Section 10: Ordered Logit and Probit, Unordered Choice

Valentine Gilbert

November 18, 2021

Overview

1. Ordered Logit and Probit
2. Aggregating Outcomes
3. Unordered Choice

Ordered logit and probit 1/3: Ordered responses and OLS

- Suppose we have a categorical outcome variables with three categories:

$$Y_i = \begin{cases} 0 \text{ if Disagree} \\ 1 \text{ if Neutral} \\ 2 \text{ if Agree} \end{cases}$$

- How can we model the relationship between the outcome variable and a set of explanatory variables, X_i?
- Two common approaches to ordered responses using OLS:
  1. Transform the ordered response into a binary dependent variable
  2. Transform the ordered response into a z-score
- Why are these preferable to regressing Y_i on X_i?

Ordered logit and probit 2/3: Latent variables

- Another approach is to model the *observed* outcomes as a function of a continuous, unobserved latent variable:

$$Y_i = \begin{cases} \text{Disagree,} & \text{if } Y_i^* \leq C_1 \\ \text{Neutral,} & \text{if } C_1 < Y_i^* \leq C_2 \\ \text{Agree,} & \text{if } C_2 < Y_i^* \end{cases}$$

- Here, $Y_i^* = X_i'\beta + u_i$
- So what's the probability that Y_i is "Disagree"?
- $P(Y_i = \text{Disagree}) = P(Y_i^* < C_1) = P(X_i'\beta + u_i < C_1) = P(u_i < C_1 - X_i'\beta)$
- How do we proceed from here?

Ordered logit and probit 3/3: Estimation

To make further progress, we have to make a distributional assumption about u_i

- If we assume that $u_i \sim \mathcal{N}(0,1)$, we get an ordered probit:

$$\begin{aligned} P(Y_i = \text{Disagree}) &= P(Y_i^* \leq C_1) \\ &= P(X_i'\beta + u_i \leq C_1) \\ &= P(u_i \leq C_1 - X_i'\beta) \\ &= \Phi(C_1 - X_i'\beta) \end{aligned}$$

- What about $P(Y_i = \text{Neutral})$ and $P(Y_i = \text{Agree})$?

$$\begin{aligned} P(Y_i = \text{Neutral}) &= P(C_1 < Y_i^* \leq C_2) \\ &= \Phi(C_2 - X_i'\beta) - \Phi(C_1 - X_i'\beta) \\ P(Y_i = \text{Agree}) &= P(C_2 < Y_i^*) \\ &= 1 - \Phi(C_2 - X_i'\beta) \end{aligned}$$

Ordered logit and probit: Estimation

We can now write out the log likelihood function. Let $y_{ij} = \mathbf{1}[Y_i = j]$. Individual i's contribution to the likelihood function is:

$$
\begin{aligned}
&P(Y_i = 0)^{y_{i0}} \times P(Y_i = 1)^{y_{i1}} \times P(Y_i = 2)^{y_{i2}} \\
&\quad = \Phi(C_1 - X_i'\beta)^{y_{i0}} \times \left[\Phi(C_2 - X_i'\beta) - \Phi(C_1 - X_i'\beta)\right]^{y_{i1}} \times \left[1 - \Phi(C_2 - X_i'\beta)\right]^{y_{i2}}
\end{aligned}
$$

The log-likelihood function is

$$
L = \sum_i y_{i0} \ln \Phi(C_1 - X_i'\beta) + y_{i1} \ln \left[\Phi(C_2 - X_i'\beta) - \Phi(C_1 - X_i'\beta)\right] + y_{i2} \ln \left[1 - \Phi(C_2 - X_i'\beta)\right]
$$

What's the next step to estimate the model?

Ordered logit and probit: Interpretation

As always, with probit and logit, you should report marginal effects rather than coefficients
For example, the average marginal effect of X_{ki} on $P(Y_i = \text{Agree})$ is:

$$\begin{aligned}
\frac{1}{N}\sum_i^N \frac{\partial}{\partial X_{ki}} P(Y_i = \text{Agree}) &= \frac{1}{N}\sum_i^N \frac{\partial}{\partial X_{ki}} \left[1 - \Phi(C_2 - X_i'\beta)\right] \\
&= \frac{1}{N}\sum_i^N \left[1 + \beta_k \phi(C_2 - X_i'\beta)\right]
\end{aligned}$$

What's the interpretation of this marginal effect?
What about the average marginal effect of X_{ki} on $P(Y_i = \text{Neutral})$?

- What would the interpretation be?
- Is this marginal effect useful?
- What would be a more useful marginal effect to calculate?

Aggregating outcomes into an index

- If you have multiple outcomes of interest that are related to each other, you can **aggregate them into an index**
- This is more common for experiments
- The index is the average z-score across outcomes, where the z-score is constructed using the control group mean and standard deviation, and outcomes are signed so that desirable outcomes all go in the same direction.

With two outcomes Y_{1i} and Y_{2i} that both go in the same direction (e.g. "more is better") we can construct the following index:

$$W_i = \underbrace{\frac{Y_{1i} - \bar{Y}^0_{1i}}{\widehat{sd}(Y_{1i}|D_i = 0)}}_{\text{z-score for } Y_{1i}} + \underbrace{\frac{Y_{2i} - \bar{Y}^0_{2i}}{\widehat{sd}(Y_{2i}|D_i = 0)}}_{\text{z-score for } Y_{2i}}$$

Then estimate via OLS: $W_i = \beta_0 + \beta_1 D_i + u_i$

Aggregating outcomes into an index

- Example: Kling, Liebman, and Katz (ECMA 2007) analyze the effects of MTO by combining outcomes into multiple indices.
  - Adult economic self-sufficiency: + adult not employed and not on TANF + employed + 2001 earnings - on TANF - 2001 government income
  - Adult mental health: - distress index - depression symptoms - worrying + calmness + sleep
  - Adult physical health: - self-reported health fair/poor - asthma attack past year - obesity - hypertension - trouble carrying/climbing
- This approach helps to avoid problems with multiple hypothesis testing
- Aggregation also increases power to detect effects

Unordered choice: Motivation

There are many settings in which agents choose from among a discrete set of options with no clear ordering:

- Household choice of what neighborhood to live in
- Student choice of what college to attend
- Individual choice of health insurance plan
- Commuter choice of mode of transportation

In these settings, choice probabilities are modeled as a function of choice characteristics and decision-maker characteristics

$$P(Y_i = j) = f(X_i, Z_i; \theta)$$

Unordered choice: Motivation

Estimating choice probabilities amounts to **estimating demand** for differentiated products

Why estimate demand?

- Want to know how much demand there will be for a new "product"
- Want to know how changing characteristics (e.g. price, quality) of some choices will change demand for other products
- Want to do estimate welfare effects of counterfactual policies

Example: How does building new housing affect incumbent residents? Does it increase affordability or spur gentrification?

- Neighborhoods are described by endogenous housing prices, sociodemographic composition, exogenous and endogenous amenities
- Choices depend on neighborhood characteristics, household income, level of education
- Effect of new housing depends on how a change in housing prices affects neighborhood demand, which in turn affects prices and amenities

Unordered choice: Choice probabilities

Choice probabilities can be derived from a model of utility maximization. The utility individual i obtains from choosing option j is:

$$U_{ij} = v(X_j, Z_i; \theta) + \varepsilon_{ij}$$

- The first term is the part of utility that's common to all observably identical individuals
- The second term is the part that's idiosyncratic to individual i and choice j

If ε_{ij} follows a Type-1 Extreme Value distribution, the probability that U_{ij} maximizes utility is given by

$$\pi_{ij} \equiv P(U_{ij} = \max\{U_{i1}, \ldots, U_{iJ}\}) = \frac{\exp(v(X_j, Z_i; \theta))}{\sum_{j'} \exp(v(X_{j'}, Z_i; \theta))}$$

Unordered choice: Choice probabilities

Often, $v(X_j, Z_i; \theta)$ is modeled as linear in the parameters: $v(X_j, Z_i; \theta) = X_j'\beta + Z_i'\gamma_j$

So choice probabilities are:

$$\pi_{ij} = \max\{U_{i1}, \ldots, U_{iJ}\}) = \frac{\exp(X_j'\beta + Z_i'\gamma_j)}{\sum_{j'} \exp(X_{j'}'\beta + Z_i'\gamma_j)}$$

We can then use MLE to estimate the preference parameters β and γ_j

- For example, if X_{1ij} is rent and X_{2ij} is school quality, then $\hat{\beta}_2/\hat{\beta}_1$ would give average **willingness to pay** for school quality

Unordered choice: Independence of irrelevant alternatives

One consequence of these choice probabilities is that they exhibit **independence of irrelevant alternatives** (IIA)

- IIA implies that the relative probability of choosing between two alternatives doesn't depend on the availability or characteristics of any other alternatives
- E.g. the relative probability of choosing between Flour and Shake Shack for lunch doesn't depend on whether Tatte is busy

The probability of choosing alternative j relative to k is simply

$$\frac{\pi_{ij}}{\pi_{ik}} = \frac{\exp(X_j'\beta + Z_i'\gamma_j)}{\exp(X_k'\beta + Z_i'\gamma_k)}$$

Notice that this doesn't depend at all on the presence of the other alternatives

Unordered choice: Independence of irrelevant alternatives

IIA may sound reasonable, but it implies unrealistic substitution patterns. Let's review the red bus / blue bus problem:

- Suppose 10% of commuters go by red bus, 10% go by blue bus, and 80% drive
- $P(Y_i = \text{blue bus}) / P(Y_i = \text{drive}) = 1/8$
- What happens to red bus riders if the red bus is taken out of service?
- Only 1/8 will switch to the blue bus; the rest drive!

Unordered choice: Independence of irrelevant alternatives

There are a few ways of allowing for more realistic substitution patterns

- Use a nested logit – IIA will hold across alternatives within the same nest but not across alternatives in different nests
- Allow preferences to vary with observable characteristics – IIA will hold conditional on observables
- Allow preferences to vary with unobservable characteristics