Toulouse School of Economics

M1 INTERMEDIATE ECONOMETRICS

Binary choice models

Koen Jochmans   François Poinas

2024 — 2025

Overview

This deck of slides goes over the binary-choice model and estimation by maximum likelihood.

The relevant chapter in Hansen is 25.

H25.7 and H25.8 concern asymptotics which we will deal with later.

Conditional on covariates X, Bernoulli variable $Y \in \{0,1\}$ have success probability

$$\mathbb{P}(Y = 1|X) = \mathbb{E}(Y|X).$$

This CEF is bounded between zero and one and so is nonlinear, in general.

The exception is when the CEF can be saturated.

The simplest example has a single binary regressor $X \in \{0,1\}$. Then

$$\mathbb{E}(Y|X) = \beta_1 X + \beta_0 (1 - X) = \beta_0 + (\beta_1 - \beta_0)X$$

for $\beta_0 = \mathbb{P}(Y = 1|X = 0)$ and $\beta_1 = \mathbb{P}(Y = 1|X = 1)$.

This can be generalized when all regressors take on a small number of possible values, but the number of parameters to estimate grows very quickly.

A parsimonious parametrization for binary choice that restricts the CEF to the unit-interval is

$$\mathbb{P}(Y = 1|X) = G(\varphi(X, \beta))$$

for a chosen function φ and CDF G.

In practice the most popular choice for $\varphi(x, \beta)$ is $x'\beta$, in which case we get

$$\mathbb{P}(Y = 1|X) = G(X'\beta);$$

the logit model and the probit model are the most popular versions of this specification.

They differ only in the choice of G.

We have previously shown that NLLS can be used to estimate the parameters in binary-choice models.

Also know that NLLS is not the optimal choice.

The optimal choice is maximum likelihood.

Maximum likelihood is a general technique that we introduce through this example.

The likelihood function

Suppose that we observe a random sample $(Y_1, X_1), \ldots, (Y_n, X_n)$.

The sequence $Y_1, \ldots, Y_n$ is a sequence of zeros and ones.

Conditional on $X_1, \ldots, X_n$ the probability of observing this particular sequence is

$$\prod_{i=1}^{n} \mathbb{P}(Y_i = 1|X_i)^{\{Y_i=1\}} \left(1 - \mathbb{P}(Y_i = 1|X_i)\right)^{\{Y_i=0\}}.$$

With our model for the conditional probability this becomes

$$L_n(\beta) = \prod_{i=1}^{n} G(X_i'\beta)^{\{Y_i=1\}} \left(1 - G(X_i'\beta)\right)^{\{Y_i=0\}}.$$

For any b, $L_n(b)$ given the probability of observing the sample in front of us if the data were generated with $\mathbb{P}(Y = 1|X) = G(X'b)$. This is the likelihood function.

We estimate β by maximizing this probability.

The MLE of β is $\hat{\beta}_{\text{mle}}$. Thus,

$$L_n(\hat{\beta}_{\text{mle}}) \geq L_n(b)$$

for any b.

Log-likelihood and optimization

It is usually easier to work with

$$\ell_n(b) = \log L_n(b) = \sum_{i=1}^{n} Y_i \log G(X_i'b) + (1 - Y_i) \log(1 - G(X_i'b)).$$

This is the log-likelihood function.

In regular situations, $\hat{\beta}_{\text{mle}}$ solves the first-order condition

$$\frac{\partial \ell_n(b)}{\partial b} = 0.$$

Here,

$$\begin{aligned}
\frac{\partial \ell_n(b)}{\partial b} &= \sum_{i=1}^{n} Y_i X_i \frac{g(X_i'b)}{G(X_i'b)} - (1 - Y_i) \frac{g(X_i'b)}{1 - G(X_i'b)} \\
&= \sum_{i=1}^{n} X_i \frac{g(X_i'b)}{G(X_i'b)\,(1 - G(X_i'b))} (Y_i - G(X_i'b)).
\end{aligned}$$

This you will recognize from the set of slides on NLLS.

Average marginal effects

The parameter β may not be the ultimate object of interest.

In our model (for continuous X)

$$\frac{\partial \mathbb{P}(Y=1|X)}{\partial X} = \beta g(X'\beta),$$

and we may be interested in such things as

$$\theta = \mathbb{E}\left(\frac{\partial \mathbb{P}(Y=1|X)}{\partial X}\right) = \mathbb{E}\left(\beta\, g(X'\beta)\right).$$

(This is not the same as $\beta\, g(\mathbb{E}(X')\beta)$, which is not very interesting.)

An estimator of θ is

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_{\text{mle}}\, g(X_i'\hat{\beta}_{\text{mle}}).$$