Lecture 18

Linear probability model

$$y_i = \alpha + \beta x_i + u_i$$

where $y_i \in \{0, 1\}$.

Problems:

1. u_i not normal;
2. u_i heteroskedastic.

Solution: as usual, re-weight by the reciprocal of the estimated std. deviation.

Probit model

$$y_i^* = \alpha + \beta x_i + u_i$$

where y_i^* is unobservable. What is actually observed is:

$$y_i = \begin{cases} 1, & \alpha + \beta x_i + u_i > 0 \\ 0, & \alpha + \beta x_i + u_i < 0 \end{cases}$$

This has likelihood function:

$$\mathcal{L} = \prod_{y_i=0} F\left(\frac{-\alpha - \beta x_i}{\sigma_i}\right) \prod_{y_i=1} \left[1 - F\left(\frac{-\alpha - \beta x_i}{\sigma_i}\right)\right]$$

Solution: MLE.

Logit model

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta x + u$$

where $0 < p < 1$.

Estimate this equation, then find:

$$p = \frac{1}{1 + e^{-(\alpha + \beta x + u)}}$$

Advantage: guaranteed that $0 < \hat{p}_i < 1$.

Lessons:

1. If data is such that $0 < y_i < 1$, use OLS to estimate $\ln\left(\frac{y_i}{1-y_i}\right)$.
2. If $y_i \in \{0, 1\}$, use MLE.

Limited Dependent Variables

Consider the model:

$$y_i = \alpha + \beta x_i + u_i$$

where y_i cannot be negative. In other terms:

$$y_i = \begin{cases} \alpha + \beta x_i + u_i, & y_i > 0 \\ 0, & y_i \leq 0 \end{cases}$$

- The truncation of the data will produce a **biased slope** and a **biased intercept**.

If $u \sim \mathcal{N}(0, \sigma^2)$, **use MLE**.

Maximum Likelihood Estimation

Suppose that $X \sim \mathcal{N}(\mu, 10)$ (known variance) and one draws a sample of, say, 500 observations.

- Suppose the sample mean is $\bar{X} = 24.5$.
- Then, most likely, the data were drawn from $\mathcal{N}(24.5, 10)$.

<u>More generally</u>:

Let X be a random variable with a probability density function depending on unknown parameters $\boldsymbol{\theta}$.

Example:

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(x-\mu)}{2\sigma^2} \right\}$$

If we assume draws are independent, for a random sample $x_1, x_2, \ldots, x_n$, then $f(x_1|\boldsymbol{\theta}) f(x_2|\boldsymbol{\theta}) \ldots f(x_n|\boldsymbol{\theta})$ is called the **likelihood function**.

$$\mathcal{L}(\boldsymbol{\theta}|x) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})$$

To maximize $\mathcal{L}$, choose $\boldsymbol{\theta}$ such that $\dfrac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} = 0, \ \dfrac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\theta}^2} < 0.$

To simplify, take the log ([monotonic transformation](), i.e., it does not change the maximum).

$$\ln \mathcal{L}(\boldsymbol{\theta}|x) = \sum_{i=1}^{n} \ln f(x_i|\boldsymbol{\theta})$$

Then, choose $\boldsymbol{\theta}$ such that $\dfrac{\partial \ln \mathcal{L}}{\partial \boldsymbol{\theta}} = 0, \ \dfrac{\partial^2 \ln \mathcal{L}}{\partial \boldsymbol{\theta}^2} < 0.$

Properties of MLE:

1. Consistent, but biased in small samples;
2. Asymptotically efficient, that is, for large n, no other consistent estimator has a smaller variance.
3. The estimates are asymptotically normal (true even if underlying distribution of X is non-normal).

Example 1:

Assume $X \sim \mathcal{N}(\mu, \sigma^2)$ and a sample $x_1, x_2, \ldots, x_n$.

For individual observations:

$$f(x_i|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\ \exp\left\{-\frac{(x_i-\mu)}{2\sigma^2}\right\}$$

The likelihood function:

$$\mathcal{L}(\hat{\mu},\hat{\sigma}^2|x) = \prod_{i=1}^{n}\left[\frac{1}{\sigma\sqrt{2\pi}}\ \exp\left\{-\frac{(x_i-\mu)}{2\sigma^2}\right\}\right]$$

The log-likelihood function:

$$\ln\mathcal{L}(\hat{\mu},\hat{\sigma}^2|x) = -n\ln\hat{\sigma} - \frac{n}{2}\ln 2\pi - \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$$

Taking the derivative for $\hat{\mu}$:

$$\begin{aligned}\frac{\partial\ln\mathcal{L}}{\partial\hat{\mu}} &= -\frac{2}{2\hat{\sigma}^2}\sum_{i=1}^{n}(x_i-\hat{\mu})(-\hat{\mu}) := 0\\ &= \sum_{i=1}^{n}(x_i-\hat{\mu}) := 0\\ &\therefore \sum x_i = n\hat{\mu} \quad\Rightarrow\quad \hat{\mu} = \frac{1}{n}\sum x_i\end{aligned}$$

Taking the derivative for $\hat{\sigma}$:

$$\begin{aligned}\frac{\partial\ln\mathcal{L}}{\partial\hat{\sigma}} &= -\frac{n}{\hat{\sigma}} - \left(-\frac{4\hat{\sigma}}{4\hat{\sigma}^4}\right)\sum(x_i-\mu)^2 := 0\\ &= -\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^3}\sum(x_i-\mu)^2 := 0\\ &\therefore\ \frac{1}{\sigma^2}\sum(x_i-\mu)^2 = N \quad\Rightarrow\quad \hat{\sigma}^2 = \frac{1}{n}\sum(x_i-\mu)^2\end{aligned}$$

For an estimated mean (and unbiased estimator of the variance): $\hat{\sigma}^2 = \frac{1}{n-1}\sum(x_i-\bar{x})$.

Example 2:

Suppose:

$$y_i = x_i\beta + u_i$$

Assuming $u_i \sim \mathcal{N}(0,\sigma^2)$, then:

$$\begin{aligned}\mathcal{L}(\hat{\beta},\sigma|x) &= \prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{\hat{u}_i^2}{2\sigma^2}\right\}\\ &= \prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(y_i-x_i\beta)^2}{2\sigma^2}\right\}\end{aligned}$$

Taking the $\log$:

$$\ln\mathcal{L}(\hat{\beta},\sigma|x) = -n\ln\sigma - \frac{n}{2}\ln 2\pi - \frac{1}{2\sigma^2}\sum(y_i-x_i\beta)^2$$

Thus, $\max_{\hat{\beta}}\ \ln\mathcal{L}$ is equivalent to $\min_{\hat{\beta}}\ \mathrm{SSE} = \sum(y_i-x_i\beta)^2$.