Advanced Econometrics

par Phu Nguyen-Van

December 3, 2024

Contents

Chapter 1

Limited Dependent Varibale Model

A Limited Dependent Variable (LDV) is a **dependent variable whose range of values is restricted**. When the dependent variable represents multiple discrete outcomes, we face *multinomial choice models* (N.B. : utile pour le travail avec Corinne !).

A dependent variable can also be constrained (*censored models*) or bounded (*truncated models*).

For instance, $y_i \in \{1, \dots, 6\}$ is a limited dependent variable.

Latent variable: unobserved variable.

[...]

1.1 Binary Models

1.1.1 Goodness-of-Fit

A goodness-of-fit is given by:

$$\text{Pseudo-}R^2 = 1 - \frac{1}{1 + 2N^{-1}(\ln L - \ln L_0)}$$

1.2 Multiresponse models

When the number of alternatives to be chosen is larger than 2. If there is a logical ordering of the alternatives, in particular when the underlying latent variable drives the choices between the alternatives, we can use the **ordered response model**.

Unordered response model arise when there is no logical ordering.

1.2.1 Ordered Response Model

Assume there is a logical ordering between M alternatives. The model is describe by:

$$y_i^* = x_i\beta + u_i$$
$$y_i = j \text{ if } \alpha_{j-1} < y_i^* \leq \alpha_j$$

In this model, the intercept is set to zero by normalization.
If we assume the error term u_i to be:

- normally distributed, we use the **ordered probit model**
- logistic distributed, we use the **ordered logit model**

Multinomial probit model : réduire à 3 (hiérarchisées) maximum pour l'interprétabilité.
Difficulté technique : dans multinomial logit, on doit supposer IIA.
Le ratio entre deux catégories ne dépend que des estimations des coefficients des deux catégories.

Plus facile à interpréter en utiilisant le calcul du *marginal effect*.

In the case where the explanatory variables are different for each category, we should use the conditional logit model:

$$P(y_i = j) = \frac{e^{\alpha_j + \beta x_{ij}}}{\sum_{k=1}^{M} e^{\alpha_k + \beta x_{ik}}}$$

1.3 Censored and Truncated Models

Standard linear models assume that we observe all values of the dependent variable, *i.e.* there is no missing informations.

However in practice, we only observe the dependent variable from a certain threshold, named **truncated variable**. Observations above (or below depending on the observation) this threshold cannot be observed.

Truncated models: observations are completely missing.

1.3.1 Sample Selection Model

Tobit II model ; Heckman's two-step procedure (Heckit model).
Heckit:

1. Estimate the selection equation
2. Estimate the main equation (Linear model with heteroskedasticity). Using the inverse Mills ratio would eliminate of the selection bias.

Chapter 2

Panel Data Models

We have both time series and cross-sectional elements. Such a dataset is known as panel data or longitudinal data.

$$y_{it} = \alpha + \beta x_{it} + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T$$

In this chapter, we only deal with balanced panel data: we have the same number of time periods by observation units.