ECON 452* -- NOTE 16

Testing Linear Coefficient Restrictions in Probit Models

This note outlines the procedures for testing hypotheses in probit models.

1. The Probit Model

- The ***unobserved*** **(or *latent*) dependent variable** Y_i^* is assumed to be generated by a classical linear regression model of the form

$$Y_i^* = x_i^T\beta + u_i \tag{1}$$

where:

$Y_i^* =$ a continuous real-valued index variable for observation i that is *unobservable*, or *latent*;

$x_i^T = (1 \;\; X_{i1} \;\; X_{i2} \;\; \cdots \;\; X_{ik})$, a 1×K row vector of regressor values for observation i;

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \text{ the K×1 coefficient vector with elements } \beta_j,\ j = 0, 1, ..., k$$

$u_i =$ an iid $N(0,1)$ random error term for observation i.

- The ***observable outcomes*** **of the binary choice problem** are represented by a **binary indicator variable Y_i** that is related to the unobserved dependent variable Y_i^* as follows:

$$Y_i = 1 \text{ if } Y_i^* > 0 \tag{2.1}$$

$$Y_i = 0 \text{ if } Y_i^* \leq 0 \tag{2.2}$$

- **The probabilities of the two binary outcomes** are formulated in terms of the cumulative distribution function of the standard normal distribution, as follows:

$$\Pr(Y_i = 1) = \Pr(Y_i^* > 0) = \Phi(x_i^T\beta) \tag{3.1}$$

$$\Pr(Y_i = 0) = \Pr(Y_i^* \leq 0) = 1 - \Phi(x_i^T\beta) \tag{3.2}$$

where

$$\Phi(x_i^T\beta) = \Pr(z_i \leq x_i^T\beta) \qquad \text{where } z_i \sim N(0,1) \text{ is a standard normal index.}$$

2. Formulation of Linear Equality Restrictions on β

The general hypothesis to be tested is that the probit coefficient vector β satisfies a set of q independent linear restrictions, where $q < K$. We formulate this general hypothesis in vector-matrix form, since this corresponds to the way in which econometric software such as *Stata* is written.

The **null hypothesis H_0** is written in general as:

$$H_0: \quad R\beta = r \quad \Leftrightarrow \quad R\beta - r = \underline{0}$$

The **alternative hypothesis H_1** is written in general as:

$$H_1: \quad R\beta \neq r \quad \Leftrightarrow \quad R\beta - r \neq \underline{0}$$

In H_0 and H_1 above:

R = a q×K matrix of specified constants;

β = the K×1 coefficient vector;

r = a q×1 vector of specified constants;

$\underline{0}$ = a q×1 null vector, i.e., a q×1 vector of zeros.

- The q×K restrictions matrix R takes the form

$$R = \begin{bmatrix} r_{10} & r_{11} & r_{12} & \cdots & r_{1k} \\ r_{20} & r_{21} & r_{22} & \cdots & r_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{q0} & r_{q1} & r_{q2} & \cdots & r_{qk} \end{bmatrix}$$

where

r_{mj} = the constant on coefficient β_j in the m-th linear restriction, m = 1, …, q.

- The q×1 restrictions vector r takes the form

$$r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_q \end{bmatrix}$$

where

r_m = the constant term in the m-th linear restriction, m = 1, …, q.

- The matrix-vector product $R\beta$ is a q×1 vector of linear functions of the regression coefficients β_0, β_1, β_2, … , β_k:

$$R\beta = \underset{(q\times K)}{\begin{bmatrix} r_{10} & r_{11} & r_{12} & \cdots & r_{1k} \\ r_{20} & r_{21} & r_{22} & \cdots & r_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{q0} & r_{q1} & r_{q2} & \cdots & r_{qk} \end{bmatrix}} \underset{(K\times 1)}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}} = \underset{(q\times 1)}{\begin{bmatrix} r_{10}\beta_0 + r_{11}\beta_1 + r_{12}\beta_2 + \cdots + r_{1k}\beta_k \\ r_{20}\beta_0 + r_{21}\beta_1 + r_{22}\beta_2 + \cdots + r_{2k}\beta_k \\ \vdots \\ r_{q0}\beta_0 + r_{q1}\beta_1 + r_{q2}\beta_2 + \cdots + r_{qk}\beta_k \end{bmatrix}}$$

- The null and alternative hypotheses can therefore be written as follows:

$$H_0\text{: } R\beta = r \quad \Rightarrow \quad \begin{bmatrix} r_{10}\beta_0 + r_{11}\beta_1 + r_{12}\beta_2 + \cdots + r_{1k}\beta_k \\ r_{20}\beta_0 + r_{21}\beta_1 + r_{22}\beta_2 + \cdots + r_{2k}\beta_k \\ \vdots \\ r_{q0}\beta_0 + r_{q1}\beta_1 + r_{q2}\beta_2 + \cdots + r_{qk}\beta_k \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_q \end{bmatrix}$$

$$H_1\text{: } R\beta \neq r \quad \Rightarrow \quad \begin{bmatrix} r_{10}\beta_0 + r_{11}\beta_1 + r_{12}\beta_2 + \cdots + r_{1k}\beta_k \\ r_{20}\beta_0 + r_{21}\beta_1 + r_{22}\beta_2 + \cdots + r_{2k}\beta_k \\ \vdots \\ r_{q0}\beta_0 + r_{q1}\beta_1 + r_{q2}\beta_2 + \cdots + r_{qk}\beta_k \end{bmatrix} \neq \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_q \end{bmatrix}$$

3. Null and Alternative Hypotheses

❑ Null and Alternative Hypotheses

- The **null hypothesis** is that the coefficient vector β satisfies a set of q independent linear coefficient restrictions:

$$H_0: \quad R\beta = r \quad \Leftrightarrow \quad R\beta - r = \underline{0}$$

- The **alternative hypothesis** is that the coefficient vector β does not satisfy the set of q independent linear coefficient restrictions specified by H_0:

$$H_1: \quad R\beta \neq r \quad \Leftrightarrow \quad R\beta - r \neq \underline{0}$$

❑ Two Alternative Estimators of the Probit Coefficient Vector β

1. The ***restricted*** **coefficient estimates** computed under H_0: $R\beta - r = \underline{0}$, which are denoted as follows:

$\tilde{\beta}$ = the ***restricted*** ML estimator of the probit coefficient vector β;

$$\ln \hat{L}_0 = \sum_{i=1}^{N} Y_i \ln \Phi\left(x_i^T \tilde{\beta}\right) + \sum_{i=1}^{N} (1 - Y_i) \ln\left[1 - \Phi\left(x_i^T \tilde{\beta}\right)\right]$$

= the ***restricted*** **maximized loglikelihood value**;

$K - q$ = the number of free probit coefficients in the ***restricted*** **model**.

2. The ***unrestricted*** **coefficient estimates** computed under H_1: $R\beta - r \neq \underline{0}$, which are denoted as follows:

$\hat{\beta}$ = the ***unrestricted*** ML estimator of the probit coefficient vector β;

$$\ln \hat{L}_1 = \sum_{i=1}^{N} Y_i \ln \Phi\left(x_i^T \hat{\beta}\right) + \sum_{i=1}^{N} (1 - Y_i) \ln\left[1 - \Phi\left(x_i^T \hat{\beta}\right)\right]$$

= the ***unrestricted*** **maximized loglikelihood value**;

K = the number of free probit coefficients in the ***unrestricted*** **model**.

4. Likelihood Ratio Tests of Linear Coefficient Restrictions

❑ The Likelihood Ratio Statistic

The LR statistic essentially compares the maximized loglikelihood values for the restricted and unrestricted probit coefficient estimates $\tilde{\beta}$ and $\hat{\beta}$.

$\ln \hat{L}_0$ = the ***restricted* maximized loglikelihood value** corresponding to the restricted probit coefficient estimates $\tilde{\beta}$ computed under the null hypothesis H_0.

$$\ln \hat{L}_0 = \sum_{i=1}^{N} Y_i \ln \Phi\left(x_i^T \tilde{\beta}\right) + \sum_{i=1}^{N} (1 - Y_i) \ln\left[1 - \Phi\left(x_i^T \tilde{\beta}\right)\right]$$

$\ln \hat{L}_1$ = the ***unrestricted* maximized loglikelihood value** corresponding to the unrestricted probit coefficient estimates $\hat{\beta}$ computed under the alternative hypothesis H_0.

$$\ln \hat{L}_1 = \sum_{i=1}^{N} Y_i \ln \Phi\left(x_i^T \hat{\beta}\right) + \sum_{i=1}^{N} (1 - Y_i) \ln\left[1 - \Phi\left(x_i^T \hat{\beta}\right)\right]$$

Note: $\ln \hat{L}_0 \le \ln \hat{L}_1 \quad \Rightarrow \quad \ln \hat{L}_0 - \ln \hat{L}_1 \le 0 \quad \Rightarrow \quad -2[\ln \hat{L}_0 - \ln \hat{L}_1] \ge 0$

$\Rightarrow \quad \ln \hat{L}_1 - \ln \hat{L}_0 \ge 0 \quad \Rightarrow \quad 2[\ln \hat{L}_1 - \ln \hat{L}_0] \ge 0$

- **The LR statistic**

$$LR = -2[\ln\hat{L}_0 - \ln\hat{L}_1] = 2[\ln\hat{L}_1 - \ln\hat{L}_0] \sim \chi^2[q] \text{ under } H_0$$

where

q = the **number of *independent linear coefficient restrictions*** specified by the null hypothesis H_0

$\chi^2[q]$ = the ***chi-square* distribution** with ***q* degrees of freedom**.

- **Decision Rule**

***Reject* $\mathbf{H_0}$** against H_1 at significance level α if:

$$LR > \chi^2_\alpha[q] \quad \textit{or} \quad \text{p-value of LR} < \alpha$$

***Retain* $\mathbf{H_0}$** against H_1 at significance level α if:

$$LR \le \chi^2_\alpha[q] \quad \textit{or} \quad \text{p-value of LR} \ge \alpha$$

5. Wald Tests of Linear Coefficient Restrictions

- The **Wald Test is Based on the Wald Principle of Hypothesis Testing**

  The **Wald principle** of hypothesis testing computes hypothesis tests using ***only*** **the *unrestricted* probit coefficient estimates** $\hat{\beta}$ of the model computed under the alternative hypothesis H_1: $R\beta \neq r$.

- **<u>The Wald Statistic</u>**

  The ***Wald test statistic*** **W** for testing the null hypothesis H_0: $R\beta = r$ against the alternative hypothesis H_1: $R\beta \neq r$ takes the form

$$W = \left(R\hat{\beta} - r\right)^T \left(R(\hat{V}[\hat{\beta}])R^T\right)^{-1} \left(R\hat{\beta} - r\right) \sim \chi^2[q] \qquad \text{under } H_0$$

or

$$W = \left(R\hat{\beta} - r\right)^T \left(R\hat{V}_{ML}R^T\right)^{-1} \left(R\hat{\beta} - r\right) \sim \chi^2[q] \qquad \text{under } H_0$$

where:

$\hat{\beta} = \hat{\beta}_{ML} =$ the ***unrestricted*** **ML estimator of probit coefficient vector β**;

$\hat{V}_{ML} = \hat{V}[\hat{\beta}]$ = the **ML estimator of the covariance matrix** $V = V[\hat{\beta}]$ **of** $\hat{\beta}$;

$\chi^2[q]$ = the ***chi-square*** **distribution** with ***q*** **degrees of freedom**.

Note: The ML probit coefficient estimator $\hat{\beta}$ and the ML coefficient covariance matrix estimator $\hat{V}_{ML}$ used in the Wald test statistic W are computed using only ***unrestricted*** **estimates** of the probit model under the alternative hypothesis H_1: $R\beta \neq r$.

- **Decision Rule**

Reject $\mathbf{H_0}$ against H_1 at significance level α if:

$$W > \chi^2_\alpha[q] \quad \textit{or} \quad \text{p-value of } W < \alpha$$

Retain $\mathbf{H_0}$ against H_1 at significance level α if:

$$W \leq \chi^2_\alpha[q] \quad \textit{or} \quad \text{p-value of } W \geq \alpha$$

6. Relationship Between Wald and LR Tests

❑ **The Wald and LR Statistics**

$$W = \left(R\hat{\beta} - r\right)^T \left(R\hat{V}_{ML}R^T\right)^{-1} \left(R\hat{\beta} - r\right) \sim \chi^2[q] \qquad \text{under } H_0$$

$$LR = -2[\ln\hat{L}_0 - \ln\hat{L}_1] = 2[\ln\hat{L}_1 - \ln\hat{L}_0] \sim \chi^2[q] \quad \text{under } H_0$$

❑ **Tests Based on the W and LR Statistics are Not Equivalent**

The Wald statistic W and the Likelihood Ratio statistic LR do not yield equivalent or identical tests of H_0: $R\beta = r$ against H_1: $R\beta \neq r$.

This nonequivalence follows from the fact that **the two test statistics W and LR are *not equal***; that is, they yield different calculated sample values of the test statistic.

$$W \neq LR$$

But the two test statistics W and LR do have *identical null distributions*, namely the $\chi^2[q]$ distribution.

$$W \sim \chi^2[q] \quad \text{and} \quad LR \sim \chi^2[q] \qquad \text{under} \quad H_0\text{: } R\beta = r$$

- **<u>Result</u>:** The Wald and LR tests of the same null hypothesis can yield ***different* inferences**.