Discrete Outcome Models

Where are we in this course?

1. General Introduction to econometrics

2. Introduction to Stata as an econometric software

3. Understanding data and data management before and during the analysis process

4. Introduction to Regression Analysis

5. Introduction to limited dependent variable models

6. Introduction to binary choice (probability) models

7. Introduction to Discrete (categorical) choice models (ordered and nominal dependent variables)

8. Count Data and other econometric models (theoretical)

Ordered response models

- Models of ordered choices describe settings in which individuals reveal the strength of their utility with respect to a single outcome

- Possible applications
  - Consumer preferences of a company product
  - Farmer satisfaction with quality of extension service
  - Level of technology adoption (none, low, medium, high)

  - Have also been extended to ordinal Ordinal food security scales

- Let y be an ordered response taking on the values {0,1, 2, ……, J}

- We derive the Ordered Probit from a **latent variable** model

$$y^* = \beta_1 X_1 + \cdots + \beta_k X_k + \mu_i$$

- where μ is a normally distributed variable with the variance normalized to one

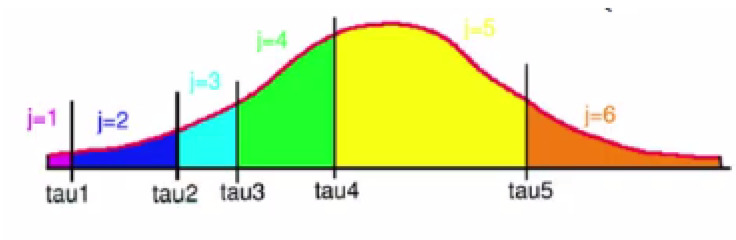
- Notice that this model does not contain a constant

Ordered depended variable models

- Ordered probit: $Y_i^* \sim STN(y_i^* | \mu_i), \mu_{i=x_i\beta}, \quad Y_i \perp Y_{i'} | X$

If instead $Y_i^* \sim STL(y_i^* | \mu_i)$: ordered logit model

- Observation mechanism: $y_i = \begin{cases} 0; if\ U_i^* \le \mu_0 \\ 1; if\ 0 < U_i^* \le \mu_1 \\ 2; if\ \mu_1 < U_i^* \le \mu_2 \\ 3; if\ \mu_2 < U_i^* \le \mu_3 \\ \vdots \\ J; if\ \mu_{\tau-1} < U_i^* \le \mu_\tau \end{cases}$



- $\mu_0, \ldots, \mu_j$ are estimated in the model together with the β

Deriving the ordered probit likelihood function

Assumptions

1. μj > μj-1 for all j
2. Parallel regression- relationship between each pair of outcome groups is the same

- There is only one set of coefficients (only one model

- Probability of one observation

$$\Pr(Y_i = j) = \Pr\left(\tau_{j-1} \le Y_i^* \le \tau_j\right) = \int_{\tau_{j-1}}^{\tau_j} STN(y_i^*|\mu_i)dy_i^*$$
$$= F_{stn}\left(\tau_j|\mu_i\right) - F_{stn}\left(\tau_{j-1}|\mu_i\right) = F_{stn}\left(\tau_j|x_i\beta\right) - F_{stn}\left(\tau_{j-1}|x_i\beta\right)$$

- Joint probability: $P(Y) = \prod_{i=1}^{n}[\Pr(Y_i = j)]$
- Log –likelihood

$$lnL(\beta, \tau|y) = \sum_{i=1}^{n} lnPr(Y_i = j)$$
$$= \sum_{i=1}^{n} \ln[F_{stn}\left(\tau_j|x_i\beta\right) - F_{stn}\left(\tau_{j-1}|x_i\beta\right)]$$

Careful of optimization constraints $\tau_{j-1} < \tau_j, \forall j$

Ordinal Probit interpretation

How to interpret β?

- β: linear effect of X on Y^* (in SD units)
- $\widehat{\Pr(Y_i|X)}$: on the simplex, J probabilities sum to 1
- One first difference: effects all J probabilities
- When one probability goes up: $\geq$ 1 probability must go down
- In practice concentrate on interpreting only the direction with no reference to the magnitude of the coefficients
- Alternatively, partial effects (marginal effects) can derived for each category can be interpreted
  - Only the first and last category marginal effects are meaningful, the intermediate categories often have very small marginal effects

Multinomial response models

- Suppose now the dependent variable is such that
  - **More than two outcomes** are possible,
  - The outcomes **cannot be ordered** in any natural way

- Examples
  - Cooking energy used (firewood, charcoal, LPG)
  - Occupational status (self-employed, wage-employed or unemployed)
  - Mode of going to work (Walking, cycling, riding, driving)
  - Type of herbicide used for weeds management

- Clearly, CLRM, binary probit and logit models are **ill-suited** for modelling data of this kind

- However, the logit model for binary choice can be extended to model more than two outcomes.

Multinomial response models

- Suppose there are J possible outcomes in the data, Y can then take J values, e.g. 0,1,…,J-1.

- So if we are modelling, say, **cooking fuel used**, and this is either firewood, charcoal or LPM, we have J = 3

- There is **no** natural ordering of these outcomes
  - What number goes with what category is **arbitrary**

- Suppose we decide on the following:

  y = 0 if individual uses firewood

  y = 1 if individual uses charcoal

  y = 2 if individual uses LPM

- We write the conditional probability that an individual belongs to category j = 0, 1, 2 as

$$\Pr(y_i = j|x_i$$

where x_i is a vector of explanatory variables

Multinomial response models

- Reasonable restrictions on these probabilities are:
  - that each of them is bounded in the (0,1) interval,
  - that they sum to unity (one)

- One way of imposing these restrictions is to write the probabilities in logit form:

$$\Pr(y_i = 1|x_i) = \frac{\exp(x_i\beta_1)}{1 + \exp(x_i\beta_1) + \exp(x_i\beta_2}$$

$$\Pr(y_i = 2|x_i) = \frac{\exp(x_i\beta_2)}{1 + \exp(x_i\beta_1) + \exp(x_i\beta_2}$$

$$\Pr(y_i = 0|x_i) = 1 - \Pr(y_i = 1|x_i) - \Pr(y_i = 2|x_i)$$

$$\Pr(y_i = 0|x_i) = \frac{1}{1 + \exp(x_i\beta_1) + \exp(x_i\beta_2}$$

- There are now two parameter vectors, β_1 and β_2

- In the general, with J possible outcomes, there are J -1 parameter vectors

MNL Interpretation

Continuous:

- A one-unit increase in X is associated with a B change the relative log odds of being in category J vs Base category

Categorical (ordinal):

- The relative log odds of being in category J vs base category change by B if moving from the lowest level of X to the highest level of X

Categorical (nominal/dummy):

- The relative log odds of being in category J vs base category change by B for X_cat compared to X_base_cat

- Marginal effects can also be used for the interpretations
  - The marginal effect may be negative even if β_{1k} is positive, and vice versa
  - The point is that whether the probability that y falls into, say, category 1 rises or falls as a result of varying x_{ik}, depends not only on the parameter estimate β_{1k}, but also on β_{2k}.

MNL Estimation problems

- Simultaneously runs binary logit models
  - One for each pair of outcomes (with one base)
- Each analysis is potentially run on a different sample
- Without constraining the logistic models, we can end up with the probability of choosing all possible outcome categories greater than 1
- Collapsing number of categories to two and then doing a logistic regression
  - suffers from loss of information and changes the original research questions to very different ones.
- Ordinal logistic regression: If the outcome variable is truly ordered and if it also satisfies the assumption of parallel regression, it could simplify the estimations

MNL Independence of irrelevant alternatives (IIA)

- MNL assumes that the ratio of any two probabilities j and m depends only on
  1. the parameter vectors β_j and β_m, and
  2. the explanatory variables x_i

$$\frac{\Pr(y_i = 1|x_i)}{\Pr(y_i = 2|x_i)} = \frac{\exp(x_i\beta_1)}{\exp(x_i\beta_1)}$$

$$= \exp(x_i(\beta_1 - \beta_2)$$

- It follows that the inclusion or exclusion of other categories must be irrelevant to the ratio of the two probabilities that y = 1 and y = 2.
  - This is potentially restrictive, in a behavoral sense.

- Individuals can use firewood, charcoal or LPM to cook

- Modelling this decision, and obtains an estimate of

$$\frac{\Pr(y_i = firewood|x_i)}{\Pr(y_i = LPM|x_i)}$$

MNL Independence of irrelevant alternatives (IIA)

- Suppose government bans use of charcoal
- Individuals choose between firewood and charcoal
  - Do you think this ratio will remain the same using data from the same regime

- If not, this suggests the MNL modelling the choice between firewood, charcoal and LPM is **mis-specified**:
  - the presence of a charcoal alternative is not irrelevant for the above probability ratio, and thus for individuals decisions more generally

Hausman test for IIA

1. Estimate the **full model**. And store the coefficients
2. Omit one category and re-estimate the model
3. Compare the coefficients from (1) and (2) above using the usual Hausman formula.

   Under the null that IIA holds, the coefficients should not be significantly different from each other.

Alternative-specific multinomial probit and nested logit models relax the IIA assumptions but requires data be choice specific

Censored and Count Data Models

Tobit and Poisson Models

Tobit Model…. Motivation

- An extension of the probit model is the **Tobit model**

- Assume household produce storage
- In the probit model our concern would be estimating the probability of storing produce after harvest a function of some **socioeconomic variables**

- In the Tobit model our interest is in finding **out the amount of produce stored** by a farmer in relation to socioeconomic variables

- Where if the dilemma?
  - If a household does not store, we have no data on storage for such household;
  - We have such data only on households that actually store

- This divides our farmers into two groups:
  1. n_1 consumers about whom we have information on the regressors (income, yield...) as well as the regressand (quantity stored)
  2. n_2 consumers about whom we have information only on the regressors but **NOT** on the regressand.

Tobit Model…. Motivation

- A sample in which information on the regressand is available only for some observations is known as a **censored sample**

- Therefore, the Tobit model is also known as a **censored regression model**

- **Censoring defined:**
  - occurs when we observe the independent variables for the entire sample, but for some observations we have only limited information about the dep variable

- **Truncated defined:**
  - In contrast a censored sample should be distinguished from a **truncated sample** in which information on the regressors is available only if the dep variable is observed

- Other applications of Tobit
  - Amount of charitable donation
  - Hours worked by married women

Tobit Model specification

- We can express the Tobit model as (**equation a.1**)

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i \quad if\ RHS > 0$$

$$= 0 \qquad\qquad otherwise$$

where RHS = right-hand side.

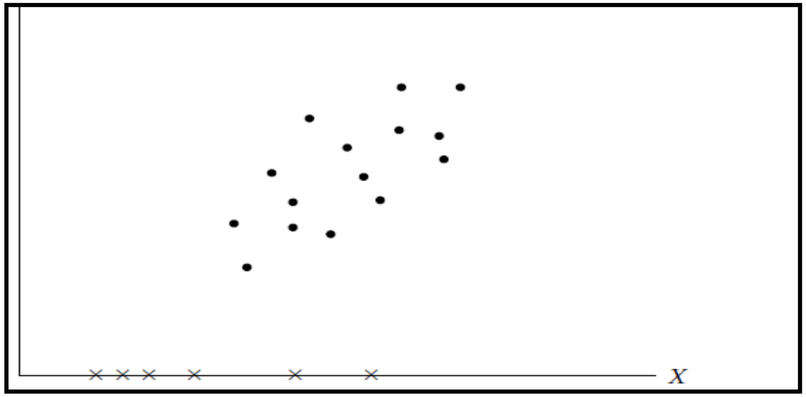
- If we estimate regression (**equation a.1**) using only n_1 observations:
  - The OLS estimates of the parameters obtained from the subset of n_1 observations will be *biased as well as inconsistent*
  - Because if we consider only the n_1 observations and omit the others, there is **no guarantee** that $E(\mu_i)$ will be necessarily zero.
  - without $E(\mu_i) = 0$ we cannot guarantee that the OLS estimates will be unbiased

- If *Y* is not observed (because of censoring), all such observations (= n_2), denoted by the crosses will lie on the horizontal axis (in figure next slide)

- If *Y* is observed, the observations (= n_1), denoted by dots, will lie in the *X*–*Y* plane.

Tobit Model specification

- If we estimate a regression line based on the n_1 observations only,
  - the resulting intercept and slope coefficients are bound to be different than if all the ($n_1 + n_2$) observations were taken into account.



- Tobit is estimated using the maximum likelihood estimator or the Heckman 2-step procedure
- In practice, most software are built with the MLE algorithm so its more commonly used

Heckman 2 step

1. Estimate the probability of a farmer storing produce using a probit
2. Estimate the model (equation a.1) by adding to it a variable (called the **inverse Mills ratio or the hazard rate**) that is derived from the probit estimate

Coefficients of Tobit are interpreted as in linear regression model

Count data models … Poisson

- What is count data?

- Phenomena where the regressand is of the **count type,**
  - number of vacations taken by a family per year,
  - the number of new crop varieties released by research institutes year,
  - Number of coups d'etat in African States per year
  - The number of medical consultations for each survey respondent

- Sometimes count data can also refer to *rare,* or *infrequent,* occurrences such as
  - Getting hit by lightning in a span of a week
  - winning more than one lottery within 2 weeks
  - Having two or more heart attacks in a span of 4 weeks.
  - How do we model such phenomena?

- **Poisson** probability distribution is specifically suited for count data

Count data models … Poisson

- **Poisson distribution first principles**
  - The underlying variable in each case is discrete
  - Event count: Number of events in a time period for unit
  - No upper limits on the number of events

- **Assumptions:**
  - Events are only observed at the end of the specified period
  - 0 events occur at the start of the period
  - No 2 events occur at the same time
  - Pr(event at time $t|events$ up to $t-1$) constant $\forall t$.

Count data models … Poisson

- For estimation purposes, we write the model as:

$$Y_i = \frac{\lambda^Y e^{-\lambda}}{Y!} + \mu_i$$

- As you can see, the resulting regression model will be nonlinear in the parameters, necessitating nonlinear regression estimation

- The pdf of the Poisson distribution is given by

$$f(Y_i) = \frac{\lambda e^{-\lambda}}{Y!} \qquad Y = 0,1,2 \ldots \ldots$$

- where $f(Y)$ denotes the probability that the variable Y takes non-negative integer values, and where $Y!$ (read Y factorial) stands for $Y! = Y \times (Y - 1) \times (Y - 2) \ldots\ldots \times 2 \times 1$.

- It can be proved that:

$$E(y) = \lambda$$
$$Var(Y) = \lambda$$

Its variance is the same as its mean value

Count data models … Poisson coefficients

- The Poisson regression model may be written as:

$$Y_i = E(Y_i) + \mu_i = \lambda_i + \mu_i$$

- where the *Y*'s are independently distributed as Poisson random variables with mean λ_i for each individual expressed as;

$$\lambda_i = E(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}$$

Interpretation

- Derivative method

$$\frac{\partial \lambda_i}{\partial X_i^1} = \exp(x_i \beta)\, \beta_1 = \lambda_i \beta_1$$

How much does the expected number of events change as one of the explanatory variable changes.

The change is therefore lambda time beta

- So we could use the mean of $y\ (\bar{y})\ times\ \beta$ for an approximate linearized effect
- That is, the rate of change of the mean value with respect to a regressor is
  - equal to the coefficient of that regressor times the mean value.
  - Of course, the mean value μ will depend on the values taken by all the regressors in the model.

Count data models … Poisson coefficients

The Poisson model makes **restrictive assumptions**:

- in that the mean and the variance of the Poisson process are the same

$V(Y_i|X_i) = E(Y_i|X_i)$, heteroskedastic and fixed

If : $V(Y_i|X_i) > E(Y_i|X_i)$

We have overdispersion and the SEs will be too small (very common)

If: $V(Y_i|X_i) < E(Y_i|X_i)$

We have under-dispersion and the SEs will be too big

Possible solutions

1. For over-dispersed data (conditional on X)-**Negative Binomial**

The model: $Y_i \sim NegBin(y_i|\phi, \sigma^2), E(Y_i \equiv \phi = e^{x_i\beta}, Y_i \perp Y_j|X$

2. **Generalized event count model:** An event count model with under-, Poisson and over-dispersion