12 Probit Asymptotics. Testing Inequality Restrictions.

12.1 Asymptotics of Probit

The conditional likelihood for the probit model is

$$L(\beta) = \prod_{i=1}^{n} f(y_i|x_i, \beta)$$

$$= \prod_{i=1}^{n} P(y_i = 1|x_i, \beta)^{y_i} P(y_i = 0|x_i, \beta)^{1-y_i}$$

$$= \prod_{i=1}^{n} \Phi(x_i'\beta)^{y_i} (1 - \Phi(x_i'\beta))^{1-y_i}$$

and the conditional log-likelihood is

$$l(\beta) = \sum_{i=1}^{n} y_i \ln \Phi(x_i'\beta) + (1 - y_i) \ln(1 - \Phi(x_i'\beta))$$

Given x_i, y_i we can find $\hat{(}\beta_{ML})$ that maximises this function.

Theorem 12.1.1. The Probit Estimator
(i) $(\hat{\beta}_{prob}) \xrightarrow{p} \beta_0$
(ii) $\sqrt{n}(\hat{\beta}_{prob} - \beta_0) \xrightarrow{d} N(0, I_1^{-1})$ where,

$$I_1 = E\left(\frac{\phi^2 x_i x_i'}{\Phi(1 - \Phi)}\right)$$

where $\Phi = \Phi(x_i'\beta_0)$ and $\phi = \frac{d}{dt}\Phi(t)\big|_{t=(x_i'\beta_0)}$
Under the following assumptions:

- (Prob 0) $\{y_i, x_i\}_{i=1}^n$ is an iid sequence with binary y_i
- (Prob 1) $E(x_i x_i')$ is finite nonsingular
- (Prob 2) $Pr(y_i = 1|x_i) = \Phi(x_i'\beta)$

Proof. The theorem follows from the fact that $\hat{\beta}_{prob}$ is an MLE estimator. Indeed, the consistency statement is implied (as we have assumed correct distribution).

For the asymptotic normality, recall:

$$\sqrt{n}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{d} N(0, n(I^{-1}(\theta_0)))$$

where $\hat{\theta}_{ML} = \hat{\beta}_{prob}$ and $\theta_0 = \beta_0$

$$I(\beta) = Var\left(\frac{d}{d\beta}L(\beta)\right)\left(\overset{\beta=\beta_0}{=} E\left((\frac{d}{d\beta}L(\beta)\frac{d}{d\beta'}L(\beta))\right)\right)$$

Since $E(\frac{d}{d\beta}L(\beta))\big|_{\beta=\beta_0} = 0$, as we have seen previously that this maximises the likelihood function.

$$\begin{aligned}\frac{dL(\beta)}{d\beta} &= \sum_{i=1}^{n}\left(\frac{y_i}{\Phi(x_i'\beta)}\phi(x_i'\beta)x_i - \frac{(1-y_i)}{1-\Phi(x_i'\beta)}\phi(x_i'\beta)x_i\right) \\ &= \sum_{i=1}^{n}\frac{y_i-\Phi(x_i'\beta)}{\Phi(x_i'\beta)(1-\Phi(x_i'\beta))}\phi(x_i'\beta)x_i\end{aligned}$$

$$Var\frac{dL}{d\beta} = E(Var\frac{dL}{d\beta}|x_i) + Var(E(\frac{dL}{d\beta}|x_i))$$

$$Var(E(\frac{dL}{d\beta}|x_i))\big|_{\beta=\beta_0} = Var(0|x_i) = 0$$

$$E(Var\frac{dL}{d\beta}|x_i)\big|_{\beta=\beta_0} = EVar\left(\sum_{i=1}^{n}\frac{y_i-\Phi(x_i'\beta)}{\Phi(x_i'\beta)(1-\Phi(x_i'\beta))}\phi(x_i'\beta)x_i|x_i\right)$$

Because iid,

$$= nEVar\left(\frac{y_i-\Phi(x_i'\beta)}{\Phi(x_i'\beta)(1-\Phi(x_i'\beta))}\phi(x_i'\beta)x_i|x_i\right)$$

since $\Phi(x_i'\beta)$ is constant conditioning on x_i:

$$= nEVar\left(\frac{y_i}{\Phi(x_i'\beta)(1-\Phi(x_i'\beta))}\phi(x_i'\beta)x_i|x_i\right)$$

$$= nE(\frac{\phi}{\Phi(1-\Phi)}x_iVar(y_i|x_i)\frac{\phi}{\Phi(1-\Phi)}x_i')$$

$Var((y_i|x_i))$ is given by:

$$E(y_i^2|x_i) - E(y_i|x_i)^2$$

by Prob 2:

$$= (\Phi(1^2)) - (\Phi(1))^2 = \Phi - \Phi^2$$

Thus:

$$I(\beta_0) = nE\left(\frac{\phi^2 x_i x_i'}{\Phi(1-\Phi)}\right)$$

$$I_1(\beta_0) = E\left(\frac{\phi^2 x_i x_i'}{\Phi(1-\Phi)}\right)$$

□

12.1.1 Interpreting coefficients in Probit

Unlike for linear regression, $\beta = \beta_0$ cannot be interpreted as the marginal effect of x on y. Here the coefficient measures the drect impact of regressors only on the (unobserved and scaled) underlying index. We are not typically interested in this slope but actually:

Definition 12.1.1

Probit Marginal Effects

$$\frac{\partial Pr(y=1|x)}{\partial x_j} = \frac{\partial}{\partial x_j}(x'\beta) = \phi(x'\beta)\beta_j$$

here x_j refers to the j-th component of vector x as opposed to the j-th observation of this vector.
Note that the effect depends not only on the value of x_j but also other variables.
Marginal Effect at the Average:

$$\frac{\partial Pr(y=1|x)}{\partial x_j}\Big|_{x=\bar{x}} = \phi(\bar{x}'\beta)\beta_j$$

Standard errors, use the delta method, where $\theta = \beta$

$$\frac{\partial Pr(y=1|x)}{\partial x_j}\Big|_{x=\bar{x}} = \phi(\bar{x}'\beta)\beta_j = g(\beta)$$

Thus we have

$$\hat{Var}(\phi(\bar{x}'\beta)\beta_j) = \frac{1}{n}\frac{\partial g(\hat{\beta})}{\partial \beta'}\hat{I}_1^{-1}\frac{\partial g(\hat{\beta})}{\partial \beta}$$

Average Marginal Effect:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial Pr(y=1|x)}{\partial x_j}\Big|_{x=x_i} = \frac{1}{n}\sum_{i=1}^{n}\phi(x_i'\beta)\beta_j$$

Note:-

The ratios of the effects of two variables are equal to the ratio of their coefficients, and are therefore comparable for probit and logit models.

$$= \frac{\frac{\partial Pr(y=1|x)}{\partial x_j}}{\frac{\partial Pr(y=1|x)}{\partial x_k}} = \frac{\phi(x'\beta)\beta_j}{\phi(x'\beta)\beta_k} = \frac{\beta_j}{\beta_k}$$

Thus, while $\hat{\beta}_j$ and $\hat{\beta}_k$ are not directly interpretable as the absolute marginal effects, their ratio can be interpreted as the ratio of the marginal effects.

12.2 Testing Inequality Constraints

Instead of simply deleting or attempting to explain inconsistent signs of parameters in the estimating equation, we can statistically test whether or not the signs of the true values of these estimates are consistent with researcher beliefs.

Consider the stylised framework, of a normal regression model with two explanatory variables:

$$Y = \beta_1 X1 + \beta_2 X2 + \varepsilon$$

where $\varepsilon|X_1, X_2 \sim N(0, \sigma^2 I_n)$ In addition, assume $\sigma^2 = 1$ known and X_1, X_2 orthonormal. That is, $X'X = I_2$ where $X = [X_1, X_2]$

Suppose that we would like to test:

$$H_0 : \beta_1 \geq 0 \text{ and } \beta_1 \geq 0 \text{ vs. } H_1 : \beta_1 < 0 \text{ or } \beta_1 < 0$$

Let's derive the LR statistic. Recall that we assumed that it is known that $\sigma^2 = 1$. Thus, the log-likelihood function is: $\log(maxL(Y, \theta|X))$ without the restrictions:

$$= -\frac{n}{2} log(2\pi) - \frac{1}{2}(Y - X\hat{\beta}_{OLS})'(Y - X\hat{\beta}_{OLS})$$

$\log(maxL(Y, \theta|X))$ with the restrictions:

$$= -\frac{n}{2} log(2\pi) - \frac{1}{2} \min_{b_1, b_2 \geq 0} (Y - Xb)'(Y - Xb)$$

Hence,

$$LR = \min_{b_1, b_2 \geq 0} (Y - Xb)'(Y - Xb) - \frac{1}{2}(Y - X\hat{\beta}_{OLS})'(Y - X\hat{\beta}_{OLS})$$

Note that

$$\begin{aligned}(Y - Xb)'(Y - Xb) &= Y'Y - 2b'X'Y + b'X'Xb \\ &= Y'Y - 2b'X'(X\hat{\beta}_{OLS} + \hat{\varepsilon}) + b'X'Xb \\ &= Y'Y - 2b'\hat{\beta}_{OLS} + b'b\end{aligned}$$

and, similarly

$$(Y - X\hat{\beta}_{OLS})'(Y - X\hat{\beta}_{OLS}) = Y'Y - \hat{\beta}'_{OLS}\hat{\beta}_{OLS}$$

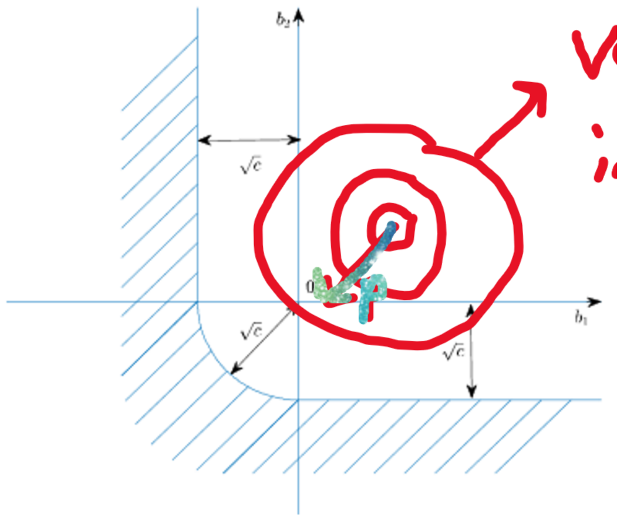
Thus,

$$LR = \min_{b_1, b_2 \geq 0} (\hat{\beta}_{OLS} - b)'(\hat{\beta}_{OLS} - b)$$

so that

$$LR = \begin{cases} 0 & \text{if } \hat{\beta}_{1,OLS} \geq 0 \text{ and } \hat{\beta}_{2,OLS} \geq 0 \\ \hat{\beta}^2_{1,OLS} & \text{if } \hat{\beta}_{1,OLS} < 0 \text{ and } \hat{\beta}_{2,OLS} \geq 0 \\ \hat{\beta}^2_{2,OLS} & \text{if } \hat{\beta}_{1,OLS} \geq 0 \text{ and } \hat{\beta}_{2,OLS} < 0 \\ \hat{\beta}^2_{1,OLS} + \hat{\beta}^2_{2,OLS} & \text{if } \hat{\beta}_{1,OLS} < 0 \text{ and } \hat{\beta}_{2,OLS} < 0 \end{cases}$$

Thus the LR statistic is the squared distance from $\hat{\beta}_{OLS}$ to the positive quadrant in $\mathbb{R}^2$:



Finding c:
If β_{OLS} ends up in the striped region (Ω), LR test rejects. For the test with 5% significance level, we need to choose the critical value c so that

$$\max_{\beta_1,\beta_2\geq 0} Pr(LR > c) = 0.05$$

We can think of $Pr(LR > c)|\beta = \beta_0$ as the volume of the probability density of $\hat{\beta}_{OLS}$ that lies inside the critical region Ω.

Recall that $\hat{\beta}_{OLS}|X \sim N(\beta_0, \sigma^2(X'X)^{-1})$ In this special case:

$$\hat{\beta}_{OLS}|X \sim N(\beta_0, I_2)$$

Thus with this geometric interpretation

$$Pr(LR > c) = \int_{\Omega} \frac{1}{2\pi} exp\{-\frac{(z-\beta)'(z-\beta)}{2}\}dz$$
$$= \int_{\Omega-\beta} \frac{1}{2\pi} exp\{-\frac{z'z}{2}\}dz$$

where $\Omega - \beta$ a linear shift of Ω. For any β from the positive quadrant (consistent with H_0), $\Omega - \beta \subseteq \Omega$, with equality only at $\beta = 0$. Therefore

$$\max_{\beta_1,\beta_2\geq 0} Pr(LR > c) = Pr(LR > c)|_{\beta_1=0,\beta_2=0}$$

This corresponds to the 'worst case' null, i.e. the case where it is hardest to differentiate between the null and alternative hypotheses.

On the other hand, if $\beta_1 = \beta_2 = 0$, then $\hat{\beta}_{OLS}$ is distributed as $N(0, I_2)$, thus $\hat{\beta}^2_{1,OLS} \sim \chi^2(1)$ and $\hat{\beta}^2_{2,OLS} \sim \chi^2(1)$, and $\hat{\beta}^2_{1,OLS} + \hat{\beta}^2_{2,OLS} \sim \chi^2(2)$

Thus,

$$LR = \begin{cases} 0 & \text{with probability } 1/4 \\ \chi^2(1) & \text{with probability } 1/2 \\ \chi^2(2) & \text{with probability } 1/4 \end{cases}$$

Thus c is the 0.95 quantile of the mixture of chi-squared distributions. This can be found numerically.

$$F_{LR}(x) = 1/4F_0(x) + 1/2F_{\chi^2(1)}(x) + 1/4F_{\chi^2(2)}(x)$$