Advanced Microeconometrics

Exercise Set 8: Binary Response Models

Andreas P. A. Pedersen, Week 46 2025

UNIVERSITY OF COPENHAGEN

Toay'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

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

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

Probit & Logit

- Index Response Models:
$$P(y_i = 1|\mathbf{x}_i) = G(\mathbf{x}_i\beta) \tag{2}$$
where $G(.) : I\!R \rightarrow [0, 1]$ is the "link function", often a cumulative distribution function (CDF)

- The Probit model assumes a standard normal CDF:
$$G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(v)dv, \;\; \phi(z) = \frac{1}{\sqrt{2\pi}} exp(-\frac{z^2}{2}) \tag{3}$$

- 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

$$f(y_i \mid x_i) = G(x_i\theta)^{1(y_i=1)}(1 - G(x_i\theta))^{1(y_i=0)} \tag{5}$$

- Therefore the log likelihood function is:

$$\ell_i(w;\theta) = y_i \log(G(x_i\beta)) + (1 - y_i)\log(1 - G(x_i\beta)) \tag{6}$$

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

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

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

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

Variance in Binary Models

- We have that

$$E(\mathbf{y}|\mathbf{x}) = Pr(\mathbf{y} = \mathbf{1}|\mathbf{x}) = G(\mathbf{x}\beta) \tag{8}$$

$$Var(\mathbf{y}|\mathbf{x}) = G(\mathbf{x}\beta)(1 - G(\mathbf{x}\beta)) \tag{9}$$

  Clearly, the variance of our error terms vary with the $\mathbf{x}$. We have heteroskedastic errors
- If you use LPM, use robust standard errors
- If in MLE this is taken care of it we believe our model is correctly specified. Why?
- Could use the sandwich estimator for robustness in MLE but that is slightly awkward. Why?

The Challenge of Non-Linearity

- In Probit/Logit models, the **Partial Effect (PE)** of a regressor x_k is a non-linear function of all estimated coefficients ($\hat{\boldsymbol{\beta}}$) and the point of evaluation ($\boldsymbol{x}$):

  $$PE_k(\boldsymbol{x}) = h(\boldsymbol{\beta}, \boldsymbol{x}) = g(\boldsymbol{x}\boldsymbol{\beta})\beta_k$$

  where $g(\cdot)$ is the density function of the link function $G(\cdot)$ (e.g., $\phi(\cdot)$ for Probit).
- Since the PE is a non-linear function of $\boldsymbol{\beta}$, its variance cannot be directly derived from the variance of $\hat{\boldsymbol{\beta}}$.

The Delta Method

The Delta Method provides an approximation of the asymptotic variance of a continuous, differentiable function of an asymptotically normally distributed estimator.
Let $\hat{\boldsymbol{\beta}}$ be the Maximum Likelihood Estimator, and let $h(\boldsymbol{\beta})$ be the function of interest (e.g., PE_k or APE_k).

- **Asymptotic Variance of** $h(\hat{\boldsymbol{\beta}})$**:**

$$\text{Avar}(h(\hat{\boldsymbol{\beta}})) = \nabla_{\boldsymbol{\beta}} h(\boldsymbol{\beta}) \cdot \text{Avar}(\hat{\boldsymbol{\beta}}) \cdot (\nabla_{\boldsymbol{\beta}} h(\boldsymbol{\beta}))'$$

Delta Method Implementation Steps

1. **Define $h(\boldsymbol{\beta})$:** Specify the function you want the variance for (e.g., PE_k at $\boldsymbol{x} = \overline{\boldsymbol{x}}$, or the Average Partial Effect (APE) by averaging $g(\boldsymbol{x}_i\boldsymbol{\beta})\beta_k$ across all observations i).
2. **Calculate the Gradient $\nabla_{\boldsymbol{\beta}} h(\boldsymbol{\beta})$:** Compute the vector of first partial derivatives of $h(\boldsymbol{\beta})$ with respect to every element of $\boldsymbol{\beta}$.
3. **Obtain Avar$(\hat{\boldsymbol{\beta}})$:** Use the asymptotic variance-covariance matrix of the MLE estimates (often the inverse of the Hessian matrix).
4. **Compute the Standard Error:** The estimated standard error is $\sqrt{\text{Avar}(h(\hat{\boldsymbol{\beta}}))}$.

Reporting Partial Effects

- Since partial effects depend on the values of $\mathbf{x}$ we need to decide how to report estimated partial effects
- **Partial Effects at the Average (PEA)**: $g(\overline{x}\beta)\beta_k$
  - For some x_k the average may ill-defined e.g. binary or categorical variables - no one has $\overline{x}_j$ in the population.
  - Even if all continuous variables, $\overline{x}$ may still not describe anyone.
- **Average Partial Effects (APE)**: $\overline{g(\mathbf{x}\beta)\beta_k}$
  - 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 LinearModels.py and estiation.py which you'll need for OLS and MLE, respectively
- Fill in probit_ante.py and logit_ante.py and solve the problem set
- Q8 on the "Delta Method" is perhaps the most important question. The delta method will be essential for inference.