Econometrics 2

Lecture 17: Likelihood Theory (Cont.) & The Probit Model

AUEB | Spring Semester 2026
Taught by: Prof. S. Arvanitis
Notes transcribed & digitized by: T. Zevlikaris & T. Kourtalis

Note: These notes digitize the handwritten derivations from class and are intended to be studied in conjunction with the slides: probit.pdf.

1. General Framework of Likelihood

As established, parametric models assign a unique distribution $p(z_n, \beta)$ to each β. The joint density for a sample z_n is decomposed as:

$$\begin{aligned} p(z_n, \beta) =& p(z_n | z_{n-1}, \dots, z_1, \beta) \cdot \\ & p(z_{n-1} | z_{n-2}, \dots, z_1, \beta) \cdots p(z_1, \beta) \end{aligned}$$

In the case of independence, this simplifies to the product of marginal distributions:

$$p(z_n, \beta) = \prod_{i=1}^{n} p(z_i, \beta)$$

The average log-likelihood function used for estimation is:

$$l_n(\beta, z_n) = \frac{1}{n} \sum_{i=1}^{n} \ln p(z_i, \beta)$$

2. Linear Model with t-distributed Errors

We consider the linear model where the error terms $\varepsilon(i)$ follow a Student's t-distribution with ν degrees of freedom ($\nu > 0$):

$$\varepsilon(i) | X_n \sim t_\nu$$

The probability density function (PDF) is given by:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

For an observation $y_{(i)} = x_{(i)}\beta + \varepsilon(i)$, the density is:

$$p(y_{(i)}, \beta) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(y_{(i)} - x_{(i)}\beta)^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

The average log-likelihood for this model is:

$$L(\beta, y_n) = \ln\left(\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\right) - \frac{\nu+1}{2n}\sum_{i=1}^{n} \ln\left(1 + \frac{(y_{(i)} - x_{(i)}\beta)^2}{\nu}\right)$$

3. The Probit Model

The Probit model is used for binary outcomes $Y \in \{0, 1\}$. It is founded on a latent utility difference Y^*:

$$Y^*_{(i)} = X_{(i)}\beta + \varepsilon(i), \quad Y(i) = \mathbb{1}\{Y^*_{(i)} > 0\}$$

Identification and Probabilities

Due to **scale invariance**, we normalize the error variance to 1 ($\text{Var}(\varepsilon|X) = 1$). Assuming $\varepsilon(i)|X_n \sim N(0, 1)$, the probability of success is:

$$\begin{aligned} P(Y(i) = 1|X_n) &= P(X_{(i)}\beta + \varepsilon(i) > 0) \\ &= 1 - \Phi(-X_{(i)}\beta) = \Phi(X_{(i)}\beta) \end{aligned}$$

where $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$. The sample log-likelihood is:

$$\ell_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \left[Y(i) \ln \Phi(X_{(i)}\beta) + (1 - Y(i)) \ln(1 - \Phi(X_{(i)}\beta))\right]$$

4. Numerical Optimization

Since $\hat{\beta}^{MLE}$ lacks a closed-form solution, we use iterative methods. The **score vector** (gradient) is defined as:

$$\nabla \ell_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \frac{(Y(i) - \Phi(X_{(i)}\beta))\phi(X_{(i)}\beta)}{\Phi(X_{(i)}\beta)(1 - \Phi(X_{(i)}\beta))} X'_{(i)}$$

Standard algorithms include **Newton-Raphson** and **Fisher Scoring** (which replaces the Hessian with the Information Matrix).