Econometrics

Week 10

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Fall 2022

Recommended Reading

For today

- Limited Dependent Variable Models
- Chapter 17.1

The next week

- Limited Dependent Variable Models
- Chapter 17.2-17.5

Today's Talk

- Limited Dependent Variables (LDV) are dependent variables whose range of values is substantially restricted (limited)
  - binary variables (values 0 or 1)
  - count variables (values 1, 2, 3...)
  - truncated variables
  - censored variables
  - etc.
- These kinds of variables need special treatment, because
  - we want to predict fitted values that are within the range
  - properties of the disturbance term are different
- During the following two lectures we will discuss models dealing with this kind of data
- We will focus on their cross-sectional applications, but these models can be used as well in panel data or time series data

Binary Dependent Variables

- Consider a binary dependent variable, e.g. voting preferences in upcoming presidential election
  $y = 1$ if a person plans voting for Babiš
  $y = 0$ if a person plans voting for someone else
- Using this variable in an OLS framework, we estimate the **linear probability model**
  $y_i = Babis_i = \beta_0 + \beta_1 x_i + u_i$
- Expected value of the dependent variable can be interpreted as probability
  - $E(Babis_i|x_i) = Prob(Babis_i = 1|x_i) = \beta_0 + \beta_1 x_i$
  - Probability that a person votes for Babiš (and not for someone else)
- While this model and it's estimates are easy to interpret, it has several drawbacks

Binary Dependent Variables

- Consider the linear probability model (LPM):

$$E(y_i|x_i) = prob(y_i = 1|x_i) = \beta_0 + \beta_1 x_i$$

- The main characteristics of this model are:
  - Fitted values (estimated probabilities) are not restricted. So, they can be less than zero or greater than one
  - By definition, disturbances are heteroscedastic
  - Partial effect of any explanatory variable is constant, i.e. the same for each value of the explanatory variable (this is true for all linear models)

Binary Dependent Variables

The limitations discussed on the last slide may be overcome by modeling the probability with a nonlinear function assuming values in the $\langle 0, 1 \rangle$ range:

Binary Response Model

$$P(y = 1|\mathbf{x}) = G(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) = G(\beta_0 + \mathbf{x}\beta),$$

where $\mathbf{x}$ is a full set of explanatory variables, G is a function taking values strictly between zero and one: $0 < G(z) < 1$.

The Logit Model

$$P(y = 1|\mathbf{x}) = G(\beta_0 + \mathbf{x}\beta)$$

- While G can be any nonlinear, strictly increasing function such that $0 < G(z) < 1$, most applications use **logistic** or **probabilistic** functions.
- The **logistic function** is a CDF for the standard logistic random variable:

The Logistic Function

$$G(z) = \exp(z)/\left[1 + \exp(z)\right] = \Lambda(z)$$

- We refer to the model where $G(z) = \Lambda(z)$ as the **logit model**:

$$P(y = 1|\mathbf{x}) = \Lambda(\beta_0 + \mathbf{x}\beta)$$

The Probit Model

$$P(y = 1|\mathbf{x}) = G(\beta_0 + \mathbf{x}\beta)$$

- Another common choice of G is the standard normal cumulative distribution function (CDF)

The Probabilistic Function

$$G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(v)dv,$$

where $\phi(z)$ is the standard normal density:

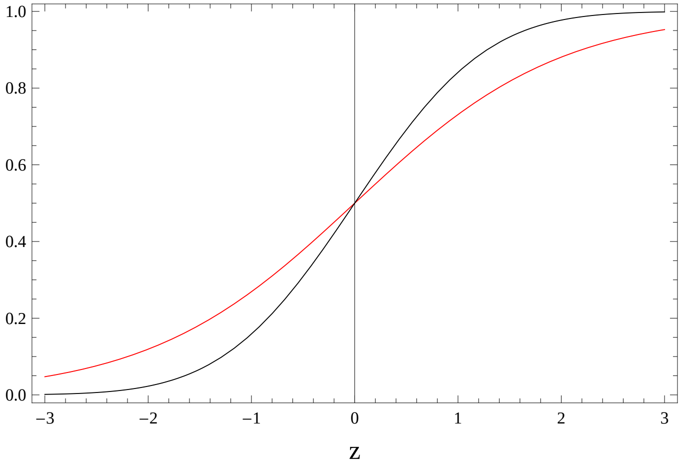
$$\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2)$$

- We refer to the model where $G(z) = \phi(z)$ as the **probit model**:

$$P(y = 1|\mathbf{x}) = \Phi(\beta_0 + \mathbf{x}\beta)$$

The Logit and Probit Models

Both G functions have very similar shapes, they are increasing in z, most quickly around 0.



Red Line is the cdf of a logistic function
Black line is the cdf of a standard normal

The Logit and Probit Models

- There is no strict reason to prefer one function over another.
- Traditionally, logit was used more often, because the logistic function leads to easier computation of the model.
- In economics, assumption of standard normal distribution is often more realistic, thus probit is preferred by economists.

Derivation of logit and probit models

- We can derive the equation for the binary response model
- Consider an unobserved (latent) variable y^*, which is linearly influenced by explanatory variable x:
  $y^* = \beta_0 + \beta_1 x + \epsilon$
- We observe only y, which is a binary variable such that:
  $y = 1[y^* > 0]$
- Then
  $P(y = 1|x) = P(y^* > 0|x) = P(\epsilon > -(\beta_0 + \beta_1 x)|x) =$
  $= 1 - G[-(\beta_0 + \beta_1 x)] = G(\beta_0 + \beta_1 x)$
- On-the-board example: voting preferences

The Logit and Probit Models Estimation

- Due to their nonlinear nature, we have to use the *Maximum Likelihood Estimation (MLE)*.
- Assume we have a random sample of size n. To obtain MLE, we need the density of y_i given x_i:

$$f(y|x_i;\beta) = [G(x_i\beta)]^y[1 - G(x_i\beta)]^{(1-y)},$$

where $y = \{0, 1\}$

MLE of β

To obtain MLE of β, we need to maximize the following log-likelihood:

$$\mathcal{L}(\beta) = \sum_{i=1}^{n} y_i \log[G(x_i\beta)] + (1 - y_i) \log[1 - G(x_i\beta)],$$

If $G(.)$ is the standard logistic cdf, $\hat{\beta}$ is the *logit estimator*, if $G(.)$ is the standard normal cdf, $\hat{\beta}$ is the *probit estimator*.

Properties of Logit and Probit Estimates

- The theory of conditional MLE for random samples implies that, under very general conditions, the MLE is **consistent, asymptotically normal and asymptotically efficient**.
- Thus we can derive asymptotic standard errors for estimates easily
- And we can use them to test single hypotheses.

Interpretation of Logit and Probit Estimates

- The most difficult aspect of these models is presenting and interpreting results.
- Usually we are interested in the effect of x on the probability that y happens, $P(y = 1|x)$.
- This effect can be obtained as: $\partial P(x)/\partial x = g(\beta_0 + x\beta)\beta_j$.
- Thus, β_j coefficients give us the sign of the partial effect of each x_j on the response probability $P(y = 1|x)$, but their magnitudes don't have direct interpretation.
- Note that the effect of x on $P(y = 1|x)$ is not constant, it depends on $\mathbf{x}$
- We can calculate these effects on sample averages to obtain the **partial effect at the average**
- Or, we can compute this effect for each observation and average it to obtain the **average marginal effect**

Interpretation of the Logit and Probit Estimates

- To measure goodness-of-fit, we cannot simply use R^2.
- One possibility is a **pseudo** R^2 based on the log-likelihood: $1 - \mathcal{L}_r/\mathcal{L}_u$.
  - $\mathcal{L}_u$ is the log-likelihood of the full model
  - $\mathcal{L}_r$ is the log-likelihood of a model with intercept only
- We can also look at the **percent correctly predicted** measure – if the model predicts a probability > 0.5 then $\hat{y} = 1$, otherwise $\hat{y} = 0$.
- **percent correctly predicted** is the percentage of times the predicted y_i ($\hat{y}_i$) matches actual y_i (which is zero or one).

Testing Multiple Hypotheses

- We can test multiple restrictions in logit and probit models.
- The easiest way is to use the **Likelihood Ratio (LR) test**:

$$LR = 2(\mathcal{L}_u - \mathcal{L}_r) \overset{a}{\sim} \chi^2_q$$

  where u is unrestricted and r restricted model and we have q restrictions.
- In this way we can simply test the significance of variables.
- If we drop a variable from the model and log-likelihood significantly decreases, we know that this variable is significant for the model.

Thank you

Thank you very much for your attention!