Lecture notes XI

Regression with a binary dependent variable

- SBW ch. 11

Let Y_i be a binary (dummy) dependent variable. For example:

$$Y_i = \begin{cases} 1 & \text{if firm } i \text{ exits/closes down} \\ 0 & \text{else} \end{cases}$$

Linear probability model

(1) $$Y_i = \beta_0 + \beta_1 X_i + u_i$$ (one regressor for simplicity)

Since Y_i can take only 2 values:

$$E(Y_i | X_i) = 0 \cdot P_r(Y_i = 0 | X_i) + 1 \cdot P_r(Y_i = 1 | X_i)$$
$$= P(Y_i = 1 | X_i)$$

If the ~~a~~ OLS-assumptions are satisfied, then $E(u_i | X_i) = 0 \Rightarrow E(Y_i | X_i) = \beta_0 + \beta_1 X_i$

Hence: $\Pr(Y_i = 1 | X_i) = \beta_0 + \beta_1 X_i$

Partial effects of a change in X_i is then

$$\frac{\partial \Pr(Y_i = 1 | X_i)}{\partial X_i} = \beta_1$$

But Y_i can take only 2 values $\Rightarrow$

$$Y_i = 1 \Leftrightarrow u_i = 1 - \beta_0 - \beta_1 X_i$$
$$Y_i = 0 \Leftrightarrow u_i = - \beta_0 - \beta_1 X_i$$

Then

$$\Pr(Y_i = 1 | X_i) = \Pr(u_i = 1 - \beta_0 - \beta_1 X_i) = \beta_0 + \beta_1 X_i$$
$$\Pr(Y_i = 0 | X_i) = \Pr(u_i = - \beta_0 - \beta_1 X_i) = 1 - \beta_0 - \beta_1 X_i$$

Thus distr. of u_i cannot be independent of X_i

Thus

$$E(u_i | X_i) = (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i) + (-\beta_0 - \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i)$$

$$= 0$$

$$Var(u_i | X_i) = E(u_i^2 | X_i) = (1 - \beta_0 - \beta_1 X_i)^2 (\beta_0 + \beta_1 X_i) + (-\beta_0 - \beta_1 X_i)^2 (1 - \beta_0 - \beta_1 X_i)$$

$$= (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i)\left[(1 - \beta_0 - \beta_1 X_i) + \beta_0 + \beta_1 X_i\right]$$

$$= (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i)$$

Thus u_i is heteroscedastic $\Rightarrow$ ols not efficient estimator

More importantly: Since $0 \leq Pr(Y_i = 1 | X_i) \leq 1$

$\Rightarrow$ $0 \leq \beta_0 + \beta_1 X_i \leq 1$, must have that $0 \leq \hat{\beta}_0 + \hat{\beta}_1 X_i \leq 0$, which may not hold for <u>all</u> possible X_i

If $0 \leq \beta_0 + \beta_1 X_i \leq 1$, then $E(u_i | X_i) \neq 0$

Probit and logit models

The Probit model is defined as

$$Pr(Y_i = 1 | X_i) = \Phi(\beta_0 + \beta_1 X_i)$$

where $\Phi()$ is cdf of standard normal distrib.
~~The generalization to k regressors is straightforward.~~

The Logit model is specified as

$$Pr(Y_i = 1 | X_i) = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)}$$

In both Probit and Logit model the probabilities are between 0 and 1

K regressors:

Let $z = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}$

$X_i = (X_{1i}, \ldots, X_{ki})$

then Probit $= \Phi(z)$

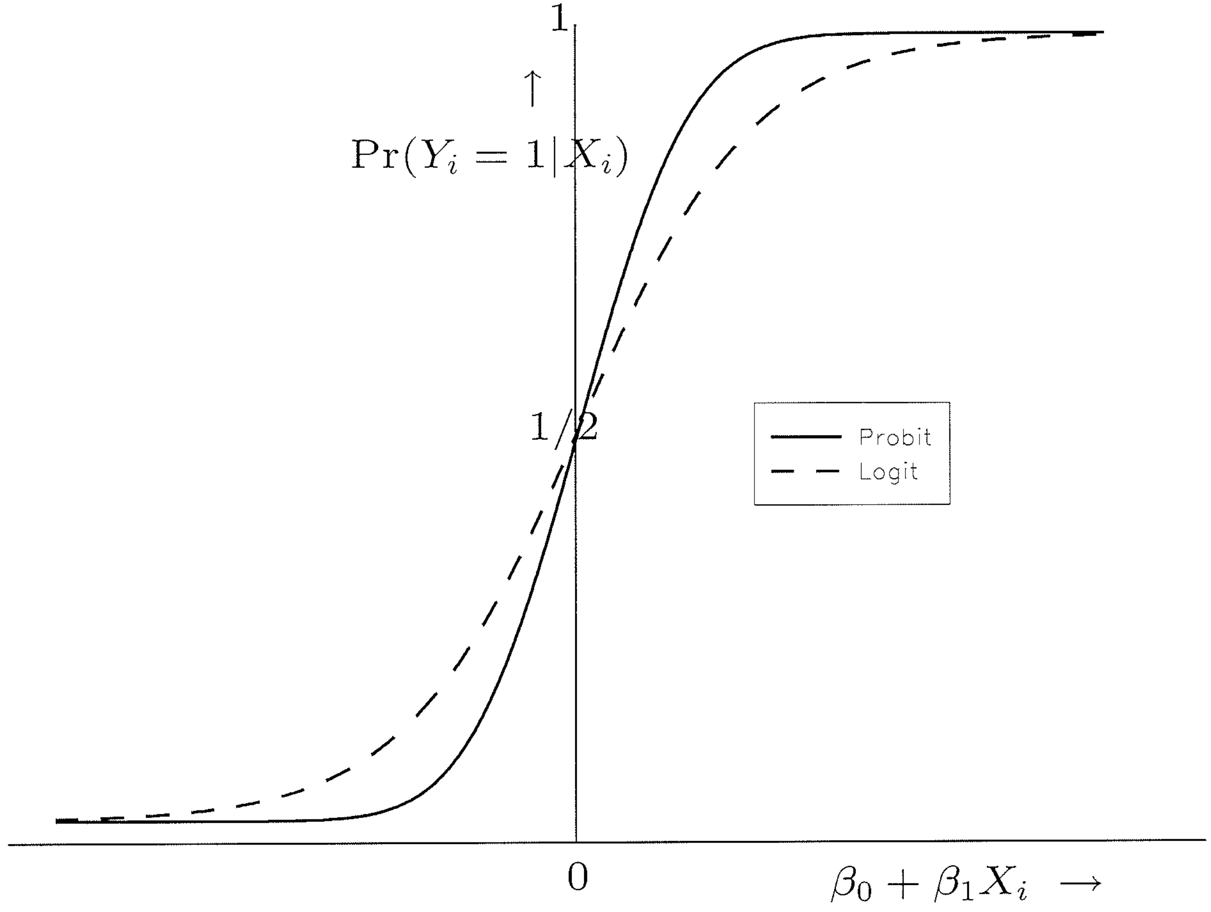
$$\text{Logit} = \frac{e^{z}}{1+e^{z}} = \frac{1}{1+e^{-z}}$$

Note that $\Phi(0) = \dfrac{e^{0}}{1+e^{0}} = \dfrac{1}{2}$

Moreover: $\lim\limits_{z \to \infty} \Phi(z) = \phi \qquad \lim\limits_{z \to -\infty} \Phi(z) = 0$

$$\lim_{z \to \infty} \frac{e^{z}}{1+e^{z}} = 0 \qquad \lim_{z \to \infty} \frac{e^{z}}{1+e^{z}} = 1$$

6



Interpreting partial effects in the Probit and Logit model is more difficult than in the linear probability model

Probit model:

$$\frac{\partial \Pr(Y_i=1 \mid X_i)}{\partial X_{1i}} = \phi(z) \cdot \beta_1$$

Logit model:

$$\frac{\partial \Pr(Y_i=1 \mid X_i)}{\partial X_{1i}} = \frac{\partial}{\partial X_{1i}}\left(\frac{1}{1+e^{-z}}\right) = \frac{-1}{(1+e^{-z})^2}\, e^{-z}(-)\beta_1$$

$$= \frac{e^{-z}}{(1+e^{-z})^2}\,\beta_1 = \frac{e^{-z}}{1+e^{-z}}\, \Pr(Y_i=1 \mid X_i)\,\beta_1$$

$$= \left(1 - \frac{1}{1+e^{-z}}\right) \Pr(Y_i=1 \mid X_i)\,\beta_1$$

$$= \left(1 - \Pr(Y_i=1 \mid X_i)\right) \Pr(Y_i=1 \mid X_i)\,\beta_1$$

Note that, as $Z \to -\infty$ or $Z \to \infty$

in both Logit and Probit models

$$\frac{\partial \Pr(Y_i = 1 \mid X_i)}{\partial X_{1i}} \to 0$$

Binary choice models are often written in terms of a latent structure, i.e. some unobserved (latent) variable Y_i^* follows a linear regression model:

$$Y_i^* = \beta_0 + \beta_1 X_i + u_i,$$

where all classical assumptions are satisfied.

The observed binary dependent (outcome) variable Y_i has the following form:

$$Y_i = \begin{cases} 1 & \text{if } Y_i^* \geq 0 \\ 0 & \text{if } Y_i^* < 0 \end{cases}$$

The probability that $Y_i = 1$ conditional on X_i equals

$$\Pr(Y_i = 1 \mid X_i) = \Pr(Y_i^* \geqslant 0 \mid X_i)$$

$$= \Pr(\beta_0 + \beta_1 X_i + u_i \geqslant 0 \mid X_i)$$

$$= \Pr(u_i \geqslant -\beta_0 - \beta_1 X_i)$$

If $u_i \sim N(0,1)$, then

$$\Pr(u_i \geqslant -\beta_0 - \beta_1 X_i) = 1 - \Phi(-\beta_0 - \beta_1 X_i)$$

$$= \Phi(\beta_0 + \beta_1 X_i) \quad \text{— i.e. Probit}$$

The Probit and Logit models are estimated using Maximum Likelihood estimation (ML)

$$\log(\mathcal{L}) = \sum_{i=1}^{n} Y_i \ln Pr(Y_i = 1 \mid X_i) + (1 - Y_i)(1 - Pr(Y_i = 1 \mid X_i))$$

There are no easy expressions for the estimator, therefore numerical optimization must be used.

The ML estimator θ is efficient, i.e. among the consistent estimators they have the lowest variance.

Furthermore, the ML estimator is normally distributed when $n \to \infty$, therefore t and F-tests can be used for hypothesis testing.