Probit Estimation

Quantitative Microeconomics

R. Mora

Department of Economics
Universidad Carlos III de Madrid

Outline

The Random Utility Model

- $U_m = \beta_m x_m + \varepsilon_m$
- $U_h = \beta_h x_h + \varepsilon_h$

$$\Updownarrow$$

$$\beta_m x_m + \varepsilon_m > \beta_h x_h + \varepsilon_h \Leftrightarrow work = 1$$

$$\Updownarrow$$

$$\beta x + \varepsilon > 0 \Leftrightarrow work = 1$$

where $\varepsilon = \varepsilon_m - \varepsilon_h$ (the unobserved net utility from participation) and $\beta x = \beta_m x_m - \beta_h x_h$ (the index function)

The Probit Assumption

- The econometrician only observes *work*, x_m, and x_h

Probit Assumption: $\varepsilon_h, \varepsilon_m \sim N(0, \Sigma)$

- $\varepsilon \equiv \varepsilon_m - \varepsilon_h \,|x \sim N\left(0, \sigma^2\right)$
- $Pr\left(work = 1\right) = Pr\left(\varepsilon > -\beta x\right) = Pr\left(\varepsilon \leq \beta x\right)$
- $Pr\left(work = 1\right) = Pr\left(\frac{\varepsilon}{\sigma} \leq \frac{\beta x}{\sigma}\right)$
- $Pr\left(work = 1\right) = \Phi\left(\frac{\beta}{\sigma} x\right)$
  where Φ is the cdf of the standard normal

Observability of σ

β and σ are observationally equivalent to $\beta^* = k\beta$ and $\sigma^* = k\sigma$

$$\Phi\left(\frac{\beta^*}{\sigma^*}x\right) = \Phi\left(\frac{k\beta}{k\sigma}x\right) = \Phi\left(\frac{\beta}{\sigma}x\right), k \neq 0$$

- an infinite number of pairs (β^*, σ^*) give the same likelihood
- ML identification conditions are violated

identification assumption: $\sigma = 1$ (hence $\varepsilon \backsim N(0,1)$)

- $Pr\,(work = 1) = \Phi\,(\beta x)$

Interpretation of the Slopes and Marginal Effects

when the control x_j appears in both utilities U_m and U_h...

- only the net effect on the index function, $\beta_{mj} - \beta_{hj}$, is identified

normality (nonlinearity) assumption

- "net slope" $\beta_{mj} - \beta_{hj}$ captures the marginal effect on index function βx of an increase of one unit of control x_j
- the marginal effect on the probability of participation is more complex
- if x_j is continuous, $\frac{\partial Pr(work=1)}{\partial x_j} = \phi(\beta x)\beta_j$
- if x_j is discrete, $\Delta Pr(work = 1) = \Phi(\beta x_1) - \Phi(\beta x_0)$ where x_1 is the controls with the final value for x_j and x_0 is the controls with the initial value for x_j

A Simple Example

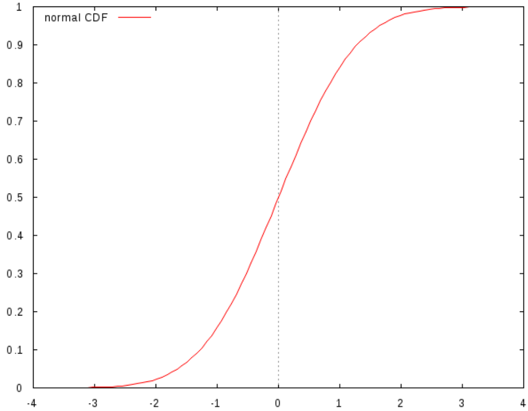
- $U_m = \beta_m^0 + \beta_m^e educ + \beta_m^k kids + \varepsilon_m$ with $\varepsilon_m \sim N\left(0, \sigma_m^2\right)$
- $U_h = \beta_h^0 + \beta_h^e educ + \beta_e^k kids + \varepsilon_h$ with $\varepsilon_h \sim N\left(0, \sigma_h^2\right)$
- $cov\left(\varepsilon_m, \varepsilon_h\right) = \sigma_{m,h}$

Probit Assumption: $\varepsilon_m - \varepsilon_h | x \sim N(0,1)$

- $Pr\left(work = 1\right) = \Phi(\beta_0 + \beta_e educ + \beta_k kids)$
- $\beta_0 = \beta_m^0 - \beta_h^0$
- $\beta_e = \beta_m^e - \beta_h^e$
- $\beta_k = \beta_m^k - \beta_h^k$
- $var\left(\varepsilon_m - \varepsilon_h\right) = \sigma_m^2 + \sigma_h^2 - 2\sigma_{m,h} = 1$

A Graphical Interpretation

The probability to participate is a nonlinear function of the index function $\beta_0 + \beta_e educ + \beta_k kids$



The Density

Assumption: iid random sample

- let the true value be β_0
- then, under the Probit model

$$Pr\left(work\left|x\right.\right) = \begin{cases} \Phi\left(\beta_0 x\right) \text{ if } work = 1 \\ 1 - \Phi\left(\beta_0 x\right) \text{ if } work = 0 \end{cases}$$

The Likelihood of an Observation

- the likelihood replaces in the density the true vector β_0 with any vector β
- then, the likelihood for individual i takes the form

$$L_i(\beta) = \begin{cases} \Phi(\beta x_i) \text{ if } work_i = 1 \\ 1 - \Phi(\beta x_i) \text{ if } work_i = 0 \end{cases}$$

- o, more conveniently,

$$L_i(\beta) = [\Phi(\beta x_i)]^{work_i} [1 - \Phi(\beta x_i)]^{1-work_i}$$

The Loglikelihood

- first, we take the logs

$$l_i(\beta) = work_i \log(\Phi(\beta x_i)) + (1 - work_i) \log(1 - \Phi(\beta x_i))$$

- then we compute the likelihood for the entire iid sample

$$l(\beta) = \sum_{i=1}^{n} l_i(\beta)$$

- hence

$$l(\beta) = \sum_i \{work_i \log(\Phi(\beta x_i)) + (1 - work_i)) \log(1 - \Phi(\beta x_i))\}$$

ML Estimation

Definition

- the MLE is the vector $\hat{\beta}^{ML}$ such that

$$\hat{\beta}^{ML} = \arg\max_{\beta} l(\beta)$$

- because of the nonlinear nature of the maximization problem, there are not explicit formulas for the probit ML estimates
- instead, numerical optimization is used, and, usually, only a few iterations are needed
- in `gretl`, a quasi-Newton algorithm is used (the BFGS algorithm)

A Perfect Classifier Control

- suppose that dummy variable D_i perfectly predicts $work_i$ in the sample in the sense that $work_i = 1 \Leftrightarrow D_i = 1$
- if $\beta x = \beta_0 + \beta_D D$, then $\beta x = \left\{ \begin{array}{ll} \beta_0 + \beta_D & \text{if } work = 1 \\ \beta_0 & \text{if } work = 0 \end{array} \right|$
- and the log-likelihood function is increasing in β_D:

$$l(\beta) = \sum_i \{work_i \log(\Phi(\beta_0 + \beta_D)) + (1 - work_i)) \log(1 - \Phi(\beta_0))\}$$

- hence, there cannot be a maximum likelihood estimator

The Perfect Prediction Problem

- more generally, suppose that vector $\widetilde{\beta}$ perfectly predicts $work_i$ in the sample in the sense that for a given scalar k, $\widetilde{\beta} x_i > k$ if and only if $work_i = 1$
- then the same thing is true for any multiple of $\widetilde{\beta}$ and the log-likelihood function will have no maximum
- this may be due to several reasons
  - one control may be a perfect classifier: drop it
  - the model may be trivially misspecified (like predicting participation among working individuals)
  - the sample may simply be not large enough

Asymptotic Properties and Testing

under general conditions, the MLE is consistent, asymptotically normal, and asymptotically efficient

- we can construct (asymptotic) t tests and confidence intervals (just as with OLS, 2SLS, and IV)
- exclusion restrictions (por ejemplo, $H_0 : \beta_j = 0 \text{ and } \beta_k = 0$)
  - the Lagrange multiplier test only requires estimating the model under the null
  - the Wald test requires estimation of only the unrestricted model
  - the likelihood ratio (LR) test requires estimation of both models

The Likelihood Ratio Test

Nested Hypothesis

- it is based on the difference in loglikelihood functions under the null and under the alternative
- restricting models cannot increase loglikelihoods

$$LR = 2\left(l_{ur}\left(\hat{\beta}_{ur}^{ML}\right) - l_r\left(\hat{\beta}_r^{ML}\right)\right) \xrightarrow{a} \chi_q$$

where q is the number of restrictions

Summary

- not all parameters of the RUM can be estimated
- the Probit model identifies how each control affects the probability of participation
- ML estimation requires numerical methods
- under general conditions, ML estimates are consistent, asymptotically normal, and asymptotically efficient
- significance tests and general restrictions tests are easy to carry out with the Probit model