion-survey responses
- **Ordered probit** (for normal error term) and **ordered logit** (for logistic error term) estimate models with ordinal dependent variables
  - oprobit and ologit in Stata
  - Both are maximum-likelihood procedures

# Latent index

- Underlying **latent index variable** $y^*$ that is continuous
- We observe ordinal variable $y$ with $M$ levels that depends on $y^*$ and threshold values μ:

$$y_i = \begin{cases} 1 & \text{if } y_i^* \leq \mu_1, \\ 2 & \text{if } \mu_1 < y_i^* \leq \mu_2, \\ 3 & \text{if } \mu_2 < y_i^* \leq \mu_3, \\ \vdots & \\ M & \text{if } \mu_{M-1} < y_i^*. \end{cases}$$

- We assume $y_i^* = x_i\beta + u_i$ and estimate the β coefficients and μ thresholds

# Estimation and interpretation

- Divide by $\hat{\sigma}$ so that we can use standard normal $\Phi$:

$$\begin{aligned}
&\Pr\left[y_i = 1 \mid x_i\right] = \Phi\left(\mu_1 - x_i\beta\right) \\
&\Pr\left[y_i = 2 \mid x_i\right] = \Phi\left(\mu_2 - x_i\beta\right) - \Phi\left(\mu_1 - x_i\beta\right) \\
&\Pr\left[y_i = 3 \mid x_i\right] = \Phi\left(\mu_3 - x_i\beta\right) - \Phi\left(\mu_2 - x_i\beta\right) \\
&\vdots \\
&\Pr\left[y_i = M \mid x_i\right] = 1 - \Phi\left(\mu_{M-1} - x_i\beta\right)
\end{aligned}$$

- Likelihood function is

$$L(\beta, \mu; y, x) = \prod_{i=1}^{n}\left[\sum_{m=1}^{M} I\left(y_i = m\right)\Pr\left[y_i = m \mid x_i; \beta, \mu\right]\right]$$

# Geometry of ordered probit

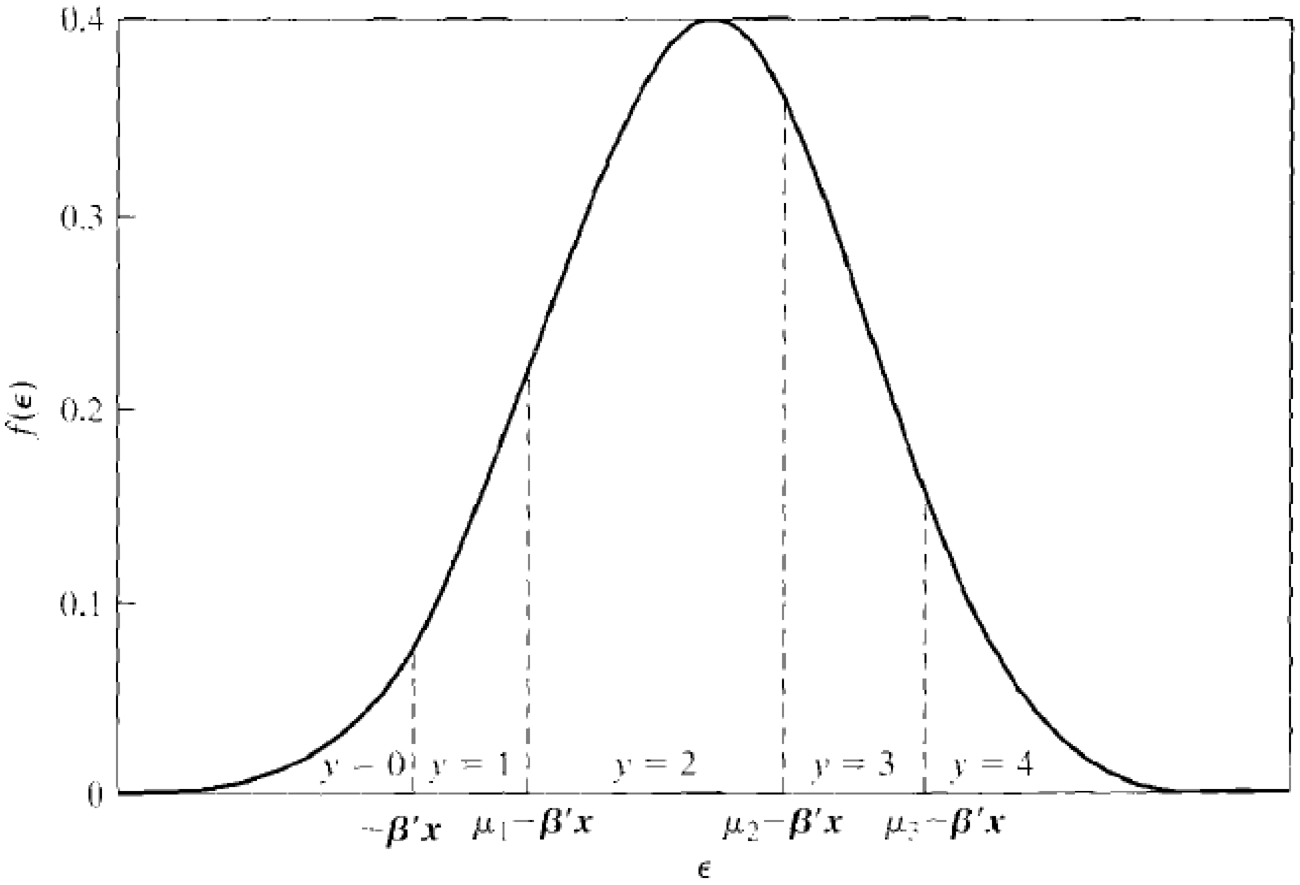



**FIGURE 23.4** Probabilities in the Ordered Probit Model.

- Diagram from Greene's text
- Distribution is of *u*
- Five outcomes here: 0 to 4
- If *u* is very negative, then $y = 0$
- If *u* is very positive, then $y = 4$
- Area in each range is probability that *y* is that value

# Effects of change in *x*

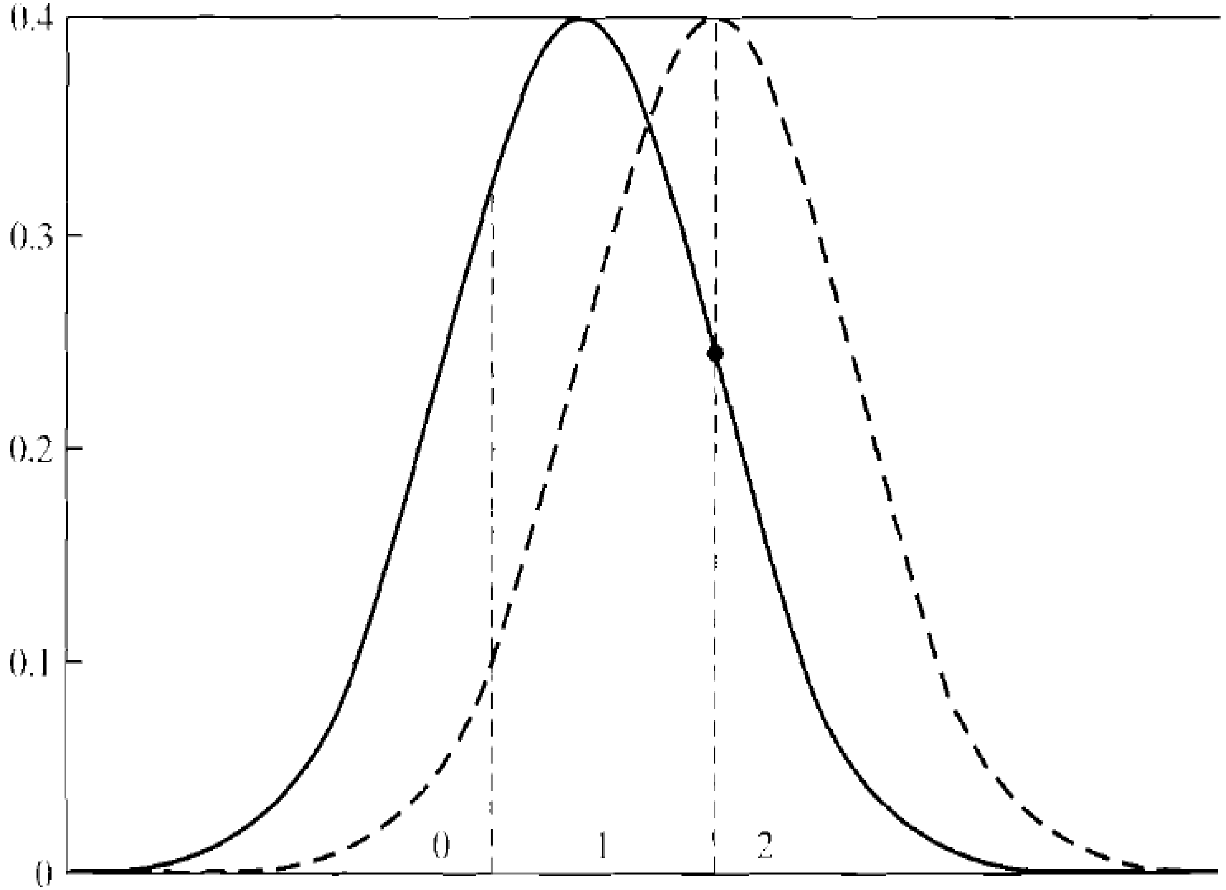



**FIGURE 23.5** Effects of Change in *x* on Predicted Probabilities.

- Change in $x$ shifts curve from solid to dashed
- Reduces likelihood of outcome $y = 0$
- Increases likelihood of outcome $y = 2$
- Changes range of $u$ for which $y = 1$ (could increase or decrease likelihood)

# Interpreting coefficients

- These equations allow Stata to calculate marginal effects: how does $\Delta x$ affect probability of each outcome?
    - margins, dydx(*) predict (outcome(#1))
    - Omit the predict clause to get all
    - predict probs* creates new variables probs1 … probsM with probabilities of outcomes for each observation

$$\frac{\partial \Pr\left[y_i = 1 \mid x_i\right]}{\partial x_{j,i}} = -\phi\left(\mu_1 - x_i\beta\right)\beta_j,$$

$$\frac{\partial \Pr\left[y_i = 2 \mid x_i\right]}{\partial x_{j,i}} = \left[\phi\left(\mu_1 - x_i\beta\right) - \phi\left(\mu_2 - x_i\beta\right)\right]\beta_j,$$

$$\vdots$$

$$\frac{\partial \Pr\left[y_i = M \mid x_i\right]}{\partial x_{j,i}} = \phi\left(\mu_{M-1} - x_i\beta\right)\beta_j.$$

# “Count” dependent variables

- Take on values 0, 1, 2, …
  - Can’t follow normal distribution
- Often modeled with **Poisson distribution**

$$\Pr\left[y_i = m \mid x_i\right] = \frac{e^{-\lambda_i}\lambda_i^m}{m!}$$

- $\lambda_i$ is mean (and variance) of distribution of *i*th observation
- Poisson regression $\lambda_i = e^{x_i\beta}$

$$\ln L = \sum_{i=1}^{n}\left[-e^{x_i\beta} + y_i x_i\beta - \ln\left(y_i!\right)\right]$$

- Stata: poisson command

# Using Poisson regression

- Interpretation of coefficients

$$\frac{\partial E[y_i \mid x_i]}{\partial x_j} = \lambda_i \beta_j = e^{x_i \beta} \beta_j$$

- Limitation: $\lambda$ plays two roles: both mean and variance of conditional distribution of $y$
  - This often doesn't fit the data
- More general alternative is **negative binomial regression** (nbreg in Stata)

# Review and summary

- We discussed three estimators:
  - Multinomial logit (probit) for unordered discrete dependent variables
  - Ordered probit (logit) for ordered dependent variables
  - Poisson (negative binomial) regression for count dependent variables
- All are easy to estimate in Stata
- All require transformations to understand the coefficients in terms of marginal effects
  - Stata's margins command is designed to do this

# Another bad economist joke …

Q: What do economists use for birth control?

A: Their personalities.

--Taken from Jeff Thredgold, *On the One Hand: The Economist's Joke Book*

# What’s next?

- We have one major class of limited dependent variable models to consider: those in which the dependent variable is continuous but has restricted range (*e.g.*, cannot be negative)
- **Tobit** models are used for corner solutions (choose zero consumption of good)
- **Censored** regression is when we can’t have an outcome above or below a certain value (sold-out performance)
- **Truncated** regression is when we can’t observe an outcome above a certain value (truncated sample)