Maximum Likelihood Estimation of Binary Response Models

By far the most common way to estimate binary response models is to use the method of maximum likelihood. Because the dependent variable is discrete, the likelihood function cannot be defined as a joint density function, as it was in Chapter 10 for models with a continuously distributed dependent variable. When the dependent variable can take on discrete values, the likelihood function for those values should be defined as the probability that the value is realized, rather than as the probability density at that value. With this redefinition, the *sum* of the possible values of the likelihood is equal to 1, just as the *integral* of the possible values of a likelihood based on a continuous distribution is equal to 1.

If, for observation t, the realized value of the dependent variable is y_t, then the likelihood for that observation if $y_t = 1$ is just the probability that $y_t = 1$, and if $y_t = 0$, it is the probability that $y_t = 0$. The logarithm of the appropriate probability is then the contribution to the loglikelihood made by observation t.

Since the probability that $y_t = 1$ is $F(\boldsymbol{X}_t\boldsymbol{\beta})$, the contribution to the loglikelihood function for observation t when $y_t = 1$ is $\log F(\boldsymbol{X}_t\boldsymbol{\beta})$. Similarly, the contribution to the loglikelihood function for observation t when $y_t = 0$ is $\log\bigl(1 - F(\boldsymbol{X}_t\boldsymbol{\beta})\bigr)$. Therefore, if $\boldsymbol{y}$ is an n–vector with typical element y_t, the loglikelihood function for $\boldsymbol{y}$ can be written as

$$\ell(\boldsymbol{y}, \boldsymbol{\beta}) = \sum_{t=1}^{n} \Bigl(y_t \log F(\boldsymbol{X}_t\boldsymbol{\beta}) + (1 - y_t) \log\bigl(1 - F(\boldsymbol{X}_t\boldsymbol{\beta})\bigr)\Bigr). \tag{11.09}$$

For each observation, one of the terms inside the large parentheses is always 0, and the other is always negative. The first term is 0 whenever $y_t = 0$, and the second term is 0 whenever $y_t = 1$. When either term is nonzero, it must be negative, because it is equal to the logarithm of a probability, and this probability must be less than 1 whenever $\boldsymbol{X}_t\boldsymbol{\beta}$ is finite. For the model to fit perfectly, $F(\boldsymbol{X}_t\boldsymbol{\beta})$ would have to equal 1 when $y_t = 1$ and 0 when $y_t = 0$, and the entire expression inside the parentheses would then equal 0. This could happen only if $\boldsymbol{X}_t\boldsymbol{\beta} = \infty$ whenever $y_t = 1$, and $\boldsymbol{X}_t\boldsymbol{\beta} = -\infty$ whenever $y_t = 0$. Therefore, we see that (11.09) is bounded above by 0.

Maximizing the loglikelihood function (11.09) is quite easy to do. For the logit and probit models, this function is globally concave with respect to $\boldsymbol{\beta}$ (see Pratt, 1981, and Exercise 11.1). This implies that the first-order conditions, or likelihood equations, uniquely define the ML estimator $\hat{\boldsymbol{\beta}}$, except for one special case that we consider in the subsection following the next one. These likelihood equations can be written as

$$\sum_{t=1}^{n} \frac{\bigl(y_t - F(\boldsymbol{X}_t\boldsymbol{\beta})\bigr) f(\boldsymbol{X}_t\boldsymbol{\beta}) x_{ti}}{F(\boldsymbol{X}_t\boldsymbol{\beta})\bigl(1 - F(\boldsymbol{X}_t\boldsymbol{\beta})\bigr)} = 0, \;\; i = 1, \ldots, k. \tag{11.10}$$

There are many ways to find $\hat{\boldsymbol{\beta}}$ in practice. Because of the global concavity