KM 32(1989)
pag 75 - 81

A Note on Logit and Probit Analysis in the Case of Separability

Piet Rietveld*

Cees Gorter*

Abstract

It is shown that when groups of observations are separable, one runs into trouble with logit and probit estimates. Some implications are discussed for the case of near-separability. Discriminant analysis is more robust to the problem of separability. It is indicated that methods from the sphere of multiple objective linear programming can be used to generate the set of parameters implied by separability.

---

* Free University
Faculty of Economics
P.O. Box 7161
1007 MC Amsterdam
Tel. 020-548 4932

1. Introduction.

Logit and probit analysis are standard tools for analyzing the influence of independent variables on a qualitative variable (cf. Maddala, 1983). For example, in the case of a referendum one group of the electorate votes 'no' ($y_i=0$), and another group votes 'yes' ($y_i=1$). Then logit and probit analysis can be used to study the role of features x of voters as determinants of their voting behaviour.

This note will be devoted to discussing some problems which will be met when the two groups are separable. We will start with a definition of separability.

Consider two groups of individuals: $S_0 = \{i|y_i=0\}$ and $S_1 = \{i|y_i=1\}$ Individuals i are characterized by a vector x_i containing J elements. Consider a linear function

$$(1) \quad Z_i = \alpha + \beta x_i$$

where β is a J dimensional vector.

Then, the groups S_0 and S_1 are called separable when coefficients $(\hat{\alpha}, \hat{\beta})$ exist for which:

$$(2) \quad \begin{aligned} \hat{Z}_i &= \hat{\alpha} + \hat{\beta}' x_i \leq 0 && \text{for all } i \epsilon S_0 \\ \hat{Z}_i &= \hat{\alpha} + \hat{\beta}' x_i > 0 && \text{for all } i \epsilon S_1 \end{aligned}$$

How relevant is the case of separability in the context of logit or probit analysis? Separability can easily be established when groups are separable on the basis of just one characteristic. In this case it is not meaningful at all to use logit or probit analysis. This does no longer hold true when separability can only be achieved on the basis of at least two characteristics. In this case, the existence of separability can usually not be established without the use of strong analytical tools. Especially when samples are relatively small, it is not impossible that separability occurs (for an example, see Rietveld and Gorter, 1988). As we will show in the next section, logit and probit analysis become problematic in the case of separability.

2. Logit and Probit in the Case of Separability.

The logit and probit model can be formulated as follows. Let y_i^* be a latent variable, defined by the relationship:

$$(3) \quad y_i^* = \alpha + \beta' x_i + u_i$$

where u_i is an error term. Further, y_i is an observed dummy variable related to y_i^* in the following way:

$$(4) \quad \begin{aligned} y_i &= 0 \qquad \text{if } y_i^* \leq 0 \\ y_i &= 1 \qquad \text{if } y_i^* > 0 \end{aligned}$$

Then, the probability that $y_i = 1$ is equal to:

$$(5) \quad \begin{aligned} \Pr\,[y_i = 1] &= \Pr\,[\alpha + \beta' x_i + u_i > 0] \\ &= \Pr\,[u_i > -\alpha - \beta' x_i] \\ &= 1 - F\,(-\alpha - \beta' x_i) \end{aligned}$$

where F is the cumulative distribution of u. The corresponding likelihood function is:

$$(6) \quad L(\alpha,\beta) = \prod_{y_i=0} F\,(-\alpha-\beta' x_i) \cdot \prod_{y_i=1} [1-F(-\alpha-\beta' x_i)]$$

Functional forms often used for F are the logistic or the cumulative normal distribution. In these cases, F satisfies the property:

$$(7) \quad \frac{dF(Z_i)}{dZ_i} > 0 \quad , \forall Z_i$$

Consider now the case of separability. Let $\hat{\alpha}$ and $\hat{\beta}$ be parameters for which (2) holds true. Then it is not difficult to see by means of (7) that when a>1:

$$(8) \quad \begin{aligned} F(-\hat{\alpha}-\hat{\beta}' x_i) &< F(-a\hat{\alpha}-a\hat{\beta}' x_i) \text{ for all } i \in S_0 \\ 1-F(-\hat{\alpha}-\hat{\beta}' x_i) &< 1-F(-a\hat{\alpha}-a\hat{\beta}' x_i) \text{ for all } i \in S_1 \end{aligned}$$

For the likelihood function this implies:

$$(9) \quad L(a\hat{\alpha}, a\hat{\beta}) > L(\hat{\alpha}, \hat{\beta}) \qquad (a>1)$$

Consequently, if S_0 and S_1 are separable, the likelihood function $L(\alpha,\beta)$ does not attain its maximum value (1) for finite values of α and β. Thus, in the case of separability, no maximum likelihood estimates will be found for the coefficients via logit or probit analysis.

As a solution to this problem, one may fix one of the coeffcients (for example, α=1) after which maximum likelihood estimates for β can be found in the ordinary way. One must be aware, however, that the estimates of β depend on the value chosen for α. More specifically, there is no proportionality between α and β in logit or probit analysis: the ratios of the coefficients depend on the value chosen for α. This makes the results of such a procedure somewhat arbitrary.

3. Discussion.

Several possibilities exist to overcome the problem of separability: adding observations, deleting independent variables, or changing the functional form of the function Z in (1).

When a sample of observations is drawn from a population, one can extend the number of observations, with the possibility that a new element of S_0 is found which does not satisfy (2), so that separability does no longer occur.

Deleting independent variables may make separable groups inseparable, as can be seen for example in Figure 1, where the groups S_0 and S_1 are inseparable according to either x_1 or x_2, but not according to x_1 and x_2 jointly. This indicates that a certain parallel exists with the problem of multicollinearity in multiple regression. Separability of groups means that it is not possible to determine the relative influence of one independent variable compared with another. Just like in the case of multicollinearity, this problem may be overcome by deleting variables. The price which must be paid for this is quite high, however: specification errors will occur. We note in passing that adding variables will never remove separability. By assigning zeros to the coefficients of the added variables, separability will persist.

Changing the functional form of Z may also remove the separability property. For example, if in Figure 1, Leontief-type relationships would be used instead of linear forms, separability would no longer occur.

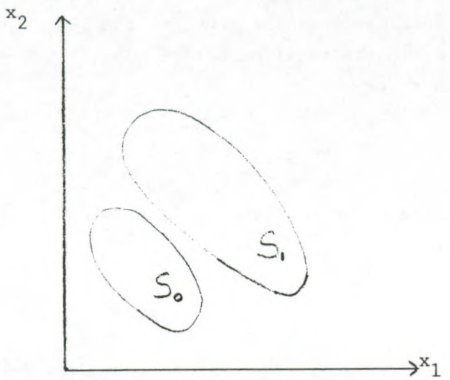

Figure 1. Separability in the case of 2 independent variables.

A small change in the value of a certain independent variable for observation i may make a separable problem inseparable. In the separability case it is impossible to obtain point estimates for the coefficients, whereas with inseparable problems the coefficients can be found without any problem. This implies that logit and probit analysis may be extremely sensitive to <u>errors in variables</u>. When independent variables are subject to measurement errors one may easily end up with inseparable groups which are separable according to the true data. Thus, in the case that logit or probit analysis only lead to a very small number of misclassifications, a special screening of data quality for these observations is recommended. Estimation results may be very sensitive for changes in these observations.

4. <u>Other Approaches in the Case of Separability</u>.

Another approach to qualitative dependent variables is discriminant analysis. The aim of linear discriminant analysis is to find a linear relation $V = \lambda' x$ that provides the best discrimination between the groups S_0 and S_1. The vector λ is chosen so that the ratio of the variance of $\lambda'x$ between groups and the variance of $\lambda'x$ within groups is as high as possible. The between group variance is related to the mean observations in both groups: $\bar{x}_o$ and $\bar{x}_1$. Then the between group variance is $[\lambda'(\bar{x}_o-\bar{x}_1)]^2$. Further, the within group variance in $\lambda'x$ is $\lambda'\Sigma\lambda$ where Σ is the weighted average of the covariance matrices for groups 0 and 1, the weights being the shares of S_0 and S_1 in the total number of observations. Thus, λ must be determined such that

$$(10) \qquad b = [\lambda'(\bar{x}_o-\bar{x}_1)]^2/\lambda'\Sigma\lambda$$

is maximized (cf. Maddala, 1983). It can be shown that the optimal solution for λ is:

$$(11) \qquad \hat{\lambda} = a\,\Sigma^{-1}\,(\bar{x}_0-\bar{x}_1)$$

where a is an arbitrary constant. Thus, a vector λ can be determined when Σ is a non-singular matrix. A closer inspection reveals that separability of groups S_0 and S_1 does not imply singularity of Σ (neither does singularity of Σ imply separability). Thus, we conclude that when logit analysis is plagued by the problem of separability, discriminant analysis can in general be carried out without any difficulty.

There is still another approach to the case of separability which is worth mentioning here. Consider the set T of all values of the coefficients α and β which satisfy (2). Thus:

$$(12) \qquad T = \{\alpha,\beta \mid \alpha+\beta'x_i \le 0 \text{ for all } i \in S_0, \text{ and } \alpha+\beta'x_i > 0 \text{ for all } i \in S_1\}$$

When S_0 and S_1 are separable, T is non-empty. However, when T is non-empty, it is also unbounded, since if $(\hat{\alpha},\hat{\beta})$ is an element of T, any vector $(a\hat{\alpha}, a\hat{\beta})$ is an element of T as long as $a > 0$. This difficulty can be solved by fixing one of the parameters, for example: $\alpha = \hat{\alpha}$.

Then the problem becomes the finding of the following set of parameters β:

$$(13) \qquad T(\beta|\hat{\alpha}) = \{\beta \mid \beta'x_i \le -\hat{\alpha} \text{ for all } i \in S_0, \text{ and } \beta'x_i > -\hat{\alpha} \text{ for all } i \in S_1\}$$

The set $T(\beta|\hat{\alpha})$ is a convex polyhedron, which can be described in terms of a finite number of nodes and faces. A method to generate these nodes and faces is described by Steuer (1986) in the context of multiple objective linear programming (MOLP). A general assumption in MOLP is that the β vector is non-negative. This restriction can easily be overcome by substituting $(\gamma-\delta)$ for β in $T(\beta|\hat{\alpha})$ and adding the restrictions $\gamma\geq 0$ and $\delta\geq 0$. We conclude that MOLP provides an operational tool for determining the whole set of coefficients β which lead to an exact separation of the two sets.

It is of special interest to see which characteristics are critical in determining separability. A variable can be defined as critical when a separable problem becomes inseparable by dropping this variable. Thus, when characteristic j is critical, the parameter β_j does not assume the value zero in $T(\beta|\hat{\alpha})$:

$$(14) \qquad T(\beta|\hat{\alpha}) \cap S(\beta|\beta_j=0) \quad = 0$$

Standard methods of linear programming can be used to find the set of critical characteristics.

References

Maddala, G.S., Limited-dependent and Qualitative Variables in Econometrics, Cambridge University Press, Cambridge, 1983.

Rietveld, P. and C. Gorter, Stimulating Innovation in Indonesian Small Scale Industry; Are Skill Upgrading Programs Effective?, Dept. of Economics, Free University, Amsterdam, 1988.

Steuer, R.E., Multiple Criteria Optimization, Wiley, New York, 1986.

Ontvangen: 09-02-1989
Geaccepteerd: 21-09-1989