Econometric Methods
PC-tutorial: Maximum Likelihood Estimation

Le Nga Tran

Winter Term 2023/24

January 23, 2024

Examples: Labor Forces Participant

Example: You want to analyse the determinants of women's labor force participation. The dependent variable y_i is a decision whether or not married woman participate the labor market.

$$y_i = \begin{cases} 1, & \text{if woman participate the job market} \\ 0, & \text{if otherwise} \end{cases}$$

The decision depends on variables x, education, age, experience, etc.

Model

The decision of participant y_i follows conditional on x_i a Bernoulli distribution

$$E(y_i|x_i) = 0 \times P(y_i = 0|x_i) + 1 \times P(y_i = 1|x_i) = P(y_i = 1|x_i)$$

$\Rightarrow$ we model the **probability** that the married woman participate in the labour force.

$$P(y_i = 1|x_i) = y_i^* = G(x_i\theta)$$

y_i^* is called latent variable and $G(x\theta)$ is a link function

$$y_i = \begin{cases} 1, & \text{if} \quad y_i^* > 0 \\ 0, & \text{if} \quad y_i^* \leq 0 \end{cases}$$

Logit and Probit model

How to choose link function $G(x\theta)$?

- Linear regression: the link function is a linear regression
$$P(y=1|x) = G(x\theta) = x\theta$$
- Probit model: standard normal distribution for e_i
$$P(y=1|x) = G(x\theta) = \Phi(x\theta) = \int_{-\infty}^{x\theta} \phi(t)dt$$
where $\Phi(x\theta)$ and $\phi(t)$ are cdf and pdf from standard normal distribution
- Logit model: logistic distribution for e_i
$$P(y=1|x) = G(x\theta) = \Lambda(x\theta) = \frac{exp(x\theta)}{1+exp(x\theta)}$$
where $\Lambda(x\theta)$ are cdf from standard logistic distribution

Interpretation

Suppose we are interested in the effect of age on woman's decision whether or not to be in the labour force.

Linear regression The link function:

$$G(x\theta) = \theta_0 + \theta_1 age + z\gamma \tag{1}$$

$\hat{\theta}_1 = -0.016$: a unit increase in woman's age leads to the probability of participating in the labor forces decrease by 1.6 percentage points.

Interpretation

In the Logit and Probit model:

Marginal effect: is not constant and depends on specific values of the regressors.

Marginal effects at means (MEA)[1] : the marginal effects evaluated the mean value of regressors.

Example: MEA for age is -0.021. Holding all the variables at their average values, a unit increase in woman's age from its average value leads to the probability of participating in the labor force decrease by 2.1 percentage points.

Average marginal effects (AME)[2] : the average of marginal effects for every observations

Example: AME for age is -0.016. A unit increase in woman's age leads to the probability of participating in labour market decrease by 1.6 percentage points on average, keeping other variables constant.

[1] partial effect of the average (PEA)

[2] average partial effect (APE)

Exe

You want to analyze the determinants of women's labor force participation. To this end, open the mroz.dta dataset in Stata.

- *inlf* is decision whether or not the woman is in the labor force, i.e. $inlf = 1$ (yes) and $inlf = 0$ (no).
- The regressors include *nwifeinc* (family income less woman's wage in 1000 dollar), *educ* (year of schooling), *age* (woman's age), *exper* and *expersq* (experience and squared experience of woman), *kidslt*6 (number of kids less than 6 years old), *kidsge*6 (number of kids from 6-18 years old)

a) Re-estimate the baseline specification presented in the textbook and in class by OLS, logit and probit. Compute the APEs and PEAs for the continuous variables.

b) Compute the partial effect of age evaluated at the first, second, and third quartile of the distribution of the other regressors.

c) Compute the average partial effect of experience both analytically (as a general function of x and θ) and empirically (for the dataset at hand). Take into account that both exper and expersq are included as regressors!

Task c)

$$y_i = G([exper_i, z_i]\beta) + e_i$$

APE of experience:

$$\begin{aligned}\widehat{APE} &= N^{-1}\sum_{i=1}^{N}\{G([exper_i + 1, z_i]\hat{\beta}) - G([exper_i, z_i]\hat{\beta})\} \\ &= N^{-1}(\sum_{i=1}^{N}\hat{G}_i^1 - \sum_{i=1}^{N}\hat{G}_i^0)\end{aligned}$$

where $\hat{G}_i^0$ is probability predicted from the model and $\hat{G}_i^1$ is probability predicted by adding one year of experience.

In case of logit model:

$$\hat{G}_i^0 = \frac{exp(\hat{\beta}_1 exper_i + \hat{\beta}_2 exper_i^2 + z_i\hat{\beta}_z)}{1 + exp(\hat{\beta}_1 exper_i + \hat{\beta}_2 exper_i^2 + z_i\hat{\beta}_z)}$$

$$\hat{G}_i^1 = \frac{exp(\hat{\beta}_1 exper_i^* + \hat{\beta}_2 exper_i^{*2} + z_i\hat{\beta}_z)}{1 + exp(\hat{\beta}_1 exper_i^* + \hat{\beta}_2 exper_i^{*2} + z_i\hat{\beta}_z)}$$

where $exper^* = exper + 1$

d) Add father's years of education, fatheduc, and mother's years of education, motheduc, as explanatory variables. Test for joint significance of these two regressors.

e) Split the quantitative variable kidslt6 into dummy variables kid0 $= 1$ if no young kids and zero else, kid1 $= 1$ if one young kid and zero else, and so on. Which specification is more restrictive? Test the more against the less restrictive specification using (a) a Wald test and (b) a likelihood ratio test.