Limited Dependent Variable Models

BS1802 Statistics and Econometrics

Jiahua Wu

382 Business School
j.wu@imperial.ac.uk

Imperial College Business School

Imperial means Intelligent Business

Roadmap

- Regression analysis with cross-sectional data
  - The multiple regression analysis
    - Basics: estimation, inference, analysis with dummy variables
    - More technically involved: asymptotics, heteroskedasticity, specification and data issues
- Advanced topics
  - Limited dependent variable models
  - Panel data analysis
  - Regression analysis with time series data

Outline (Wooldridge, Ch. 17.1)

- Binary response models: Logit and Probit models

Binary Dependent Variables

- Recall the linear probability model (LPM)

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u,$$

  where y either equals 0 or 1

  - Interpretation of the LPM

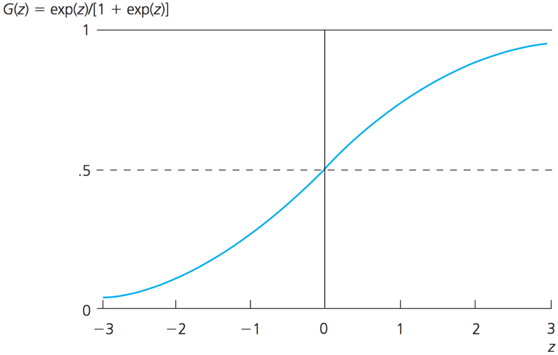
$$P(y = 1|\mathbf{x}) = E(y|\mathbf{x}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

  - Drawbacks of LPM
    - Predicted values are not constrained to be between 0 and 1
    - Violation of homoskedasticity

- An alternative is to model the probability as a function, $G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$, where $0 < G(z) < 1$

The Logit Model

- The Logit model uses the logistic function, which is the cumulative distribution function (cdf) for a standard logistic random variable
$$G(z) = \frac{\exp(z)}{1 + \exp(z)}$$



The Probit Model

- The Probit model uses the standard normal cumulative distribution function

$$G(z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right) dv$$

  - Both functions have similar shapes - they are increasing in z, most quickly around 0
  - Both models are nonlinear, and require maximum likelihood estimation (MLE)
  - No real reason to prefer one over the other
    - Traditionally we saw more of the logit, mainly because the logistic function leads to a more easily computed model
    - Today, the probit is also easy to compute with standard packages

Maximum Likelihood Estimation of Logits and Probits

- Recall the interpretation of binary response models, where

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

and

$$P(y = 0|\mathbf{x}) = 1 - G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$$

- Conditional on the explanatory variables, the density of y_i is given by

$$f(y|\mathbf{x}_i; \beta) = [G(z_i)]^{y_i} \cdot [1 - G(z_i)]^{1-y_i},$$

where $z_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$, and y_i is either 0 or 1.

Maximum Likelihood Estimation of Logits and Probits

- The log-likelihood function for observation i is obtained by taking the log of $f(y|\mathbf{x}_i;\beta)$,

$$l_i(\beta) = y_i \log[G(z_i)] + (1-y_i)\log[1-G(z_i)]$$

- The log-likelihood for a sample is obtained by summing $l_i(\beta)$ across all observations: $L(\beta) = \sum_{i=1}^{n} l_i(\beta)$
- The MLE of β,denoted as $\hat{\beta}$, maximizes $L(\beta)$
- MLE is asymptotically normal and asymptotically efficient

$$\frac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)} \sim N(0,1),$$

under the null $H_0 : \beta_j = a_j$.

Interpretation of Logits and Probits

- What is the effect of x_j on $P(y=1|\mathbf{x})$?
  - For the linear case, this is just the coefficient on x_j
  - For the nonlinear logit and probit models, β_j can no longer be interpreted as the marginal effect of x_j on y
    - To see this, differentiate $P(y=1|\mathbf{x})$ with respect to x_j, we have

      $$\frac{\partial P(y=1|\mathbf{x})}{\partial x_j} = g(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)\beta_j,$$

      where $g(z)$ is dG/dz
    - Since we are bounding the dependent variable using a non-linear function, the marginal effect depends on all the estimates and their values

Interpretation of Logits and Probits

- The effects of x_j on the response probability is roughly

$$\Delta \widehat{P}(y = 1|\mathbf{x}) \approx [g(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k)\hat{\beta}_j]\Delta x_j$$

- It is usually handy to have a single scale factor $g(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k)$ that can used to multiply each $\hat{\beta}_j$
  - If each explanatory variable is replaced with its sample average, the partial effect becomes

$$g(\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \cdots + \hat{\beta}_k \bar{x}_k)\hat{\beta}_j,$$

    and we obtain the partial effect at the average (PEA)
  - Alternatively, we can average the individual partial effects across the sample, i.e.,

$$n^{-1}\sum_{i=1}^{n} g(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik})\hat{\beta}_j$$

    and we obtain the average partial effect (APE)

The Likelihood Ratio Test

- In the LPM, we can compute F stat to test exclusion restrictions ($H_0 : \beta_{k-q+1} = 0, \ldots, \beta_k = 0$)
- In the logit and probit models, F test is no longer valid - we need the likelihood ratio test
  - Maximum likelihood estimation (MLE), will always produce a log-likelihood, $\mathsf{L}(\hat{\beta})$
  - Just as in an F test, we estimate the restricted and unrestricted model, then form the likelihood ratio statistics

$$LR = 2(\mathsf{L}_{ur} - \mathsf{L}_r) \sim \chi^2_q$$

  - Reject the null $H_0 : \beta_{k-q+1} = 0, \ldots, \beta_k = 0$ if $LR > c$ (χ^2_q critical value)

Goodness of Fit

- In the LPM, R^2 is a measure for goodness of fit
- Common goodness-of-fit measures for logit and probit models
  - Pseudo R-squared: is based on the log likelihood and defined as

    $$1 - \mathsf{L}_{ur}/\mathsf{L}_r,$$

    where L_{ur} is the log-likelihood for the estimated model, and L_r is the log-likelihood in the model with only an intercept
  - Percent correctly predicted: an observation i is correctly predicted if either
    - $y_i = 1$ and $G(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}) \geq .5$ or
    - $y_i = 0$ and $G(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}) < .5$.
    - The percentage observations that are correctly predicted is used as the measure

Goodness of Fit

- Common goodness-of-fit measures for logit and probit models
  - AIC: a measure of the relative quality of statistical models
    - For maximum likelihood estimators, it is defined as

      $$AIC = 2k - 2\mathrm{L},$$

      where k is the number of regressors and L is the log-likelihood for the model
    - Given a set of candidate models, the preferred model is the one with the minimum AIC value
  - BIC: $\ln(n)k - 2\mathrm{L}$
  - AICc: $AIC + \frac{2(k+2)(k+3)}{n-k-3}$

Estimating Logit and Probit Models in R

- We use the `glm` command to estimate logit and probit models

```
glm(formula, family(link), data, …)
```

  - Three common choices of family (and link functions)

Family	Link	Model
gaussian	identity	linear regression model
binomial	logit, probit	logit, probit models
poisson	log	poisson model

  - **Family** indicates the conditional distribution of the dependent variable y
  - A **link** function $f(\cdot)$ is defined as

$$f(E(y|\mathbf{x})) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

Understanding `glm` Output

- Deviance of the estimated model
  - It is the counterpart of sum of squared residuals in the multiple regression models
  - It is defined as
    $$-2\mathsf{L}(\hat{\beta}) + c,$$
    where c is a constant
- Test for overall significance $(H_0 : \beta_1 = 0, \ldots, \beta_k = 0)$
  - The restricted model is a model with only an intercept, and its deviance is given by null deviance
  - The deviance of the unrestricted model is given by residual deviance
  - The likelihood ratio statistics
    $$LR = 2(\mathsf{L}_{ur} - \mathsf{L}_r) = \text{null deviance} - \text{residual deviance} \sim \chi^2_k.$$
    Reject the null if $LR > c$ (χ^2_k critical value)

Logit and Probit Models: An Example

- Example 17.1: Labour Force Participation (mroz.RData)

$$\begin{aligned} \textit{inlf} &= \beta_0 + \beta_1 \textit{nwifeinc} + \beta_2 \textit{educ} + \beta_3 \textit{exper} \\ &\quad + \beta_4 \textit{exper}^2 + \beta_5 \textit{age} + \beta_6 \textit{kidslt}6 + \beta_7 \textit{kidsge}6, \end{aligned}$$

  where

  - *inlf*: a binary variable indicating labor force participation by a married woman
  - *nwifeinc*: husband's earnings (in thousands of dollars)
  - *educ*: years of education
  - *exper*: past years of labor market experience
  - *age*: age
  - *kidslt*6: # of children less than 6 years old
  - *kidsge*6: # of kids between 6 and 18 years old

Logit and Probit Models: An Example

Dependent Variable: *inlf*			
Independent Variables	**LPM (OLS)**	**Logit (MLE)**	**Probit (MLE)**
nwifeinc	−.0034 (.0015)	−.021 (.008)	−.012 (.005)
educ	.038 (.007)	.221 (.043)	.131 (.025)
exper	.039 (.006)	.206 (.032)	.123 (.019)
exper^2	−.00060 (.00018)	−.0032 (.0010)	−.0019 (.0006)
age	−.016 (.002)	−.088 (.015)	−.053 (.008)
kidslt6	−.262 (.032)	−1.443 (.204)	−.868 (.119)
kidsge6	.013 (.013)	.060 (.075)	.036 (.043)
constant	.586 (.151)	.425 (.860)	.270 (.509)
Percentage correctly predicted	73.4	73.6	73.4
Log-likelihood value	—	−401.77	−401.30
Pseudo R-squared	.264	.220	.221