1 Estimation

Probability of e.g. $y_1 = 1, y_2 = 0, y_3 = 1$........is $F(\boldsymbol{\beta}'\mathbf{x}_i) * (1 - F(\boldsymbol{\beta}'\mathbf{x}_i)) * F(\boldsymbol{\beta}'\mathbf{x}_i)$......

In general we have:

$$L = \prod_{y_i=0} [1 - F(\boldsymbol{\beta}'\mathbf{x}_i)] \prod_{y_i=1} F(\boldsymbol{\beta}'\mathbf{x}_i)$$

and

$$\ln L = \sum_{i=1}^{n} \{y_i \ln F(\boldsymbol{\beta}'\mathbf{x}_i) + (1 - y_i)[1 - F(\boldsymbol{\beta}'\mathbf{x}_i)]\}$$

FOC:

$$\frac{\partial \ln L}{\partial \beta_k} = \sum_{i=1}^{n} \left[\frac{y_i f_i}{F_i} + (1 - y_i) \frac{-f_i}{(1 - F_i)} \right] x_{ik} = 0$$

For Logit we get:

$$\frac{\partial \ln L}{\partial \beta_k} = \sum_{i=1}^{n} (y_i - \Lambda_i) x_{ik} = 0$$

$$\mathbf{H} = \frac{\partial^2 \ln L}{\partial \beta_k \partial \beta_j} = \sum_{i=1}^{n} \Lambda_i (1 - \Lambda_i) x_{ik} x_{ij}$$

Impossible to derive an ML estimator on closed form. Use numerical optimization routines.

Corresponding conditions for the Probit model.

2 Inference

A general problem with both logit and probit is that the variance is not identified. For probit $\sigma^2 = 1$ and for logit it is set to $\frac{\pi}{3}$. What we get out is actually β_k/σ, β_{k+1}/σ .....

1. Marginal Effects:

$$\frac{\partial P(y=1)}{\partial \beta_k} = \frac{\partial F(\boldsymbol{\beta}'\mathbf{x}_i)}{\partial \beta_k} = f\left(\boldsymbol{\beta}'\mathbf{x}_i\right)\beta_k$$

   Thus it is a non-linear function of the estimated coefficients. We need to use the delta method to obtain standard errors. One problem is that both logit and probit are non-linear functions. Evaluation must be in a particular point. Most common alternative is the average of the $\mathbf{x}$ vector, i.e., $f\left(\boldsymbol{\beta}'\overline{\mathbf{x}}_i\right)\beta_k$. dprobit command in STATA does this. Sometimes this is supplemented with maximum and minimum values for the index function. An alternative is to average over all possible values of the index function, i.e., i.e., $\left[N^{-1}\sum_{i=1}^{N} f\left(\boldsymbol{\beta}'\mathbf{x}_i\right)\right]\beta_k$.

2. Another, previously very often used alternative, is to use a translation formula. If the index function is 0, then we get $\phi(0) = \frac{1}{\sqrt{2\pi}} \approx 0.4$ for the probit and $g(0) = \exp(0)/[1+\exp(0)]^2 = \frac{1}{4}$ for the logistic distribution and logit. So, the logit estimates should be 0.4/0.25 =1.6 times the probit estimates.

$$\widehat{\beta}_{LPM} = 0.25 * \widehat{\beta}_{Logit} \text{ and } \widehat{\beta}_{LPM} = 0.4 * \widehat{\beta}_{\Pr obit}$$

3. Ratios between coefficients has a direct interpretation, since

$$\frac{\beta_k/\sigma}{\beta_{k+1}/\sigma} = \frac{\beta_k}{\beta_{k+1}}.$$

Again, use the Delta-method, since it is a non-linear transformation of the coefficients, to obtain standard errors.

4. Another way to present result from these kind of models is to simulate the effect of hypothetical reforms. A very simple example is the change in probability due to the change in a discrete dependent variable, i.e.,

$$F\left(\beta_1 + \beta_2 x_2 + ... + \beta_{K-1} x_{K-1} + \beta_K\right) - F\left(\beta_1 + \beta_2 x_2 + ... + \beta_{K-1} x_{K-1}\right)$$

Test of exclusion restrictions:

We can, however, interpret the signs of the coefficients. LR test the most often used testprocedure for these models, i.e.,

$$\beta_k = \beta_{k+1} = ... = 0$$

$$LR = 2\left(\ln L_U - \ln L_R\right) \sim \chi^2_Q,$$

where Q is the number of restrictions.

Goodness-of-fit:

1. Most often used goodness-of-fit measure is McFadden's Pseudo R^2 :

   $$LRI = 1 - \ln L / \ln L_0.$$

   Attractive feature: it is always between 0 and 1, since $|\ln L| < |\ln L_0|$ .Unattractive feature: does not have the interpretation of "explained" sum of squares.

2. Alternative measure: $1 - SSR/SSR_0$, where SSR is residual sum of squares and $\widehat{u}_i = y_i - F\left(\widehat{\boldsymbol{\beta}}'\mathbf{x}_i\right)$ and SSR_0 the residual sum of squares when we onlu get the mean right.

3. Useful measure: predict y_i from $F\left(\widehat{\boldsymbol{\beta}}'\mathbf{x}_i\right)$ and recode all predictions greater than 0.5 as 1 and all smaller than 0.5 as zeros. Compare the actual data with the predictions.

2.1 Omitted Variable Bias

Remember that omitted variable bias in linear models: $E\left(\widehat{\boldsymbol{\beta}}_1\right) = \boldsymbol{\beta}_1 + \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_2\boldsymbol{\beta}_2$. More complicated in logit and probit models.

Example from Wooldridge

$$P\left(y=1 \mid \mathbf{x}, c\right) = \Phi\left(\mathbf{x}\beta + \gamma c\right)$$

c is not observed. Index function: $y^* = \mathbf{x}\beta + \gamma c + e$, $y = 1$ if $y^* > 0$.

$e \mid \mathbf{x}, c \sim N(0,1)$. (simplifying assumption).

$c \sim N\left(0, \tau^2\right)$.

$Cov\left(\mathbf{x}, c\right) = \mathbf{0}$.

This implies that $\gamma c + e \sim N\left(0, \gamma^2\tau^2 + 1\right)$ and

$$P\left(y=1 \mid \mathbf{x}, c\right) = P\left(\gamma c + e > -\mathbf{x}\boldsymbol{\beta} \mid \mathbf{x}\right) = \Phi\left(\mathbf{x}\beta/\sigma\right),$$

where $\sigma^2 \equiv \gamma^2\tau^2 + 1$.

Implications:

1. $p\lim\widehat{\beta}_j = \beta_j/\sigma$, since $\sigma = \sqrt{\gamma^2\tau^2+1} > 1$, $\left|\beta_j/\sigma\right| < \left|\beta_j\right|$. Attenuation bias. But we are rarely interested in the level of β_j/σ.

2. Marginal effect: $\partial P\left(y=1 \mid \mathbf{x}, c\right)/\partial x_j = \beta_j\phi\left(\mathbf{x}\beta + \gamma c\right)$. Probit gives a consisten estimate of $\beta_j\phi\left(\mathbf{x}\beta\right)$. The estimate of the partial effect at $c = 0$ in general not consistent. β_j/σ is closer to zero and $\phi\left(\mathbf{x}\beta/\sigma\right)$ is larger than $\phi\left(\mathbf{x}\beta\right)$. Counteracting effects!!

Average partial effects (APE)

If we want the average partial effect instead of the partial effect evaluated at $c = 0$, which is uninteresting since $P(c = 0) = 0$. Plug in a certain value of the $\mathbf{x}$ matrix, say $\mathbf{x}^0$.

$$E\left[\partial P\left(y = 1 \mid \mathbf{x}^0, c\right)/\partial x_j\right] = E\left[\beta_j \phi\left(\mathbf{x}^0 \boldsymbol{\beta} + \gamma c\right)\right] = \left(\beta_j/\sigma\right)\phi\left(\mathbf{x}^0\boldsymbol{\beta}/\sigma\right).$$

This follows from the law of iterated expectations. $E_c\left[\phi\left(\mathbf{x}\beta + \gamma c\right)\right] = \phi\left(\mathbf{x}\beta/\sigma\right)$ and $\partial\phi\left(\mathbf{x}\beta/\sigma\right)/\partial x_j = \left(\beta_j/\sigma\right)\phi\left(\mathbf{x}\beta/\sigma\right)$.

This parameter is what we most often are interested in when we talk about "marginal effects".

General result from Yatchew and Griliches (1984):

1. $p\lim \widehat{\boldsymbol{\beta}}_1 = c_1\boldsymbol{\beta}_1 + c_2\boldsymbol{\beta}_2$, c_1 and c_2 are complicated functions of $\boldsymbol{\beta}_2$ the unknown parameters.
2. Heteroscedasticity will be introduced.

2.2 Heteroscedasticity

Example in Wooldridge:

$$
\begin{aligned}
y^* &= \beta_0 + \beta_1 x_1 + e \\
e &\mid x_1 \sim N\left(0, x_1^2\right)
\end{aligned}
$$

Suppose we use a probit model, this implies

$$P\left(y = 1 \mid x_1\right) = \Phi\left(\beta_0 / x_1 + \beta_1\right)$$

and the marginal effect is

$$\frac{\partial P\left(y = 1 \mid x_1\right)}{\partial x_1} = -\left(\beta_0 / x_1^2\right) \phi\left(\beta_0 / x_1 + \beta_1\right).$$

If $\beta_0{>}0$ and $\beta_1{>}0$ then $\frac{\partial P(y=1|x_1)}{\partial x_1}$ will be negative!!
Probit on 1 and $1/x_1$ will give consistent parameter estimates

Wooldridge's point: the problem is not that probit gives inconsisten estimates, but that $F(\boldsymbol{\beta}'\mathbf{x}_i) \neq \Phi(\boldsymbol{\beta}'\mathbf{x}_i)$.

Testing for heteroscedasticity:

Problem since σ^2 is not identified and ML will give inconsistent estimates and the covariance matrix will be inappropriate Suppose we have

$$P\left(y_i = 1 \mid \mathbf{x}_i\right) = F\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma g\left(\mathbf{x}_i\right)}\right)$$

or

$$P\left(y_i = 1 \mid \mathbf{x}_i\right) = F\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma \boldsymbol{\gamma}'\mathbf{x}_i}\right).$$

It is then hard to identify if the effect works through the mean or the variance.

Suppose

$$\begin{aligned} y_i^* &= \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i \\ Var\left(\varepsilon_i\right) &= \left(e^{\boldsymbol{\gamma}'\mathbf{z}_i}\right)^2 \\ \ln L &= \sum^n \left\{ y_i \ln F\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\exp\left(\boldsymbol{\gamma}'\mathbf{z}_i\right)}\right) + \left(1 - y_i\right) \ln F\left(\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\exp\left(\boldsymbol{\gamma}'\mathbf{z}_i\right)}\right)\right\} \end{aligned}$$

$\mathbf{z}_i$ cannot include a constant. A feasible test is to use a LM test and estimate this model under $H_0 : \boldsymbol{\gamma} = \mathbf{0}$. This is what is used in most statistical packages, e.g. STATA. $\mathbf{z}_i$ has to be specified, often polynomials of the $\mathbf{x}_i$ vector.

3 Non- and Semi-parametric Methods

3.1 MSCORE

$$Max_{\boldsymbol{\beta}} S_{N_\alpha}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\left[z_i - (1-2\alpha)\right] sign\left(\boldsymbol{\beta}'\mathbf{x}_i\right),$$

where α is a parameter set by the researcher and $z_i = 2y_i - 1$ ($z_i = 1$ if $y_i = 1$ and $z_i = -1$ if $y_i = 0$). If $\alpha = .5$, then

$$Max_{\boldsymbol{\beta}} S_{N.5}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} z_i * sign\left(\boldsymbol{\beta}'\mathbf{x}_i\right).$$

Choose $\boldsymbol{\beta}$ so as to maximize the number of right predictions under the constraint that $\boldsymbol{\beta}$ ' $\boldsymbol{\beta} = 1$.

No likelihood function underlying this. Standard errors has to be obtained by bootstrapping.

Good for prediction. Disadvantage that it is not possible to obtain meaningful marginal effects.

3.2 Kernel estimation of $F(\cdot)$

Idea is to estimate $F(\boldsymbol{\beta}'\mathbf{x}_i)$ by a general function $G(\boldsymbol{\beta}'\mathbf{x}_i)$.

Usual likelihood function:

$$\log L_n = \frac{1}{n}\sum_{i=1}^{n}\left\{y_i \log G_n(\boldsymbol{\beta}'\mathbf{x}_i) + (1-y_i)\log[1-G_n(\boldsymbol{\beta}'\mathbf{x}_i)]\right\},$$

where the distribution function $G_n(\cdot)$ is estimated in each iteration.

The density function for $\boldsymbol{\beta}'\mathbf{x}_i$ conditional on the $y_i = 1$ is calculated as

$$g_n(z \mid y=1) = \frac{1}{n\overline{y}h_n}\sum_{i=1}^{n} y_i K\left(\frac{z-\boldsymbol{\beta}'\mathbf{x}_i}{h_n}\right),$$

and conditional on $y_i = 0$ as

$$g_n(z \mid y=0) = \frac{1}{n(1-\overline{y})h_n}\sum_{i=1}^{n}(1-y_i)K\left(\frac{z-\boldsymbol{\beta}'\mathbf{x}_i}{h_n}\right),$$

where h_n is the bandwith and $K(\cdot)$ is the kernel function. Finally the distribution function is calculated as

$$G_n(z) = \frac{\overline{y}g_n(z \mid y=1)}{\overline{y}g_n(z \mid y=1) + (1-\overline{y})g_n(z \mid y=0)}.$$

1. Assume ε follows a uniform distribution $\rightarrow$ LPM.

2. Assume ε follows a logistic distribution $F(z) = \frac{1}{1+\exp(-z)}$ $\rightarrow$ Logit model.

3. Assume ε follows a normal distribution $\rightarrow$ Probit model.

4. Assume ε follows a general probability distribution $\rightarrow$ Kernel estimation.

5. Only specify that $y_i = 1$ if $\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon > 0$ and $y_i = 0$, where ε is a random variable with $median = 0$ $\rightarrow$ MSCORE.