Choice with multiple alternatives – 5.4 Maximum likelihood estimation

Michel Bierlaire

The maximum likelihood estimation method is exactly the same for logit with multiple alternatives, as for the binary logit model. The logit model is

$$P_n(i|\mathcal{C}_n) = \frac{e^{V_{in}}}{\sum_{j\in\mathcal{C}_n} e^{V_{jn}}}. \tag{1}$$

The log likelihood of a sample is

$$\mathcal{L}(\beta_1,\ldots,\beta_K) = \sum_{n=1}^{N}\left(\sum_{i\in\mathcal{C}_n} y_{in}\ln P_n(i|\mathcal{C}_n)\right), \tag{2}$$

where $y_{in} = 1$ if individual n has chosen alternative i, 0 otherwise. Using (1) into (2), we obtain

$$\mathcal{L}(\beta_1,\ldots,\beta_K) = \sum_{n=1}^{N}\sum_{i\in\mathcal{C}_n} y_{in}\left(V_{in} - \ln\sum_{j\in\mathcal{C}_n} e^{V_{jn}}\right). \tag{3}$$

The maximum likelihood estimation amounts to find the vector β solving the optimization problem

$$\max_{\beta\in\mathbb{R}^K} \mathcal{L}(\beta). \tag{4}$$

In the case of logit, it can be shown (McFadden, 1974) that, if the utility function is linear in the parameters, the log likelihood function is globally concave and does not exhibit local maxima (under some relatively weak conditions).

The necessary first-order optimality conditions impose that the partial derivatives with respect to each parameter is equal to zero. The kth partial derivative is

$$\frac{\partial \mathcal{L}}{\partial \beta_k} = \sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} y_{in} \left(\frac{\partial V_{in}}{\partial \beta_k} - \sum_{j \in \mathcal{C}_n} P_n(j) \frac{\partial V_{jn}}{\partial \beta_k} \right) \text{ for } k = 1, \ldots, K. \tag{5}$$

Distributing the y_{in}, we obtain

$$\frac{\partial \mathcal{L}}{\partial \beta_k} = \sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} (y_{in} - P_n(i)) \frac{\partial V_{in}}{\partial \beta_k} \text{ for } k = 1, \ldots, K. \tag{6}$$

For a linear-in-parameters logit, it is

$$\sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} (y_{in} - P_n(i)) x_{ink}, \text{ for } k = 1, \ldots, K. \tag{7}$$

Setting these equations to zero leads to the necessary first-order optimality conditions:

$$\sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} y_{in} x_{ink} = \sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} P_n(i) x_{ink}. \tag{8}$$

It means that the expected value of each attribute of the chosen alternative must be the same when computed in the sample or with the choice model.

The reader can also verify that the second derivatives of $\mathcal{L}$ are given by

$$\frac{\partial^2 \mathcal{L}}{\beta_k \beta_\ell} = - \sum_{n=1}^{N} \sum_{i \in \mathcal{C}_n} P_n(i) \left(x_{ink} - \sum_{j \in \mathcal{C}_n} x_{jnk} P_n(j) \right) \left(x_{in\ell} - \sum_{j \in \mathcal{C}_n} x_{jn\ell} P_n(j) \right). \tag{9}$$

References

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior, *in* P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, pp. 105–142.