AME

Week 9: Binary Models

Sophie Bindslev, November 2022

UNIVERSITY OF COPENHAGEN

Today's Plan

- LPM
- Probit & Logit
- Maximum Likelihood Estimation
- Variance in Binary Models
- Partial Effects
- Your time to shine!

Linear Probability Model

- The dependent variable is binary e.g. $y \in 0, 1$.
- The linear probability model:

$$P(y_i = 1 \mid \mathbf{x}_i) = \mathbf{x}_i \boldsymbol{\beta} \tag{1}$$

  where β_k is the change in probability of success $(y_i = 1)$ following a one-unit rise in $x_{i,k}$.
- Advantages of the LPM: A) approximates the average partial effect (APE) of $\mathbf{x}$ well, B) does not require assumptions regarding the entire conditional distribution $y \mid \mathbf{x}$
- Disadvantages of the LPM: A) may predict outcomes outside the $[0, 1]$ interval, B) performs poorly when considering partial effects of $\mathbf{x}$ at other bits of the $\mathbf{x}$'s distribution (partial effects are not always linear)

Probit & Logit

- Index Response Models:

$$P(y_i = 1 \mid \mathbf{x}_i) = G(\mathbf{x}_i\boldsymbol{\beta}) \tag{2}$$

  where $G(.) : \mathbb{R} \to [0,1]$ is the "link function", often a cumulative distribution function (CDF)

- The probit model assumes a standard normal CDF:

$$G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(v)dv, \quad \phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{3}$$

- The logit model assumes a logistic CDF:

$$G(z) = \Lambda(z) = \frac{\exp(z)}{1+\exp(z)} = \frac{1}{1+\exp(-z)} \tag{4}$$

Maximum Likelihood Estimation: Probit & Logit

- We have the likelihood function:

$$f(y_i \mid \mathbf{x}_i) = G(\mathbf{x}_i\beta)^{\mathbf{1}(y_i=1)}(1 - G(\mathbf{x}_i\beta))^{\mathbf{1}(y_i=0)} \tag{5}$$

- Therefore the log-likelihood function is:

$$\ell_i = \ell(\mathbf{w}_i, \boldsymbol{\theta}) = y_i \log(G(\mathbf{x}_i\beta)) + (1 - y_i)\log(1 - G(\mathbf{x}_i\beta)) \tag{6}$$

- Today we will use Maximum Likelihood to obtain parameter estimates:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{argmin} - \frac{1}{N}\sum_{i=1}^{N} \ell(\mathbf{w}_i, \boldsymbol{\theta}) \tag{7}$$

where $\mathbf{w}_i = \{y_i, \mathbf{x}_i\}$ and $\boldsymbol{\theta} = \boldsymbol{\beta}/\sigma = \boldsymbol{\beta}$ as we set $\sigma = 1$

- Are β and σ separately identified? Why/why not?

Variance in Binary Models

- We have that

$$E(\mathbf{y} \mid \mathbf{x}) = Pr(\mathbf{y} = 1 \mid \mathbf{x}) = G(\mathbf{x}\beta) \tag{8}$$
$$Var(\mathbf{y} \mid \mathbf{x}) = G(\mathbf{x}\beta)(1 - G(\mathbf{x}\beta)) \tag{9}$$

  Clearly, the variance of our error terms vary with the $\mathbf{x}$. We have heteroskedastic errors

- If using the LPM, should use robust standard errors
- If in MLE this is taken care of if we believe our model is correctly specified. Why?
- Could use the sandwich estimator for robustness in MLE but that is slightly awkward. Why?

Partial Effects

- Partial effects in binary models depend on $\mathbf{x}$:

$$\frac{\partial P(\mathbf{y}=1\mid\mathbf{x})}{\partial\mathbf{x}_k}=\frac{\partial G(\mathbf{x}\boldsymbol{\beta})}{\partial\mathbf{x}_k}$$

$$=\begin{cases} g(\mathbf{x}\boldsymbol{\beta})\beta_k & \textit{if continuous} \\ G(\mathbf{x}_{j\neq k}\boldsymbol{\beta}_{j\neq k}+\beta_k)-G(\mathbf{x}_{j\neq k}\boldsymbol{\beta}_{j\neq k}) & \textit{if discrete} \end{cases} \tag{10}$$

- In binary models the sign of β_k determines the sign of the partial effect of x_k
- The ratio of partial effects does not depend on $\mathbf{x}$:

$$\frac{\partial P(\mathbf{y}=1\mid\mathbf{x})/\partial\mathbf{x}_k}{\partial P(\mathbf{y}=1\mid\mathbf{x})/\partial\mathbf{x}_j}=\frac{g(\mathbf{x}\boldsymbol{\beta})\beta_k}{g(\mathbf{x}\boldsymbol{\beta})\beta_j}=\frac{\beta_k}{\beta_j} \tag{11}$$

Inference on Partial Effects

- For inference about the partial effects of different regressors we can no longer just look at the estimated coefficients $\hat{\boldsymbol{\beta}}$
- To obtain standard errors of partial effects we can use the delta method when $h(\boldsymbol{\beta})$ is a **continuous** transformation of $\boldsymbol{\beta}$:

$$h(\boldsymbol{\beta}) := PE_k = g(\mathbf{x}\boldsymbol{\beta})\beta_k \tag{12}$$

$$Avar(h(\boldsymbol{\beta})) = \nabla_{\beta} h(\boldsymbol{\beta}) Avar(\boldsymbol{\beta})(\nabla_{\beta} h(\boldsymbol{\beta}))' \tag{13}$$

Reporting Partial Effects

- Since partial effects depend on the values of $\mathbf{x}$ we need to decide how to report estimated partial effects
- **Partial Effect at the Average (PEA)**: $g(\bar{\mathbf{x}}\boldsymbol{\beta})\beta_k$
  - For some x_j the average may be ill-defined e.g. binary or categorical variables - no one has $\bar{x}_j$ in the population.
  - Even if all continuous variables, $\bar{\mathbf{x}}$ still may not describe anyone.
- **Average Partial Effect (APE)**: $\overline{g(\mathbf{x}\boldsymbol{\beta})\beta_k}$
  - Can get complicated if some variables are functions of each other e.g. age, age^2
- Partial Effects for interesting subgroups

Your time to shine!

- Take a look at `LinearModel.py` and `estimation.py` which you'll need for OLS and MLE, respectively
- Fill in `probit_ante.py` and `logit_ante.py` and solve the problem set
- The "Delta Method" Note by Jesper and Anders is helpful for Q8