Volume Title: Annals of Economic and Social Measurement, Volume 5, number 4

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm76-4

Publication Date: October 1976

Chapter Title: A Comment on Discriminant Analysis "Versus" Logit Analysis

Chapter Author: Daniel McFadden

Chapter URL: http://www.nber.org/chapters/c10493

Chapter pages in book: (p. 511 - 523)

*Annals of Economic and Social Measurement*, 5/4, 1976

# A COMMENT ON DISCRIMINANT ANALYSIS "VERSUS" LOGIT ANALYSIS[1]

BY DANIEL MCFADDEN

*This note contrasts discriminant analysis with logit analysis. In causal models, it is seen that forecasting leads to classification problems based on selection probabilities. The posterior distributions implied by the selection probabilities and prior distribution may provide a useful starting point for estimation of the selection probability parameters in a discriminant-type analysis, but this procedure does not tend to be robust with respect to misspecification of the prior. In conjoint models, on the other hand, the posterior distributions and selection probabilities are alternative conditional distributions characterizing the joint distribution. In these models, it is generally not meaningful to examine the effects of shifts in explanatory variables.*

## I. INTRODUCTION

Consider an experiment in which individual characteristics, attributes of possible responses, and actual responses are observed for a sample of subjects. Suppose the sets of possible responses are finite, so the problem is one of quantal response.

One approach to the analysis of such data is the *logit model*, which postulates that the actual responses are drawings from multinomial distributions with selection probabilities conditioned on the observed values of individual characteristics and attributes of alternatives, with the logistic functional form. A second approach is *discriminant analysis*, which postulates that the observed values of individual characteristics and attributes of alternatives are drawings from posterior distributions conditioned on actual responses.

When the posterior distributions in discriminant analysis are taken to be multivariate normal with a common covariance matrix, one obtains the implication that the relative odds that a given vector of observations is drawn from one posterior distribution or the other are given by a logistic formula.[2] This seems to have led to some confusion as to whether these two approaches provide equally satisfactory interpretations of the logit model, and whether the statistical estimators and applications which seem natural for one of the models have some reasonable interpretation in the other model. In this comment, I will write down a common probability model for the two approaches, and use it to clarify these issues.

## II. OBSERVED VARIABLES

Consider a typical quantal response experiment, for example a study of travel mode choice. The possible responses of a subject in a particular experimental setting are indexed by a finite set $B = \{1, \dots, J\}$. With each response $j \in B$ is associated a vector $z_j$ of observed variables and vector $\xi_j$ of unobserved variables. We define $z = (z_1, \dots, z_J)$ and $\xi = (\xi_1, \dots, \xi_J)$.

[1] This research is supported by NSF Grant No. GS-35890X. The question addressed in this comment was raised during the NSF-NBER Conference on Individual Decision Rules, University of California, Berkeley, March 22–23, 1974. I benefited from discussions at that time with R. Hall, J. Heckman, J. Houseman, J. Press, and R. Westin. I retain sole responsibility for errors.

[2] A discussion of the discriminant model and of this and related properties has been given by Ladd (1966).

Some discussion is required on the interpretation of the response index $j$ and the data vector $z_j$. In applications such as mode choice, it is usually natural to associate a particular index with a particular response; e.g., $j = 1$ may be the "walk" mode. In other applications such as destination choice, there will be no natural indexing, so that the index $j$ associated with a particular response is arbitrary. The data vector $z_j$ can be interpreted as a transformation of observations $x_i^0$ on the attributes of each alternative $i$ and $s^0$ on the characteristics of the subject; i.e.,

$$z_j = Z(x_j^0; x_1^0, \ldots, x_{j-1}^0, x_{j+1}^0, \ldots, x_J^0; s^0), \tag{1}$$

where $Z$ is a vector of known functions. Note that the components of $z_j$ may be components of observed attributes of alternatives or characteristics of individuals, or may be interaction terms involving products or more complex functions of these variables. In the case that there is a natural indexing of responses, we can include the index $j$ as a component of the vector $x_j^0$; this allows the inclusion of components of $z_j$ which are interactions between components of the $x_j^0$ or of $s^0$ and a dummy variable for a particular index $i$; i.e.,

$$z_{jl} = \begin{cases} x_{im}^0 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} \quad \text{or} \quad z_{jl} = \begin{cases} s_m^0 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases}. \tag{2}$$

On the other hand, when there is no natural indexing, variables such as those in Equation (2) are not meaningful. It is for this reason that the function $Z$ in Equation (1) is assumed to depend on the response index $j$ only via its effect on $x_j^0$. We note further in this equation that in most applications, $z_j$ will depend solely on $x_j^0$ and $s^0$. More generally, dependence across alternatives is possible. However, in keeping with the stipulation above that $z_j$ depends on the index $j$ only if the index itself is an attribute of the alternative, we require that $Z$ be invariant with respect to the order of the sub-vectors $x_1^0, \ldots, x_{j-1}^0, x_{j+1}^0, \ldots, x_J^0$. Analogously to the interpretation of the observed variables $z$, we can interpret the unobserved variables $\xi$ as coming from unobserved attributes of alternatives $x_j^u$ and unobserved individual characteristics $s^u$.

## III. Selection Probabilities

Provided we take a sufficiently general definition of the unobserved variables $\xi$, the subject's actual response is completely determined by the alternative set $B$ and the observed and unobserved variables $(z, \xi)$; let

$$j = D(B, z, \xi) \tag{3}$$

denote this relationship, and define

$$E_j(B, z) = \{\xi \mid D(B, z, \xi) = j\} \tag{4}$$

to be the set of unobserved vectors giving response $j$.

We now assume the variables $z$, $\xi$ are jointly distributed with a frequency function $f(z, \xi)$. In general, we can allow some components of $(z, \xi)$ to be continuous and others to be discrete, taking the corresponding components of the product measure $(\nu, \eta)$ on $(z, \xi)$ to be Lebesgue or counting measure, respectively. We can also allow $f$ to be degenerate, and restrict our attention to a suitable manifold. For example, the case where some components of $z$ involve interactions of variables with alternative dummies will correspond to a degenerate $f$ distribution.

We first define the selection probability that response $j$ occurs, conditioned on the response set $B$ and observed data $z$. Let

$$g(z) \equiv g(z; B) = \int f(z, \xi)\eta(d\xi) \tag{5}$$

be the marginal frequency for $z$. Then the selection probability is given by the conditional probability formula

$$p_j(B, z) = \int_{E_j(B,z)} f(z, \xi)\eta(d\xi)/g(z). \tag{6}$$

We note that the expression

$$h(j, z; B) \equiv p_j(B, z)g(z) = \int_{E_j(B,z)} f(z, \xi)\eta(d\xi)$$

is the *joint* distribution of $(j, z)$ conditioned on $B$. Equation (6) is meaningful whether or not there is a natural indexing of alternatives. This implies in particular that models formulated and analyzed solely in terms of the selection probabilities do not require natural indexing. However, the concepts to be introduced next require natural indexing in order to be meaningful.

## IV. Classification Models

Assume hereafter that there is a natural indexing $j$ of alternatives. Define *mean* selection probabilities

$$\begin{aligned} P_j \equiv P_j(B) &= \int p_j(B, z)g(z)\nu(dz) \\ &= \int \left\{ \int_{E_j(B,z)} f(z, \xi)\eta(d\xi) \right\} \nu(dz). \end{aligned} \tag{7}$$

Next, define the posterior distribution of the observed variables given the actual response $j$. This frequency is clearly proportional to the probability of actual response $j$ conditioned on the observed data, multiplied by the marginal frequency function for the observed data, or

$$q_j(B, z) = p_j(B, z)g(z)/P_j = h(j, z, B)/P_j \tag{8}$$

with the normalizing constant obtained from Equation (7). An obvious implication of this equation is that any specification of the selection probabilities $p_j$ and frequency function $g$ of the observations determines specific posterior distributions $q_j$. In this sense, every model for the selection probabilities combined with a "prior" distribution $g$ on the explanatory variables yields a classification model to which some sort of discrimination analysis could be applied. However, the case of multinomial logit selection probabilities and a multivariate normal prior will *not* yield multivariate normal posterior distributions. (In the binary response case, the posterior distributions are transformations of the $s_B$ distribution; see Johnson (1949) and Westin (1974).)

## V. Consistency of Selection Probabilities and Posterior Distributions

We next consider the question of whether particular parametric specifications for the selection probabilities and posterior distributions are consistent, or equivalently whether there exists a prior distribution $g$ satisfying

$$g(z) = q_j(B, z)P_j/p_j(B, z) \tag{9}$$

for all $j$. (In this construction, the $P_j$ can be treated as constants to be determined.) It is obvious that (9) need not have a solution; clearly, $q_j(B, z)/p_j(B, z)$ must be integrable, and $q_j$ must equal $p_j$ except for a multiplicative constant depending on $j$ and a multiplicative function independent of $j$.[3]

Suppose the selection probabilities are specified to be multinomial logit,

$$p_j(B, z) = \frac{e^{\gamma_j + \beta' z_j}}{\sum_{i\in B} e^{\gamma_i + \beta' z_i}}, \tag{10}$$

where $\beta, \gamma_1, \ldots, \gamma_J$ are parameters and we impose the normalization $\gamma_1 + \ldots + \gamma_J = 0$. Note that when the $z_i$ variables are of the form in Equation (2), Equation (10) specializes to

$$p_j(B, z) = \frac{e^{\gamma_j + \beta'_{(j)} z_{(j)}}}{\sum_{i\in B} e^{\gamma_i + \beta'_{(i)} z_{(i)}}}, \tag{11}$$

where the $\beta_{(j)}$ and $z_{(j)}$ are subvectors of $\beta$ and $z$. An important special case of Equation (11) occurs when the variables $z_{(j)}$ are the same for each alternative,

$$p_j(B, z) = \frac{e^{\gamma_j + \beta'_{(j)} z_{(1)}}}{\sum_{i\in B} e^{\gamma_i + \beta'_{(i)} z_{(1)}}} \tag{12}$$

and the normalization $\sum_{i\in B} \beta'_{(i)} = 0$ is imposed. This formulation is common when attributes of alternatives are absent and only characteristics of subjects are observed. However, note that $z_{(1)}$ may contain attributes of all alternatives, making Equation (12) as general as Equation (10).

Next suppose the posterior distributions $q_j$ to be multivariate normal with a common covariance matrix. In order to include the possibility that $g$ is degenerate, we assume (by a translation of the origin if necessary) that $z$ varies in a subspace $L$. Then, $q_j$ has a mean $\mu^j \in L$ and a covariance matrix $\Omega$ that is positive

[3] A question with a trivial affirmative answer is whether, given posterior distributions $q_j$ and mean selection probabilities $P_j$, one can find a prior distribution $g$ and selection probabilities $p_j$ such that Equation (9) holds. From Equation (9) define $p_j = P_j q_j / g$. Then $\sum_j p_j = 1$, $g = \sum_j P_j q_j$. Then, a prior $g$ which is a $P_j$ probability mixture of the posterior distributions is necessary and sufficient to give a solution. Compare this result with the analysis following where $p_j$ is restricted.

semi-definite and definite with respect to the subspace $L$.[4] The frequency functions can then be written (suppressing $B$)

$$q_j(B, z) \equiv q_j(z) = K \exp\left[-\tfrac{1}{2}(z-\mu^j)'\Lambda(z-\mu^j)\right], (z \in L) \tag{13}$$

where $K$ is a constant independent of $j$ and $\Lambda$ is the generalized inverse of $\Omega$.[5] Define a vector $\beta^j = (0, \ldots, \beta, \ldots, 0)$ commensurate with $z = (z_1, \ldots, z_J)$ and with the $j$-th subvector equal to $\beta$.

*Theorem* 1. Suppose the selection probabilities satisfy Equation (10) and the posterior distributions satisfy Equation (13). Then the conditions for consistency are that the prior distribution be a probability mixture of the posterior distributions.

$$g(z) = \sum_{i \in B} P_i q_i(z), \tag{14}$$

with the means $\mu^j$ in Equation (13) satisfying

$$\mu^j = \Omega(\beta^j + \delta), \tag{15}$$

$\delta$ an arbitrary vector, and with

$$\begin{aligned} P_j &= \exp\left[\gamma_j + \tfrac{1}{2}\mu^{j\prime}\Lambda\mu^j\right] \Big/ \sum_{i \in B} \exp\left[\gamma_i + \tfrac{1}{2}\mu^{i\prime}\Lambda\mu^i\right] \\ &= \exp\left[\gamma_j + \tfrac{1}{2}(\beta^j + \delta)'\Omega(\beta^j + \delta)\right] \Big/ \sum_{i \in B} \exp\left[\gamma_i + \tfrac{1}{2}(\beta^i + \delta)'\Omega(\beta^i + \delta)\right]. \end{aligned} \tag{16}$$

*Corollary* 1.1. Suppose the selection probabilities satisfy Equation (10) with given $\beta, \gamma_1, \ldots, \gamma_J$. Suppose the posterior distributions are multivariate normal with a common positive semidefinite covariance matrix $\Omega$. Then there exist posterior means satisfying Equation (15), mean selection probabilities satisfying Equation (16), and a prior distribution which is a mean selection probability mixture of the posterior distributions, such that

$$p_j(B, z) = q_j(z) \Big/ \sum_{i \in B} q_i(z).$$

[4] Let $K$ denote the dimension of $z_j$. Then $z$ is of dimension $JK$, where $J$ is the number of alternatives. The subspace $L$ is given by $L = \{\Omega z | z \in R^{JK}\}$, and its orthogonal complement $L^c$ is the null-space of $\Omega$, i.e., $L^c = \{z \in R^{JK} | \Omega z = 0\}$. Then $z \in L$ and $z \neq 0$ implies $z'\Omega z > 0$. Every vector $y \in R^{JK}$ has a unique representation $y = v + w$ with $v \in L^c$. Since $\Omega$ is symmetric and positive semidefinite, there exists an orthonormal matrix $A$ such that $AA' = I$ and

$$A'\Omega A = \begin{bmatrix} W & 0 \\ 0 & 0 \end{bmatrix},$$

where $W$ is a diagonal matrix with positive diagonal elements and rank equal to the dimension of $L$.

[5] The generalized inverse of $\Omega$ is defined to be the matrix

$$\Lambda = A \begin{bmatrix} W^{-1} & 0 \\ 0 & 0 \end{bmatrix} A',$$

in the notation of footnote 4. It is simple to verify using this formula that the system of equations $y = \Omega z$ has a solution if and only if $y \in L$, and that $y \in L$ implies $z = \Lambda y \in L$ is a solution, as is $z + w$ for any vector $w$ in the orthogonal complement of $L$.

*Corollary* 1.2. Suppose the selection probabilities satisfy Equation (10) with given $\beta$. Suppose the posterior distributions are multivariate normal with a common positive semidefinite covariance matrix $\Omega$. Suppose the mean selection probabilities $P_1, \ldots, P_J$ are given. Then there exists posterior means satisfying Equation (15), selection probability parameters $\gamma_1, \ldots, \gamma_J$ satisfying

(17) $$\gamma_j = \ln P_j - \frac{1}{2}\mu^{j\prime}\Lambda\mu^j - \frac{1}{J}\sum_{i\in B}\left(\ln P_i - \frac{1}{2}\mu^{i\prime}\Lambda\mu^i\right),$$

and a prior distribution which is a mean selection probability mixture of the posterior distributions, such that $p_j(B, z) = q_j(z)/\sum_{i\in B} q_i(z)$.

Proof: Substituting Equations (10) and (13) into formula (9) for $g$ yields

(18) $$g(z) = \sum_{i\in B} \exp\left[(z_i - z_j)'\beta + \gamma_i - \gamma_j\right]P_jK \exp\left[-\tfrac{1}{2}(z-\mu^j)'\Lambda(z-\mu^j)\right]$$
$$= \sum_{i\in \beta} \exp\left[z_i'\beta + \gamma_i + \log K - \tfrac{1}{2}z'\Lambda z - z_j'\beta + z'\Lambda\mu^j - \gamma_j + \log P_j - \tfrac{1}{2}\mu^{j\prime}\Lambda\mu^j\right].$$

Since the right-hand-side of this equation cannot depend on $j$, consistency requires

(19) $$-\gamma_j + \log P_j - \tfrac{1}{2}\mu^{j\prime}\Lambda\mu^j = \lambda,$$

where $\lambda$ is a constant, and

(20) $$-z_j'\beta + z'\Lambda\mu^j = z'\delta,$$

where $\delta$ is a vector of constants.

Equation (20) can be written

(21) $$z'\Lambda\mu^j = z'(\beta^j + \delta)\cdot(z \in L)$$

Taking $z = \Omega w$ for any real vector $w$, this implies $w'\mu = w'\Omega(\beta^j + \delta)$, or

(22) $$\mu^j = \Omega(\beta^j + \delta).$$

Substituting these expressions in Equation (18) yields

(23) $$g(z) = \sum_{i\in B} \exp\left[\log K - \tfrac{1}{2}z'\Lambda z + z'(\beta^i + \delta) + \gamma_i + \lambda\right]$$
$$= \sum_{i\in B} \exp\left[\log K - \tfrac{1}{2}z'\Lambda z + z'\Lambda\mu^i + \gamma_i + \lambda\right]$$
$$= \sum_{i\in B} \exp\left[-\tfrac{1}{2}z'\Lambda z + z'\Lambda\mu^i - \tfrac{1}{2}\mu^{i\prime}\Lambda\mu^i + \gamma_i + \tfrac{1}{2}\mu^{i}\Lambda\mu^i + \lambda\right]$$
$$= \sum_{i\in B} P_iK \exp\left[-\tfrac{1}{2}(z-\mu^i)'\Lambda(z-\mu^i)\right].$$

Q.E.D

## VI. The Consistency of Given Posterior Distributions

Suppose one is given multivariate normal posterior distributions with a common positive semidefinite covariance matrix. We seek conditions for the existence of multinomial logit selection probabilities of the form given in Equation (11). It will be convenient for this analysis to change notation slightly, defining $z' = (z'_{(1)}, \ldots z'_{(J)})$ and $\beta' = (\beta'_{(1)}, \ldots, \beta'_{(J)})$. In general, $z_{(j)}$ and $\beta_{(j)}$ vary with $j$. However, we consider also the cases where $z_{(j)}$ or $\beta_{(j)}$ are uniform across $j$. In the last of these cases, the multinomial logit equation (11) reduces to equation (10).

*Theorem* 2. Suppose the posterior distributions satisfy Equation (13) with given means $\mu^{j'} = (\mu^{j'}_{(1)}, \ldots, \mu^{j'}_{(J)})$ and a common positive semidefinite covariance matrix $\Omega$. Suppose the mean selection probabilities $P_j$ are given. Suppose the selection probabilities are required to have the form specified in Equation (11). Then the following conditions are necessary for consistency:

(1) The prior distribution is a probability mixture of the posterior distributions satisfying Equation (14).

(2) The parameter vector $\beta' = (\beta'_{(1)}, \ldots, \beta'_{(J)})$ satisfies

$$\beta_{(i)} = -J[\Lambda_i(\mu^j - \bar{\mu}) + q^j_{(i)}], \qquad (j \neq i) \tag{24}$$

where

$$\Lambda = \begin{bmatrix} \Lambda_1 \\ \vdots \\ \Lambda_J \end{bmatrix}$$

is a partition of $\Lambda$ left-commensurate with the partition of $\beta$.

$$\bar{\mu} = \frac{1}{J} \Sigma_{j \in B} \mu^j, \tag{25}$$

and the $q^{j'} = (q^{j'}_{(1)}, \ldots, q^{j'}_{(J)})$ are some vectors in the null-space of $\Omega$ (i.e., $\Omega q^j = 0$) satisfying

$$\sum_{j \in B} q^j = 0. \tag{26}$$

(3) The parameters $\gamma_1, \ldots, \gamma_J$ satisfy

$$\gamma_j = -\left(\ln P_j - \frac{1}{J} \sum_{i \in B} \ln P_i\right) + \frac{1}{2}\left(\mu^{j'} \Lambda \mu^j - \frac{1}{J} \sum_{i \in B} \Lambda \mu^i\right). \tag{27}$$

*Remark*. Equations (24), (25), and (26) imply

$$\beta_{(i)} = \frac{J}{J-1}\left[\Lambda_i(\mu^i - \bar{\mu}) + q^i_{(i)}\right] \tag{28}$$

Combining Equations (24) and (28) yields

$$\beta_{(i)} = \Lambda_i(\mu^i - \mu^j) + q^i_{(i)} - q^j_{(i)} \qquad (j \neq i) \tag{29}$$

Equations (26) and (29) plus the conditions $\Omega q^j = 0$ give $2J^2$ equations in the $J + J^2$ unknowns $\beta_{(i)}$ and $q^i_{(j)}$. Hence, the existence of a solution requires, in general, conditions yielding dependencies between equations. For example, if $\Omega$ is an identity matrix, then Equation (29) implies that a necessary condition for consistency is $\mu^j_{(i)} = \mu^k_{(i)}$ for $i \neq j, k$.

*Corollary* 2.1. If $\Omega$ is non-singular, then a necessary condition for consistency is

$$\Lambda_i(\mu^j - \mu^k) = 0 \qquad \text{for } i \neq j, k. \tag{30}$$

*Corollary* 2.2. If $J = 2$, then the solution

$$\begin{bmatrix} \beta_{(1)} \\ -\beta_{(2)} \end{bmatrix} = \Lambda(\mu^1 - \mu^2) \tag{31}$$

is consistent.

*Corollary* 2.3. If $\beta_{(1)} = \ldots = \beta_{(i)} = \ldots = \beta_{(J)}$, then a necessary condition for sufficiency is that $\Lambda_i(\mu^i - \mu^j) + q^i_{(i)} - q^j_{(i)}$ be independent of $i$ and $j$ for $i \neq j$.

*Corollary* 2.4. If $z_{(1)} = \ldots = z_{(i)} = \ldots = z_{(J)}$, then the solution

$$\beta_{(j)} = \Lambda_{11}\mu^j_{(1)}, \tag{32}$$

with $\Lambda_{11}$ the generalized inverse of the covariance matrix $\Omega_{11}$ of $z_{(1)}$, is consistent.

*Remark*. By defining $z_{(1)}$ in Corollary 2.4 to contain all the variables of the original problem, we obtain the general result that any multivariate normal posterior distributions with a common covariance matrix are consistent with a multinomial logit model of the form of Equation (11) with every variable appearing in the attributes of each alternative. The preceding results show that additional conditions on the posterior distributions are required to obtain multinomial logit models with added structure on the independent variables, as in Equation (10).

Proof: Equations (14) through (17) continue to be necessary and sufficient for consistency with

$$\beta^j = (0, \ldots, 0, \beta_{(j)}, 0, \ldots, 0).$$

In order to express Equation (15) in more detail, partition $\Lambda$ into submatrices $\Lambda_{ij}$, each square and of the same dimension as $\beta_{(j)}$, and write $\mu^j = (\mu^j_{(1)}, \ldots, \mu^j_{(J)})$ and $\delta = (\delta_{(1)}, \ldots, \delta_{(J)})$ commensurately with $z = (z_{(1)}, \ldots, z_{(J)})$. Then

$$\beta_{(j)} = \sum_k \Lambda_{jk}\mu^j_{(k)} - \delta_{(j)} + q^j_{(j)} \tag{33}$$

$$0 = \sum_k \Lambda_{ik}\mu^j_{(k)j} - \delta_{(i)} + q^j_{(i)} \qquad (i \neq j) \tag{34}$$

or

$$\beta_{(j)} = \sum_k \Lambda_{jk}(\mu^j_{(k)} - \mu^i_{(k)}) + q^j_{(j)} - q^i_{(j)} \qquad (i \neq j) \tag{35}$$

where as before the $\mu^j$ are assumed to lie in the non-null space $L$ of $\Omega$ and $q^j$ is a vector in the null space of $\Omega$ such that Equation (26) holds. Summing Equation (35) over $i \neq j$ yields

$$(J-1)\beta_{(j)} = \sum_k \Lambda_{jk}\Big(J\mu^j_{(k)} - \sum_{i\in B} \mu^i_{(k)}\Big) + Jq^j_{(j)} - \sum_{i\in B} q^i_{(j)}.$$

Using Equation (26), this implies Equation (28). Subtracting $J$ times Equation (35) from Equation (28) yields Equation (24). Equation (27) follows from Equation (17). This completes the proof of the theorem.

Corollaries 2.1 and 2.3 follow from Equation (35) and the observation that $\Omega q^j = 0$ and $\Omega$ non-singular implies $q^j = 0$. Corollary 2.2 is proved by verifying that the proposed solution satisfies

$$\begin{aligned} g(z) &= q_j(z)P_j/p_j(z) \\ &= \sum_{i\in B} \exp[z'_{(i)}\beta_{(i)} + \gamma_i + \log K - \tfrac{1}{2}z'\Lambda z - z'_{(j)}\beta_{(j)} + z'\Lambda\mu^j \\ &\quad -\gamma_j + \log P_j - \tfrac{1}{2}\mu^{j'}\Lambda\mu^j] \end{aligned}$$

with the right-hand-side independent of $j$. One has

$$\begin{aligned} -z'_{(1)}\beta_{(1)} + z'\Lambda\mu^1 = z'_{(1)}(\Lambda_{11}\mu^2_{(1)} + \Lambda_{12}\mu^2_{(2)}) \\ + z'_{(2)}(\Lambda_{22}\mu^1_{(2)} + \Lambda_{21}\mu^1_{(1)}) \equiv z'\delta \end{aligned}$$

and

$$\begin{aligned} -z'_{(2)}\beta_{(2)} + z'\Lambda\mu^2 = z'_{(1)}(\Lambda_{11}\mu^2_{(2)}) \\ + z'_{(2)}(\Lambda_{22}\mu^1_{(2)} + \Lambda_{21}\mu^1_{(1)}) \equiv z'\delta, \end{aligned}$$

where

$$\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}, \text{ yielding the result.}$$

Corollary 2.4 is established by considering

$$\begin{aligned} g(z_{(1)}) &= q_j(z_{(1)})P_j/p_j(z_{(1)}) \\ &= \sum_{i\in B} \exp[z'_{(1)}] + + \gamma_i + \log K - \tfrac{1}{2}z'_{(1)}\Lambda_{11}z_{(1)} - z'_{(1)}\beta_{(j)} \\ &\quad + z'_{(1)}\Lambda_{11}\mu^j_{(1)} - \gamma_j + \log P_j - \tfrac{1}{2}\mu^{j'}_{(1)}\mu^j_{(1)}] \end{aligned}$$

where $\Lambda_{11}$ is the generalized inverse of the covariance matrix $\Omega_{11}$ of $z_{(1)}$. When $\beta_{(j)} = \Lambda_{11}\mu^j_{(1)}$, the right-hand-side of this equation is independent of $j$. Q.E.D.

## VII. The Robustness of Discriminant Estimates of the Logit Model

We have established conditions under which statistics derived from posterior distributions under the postulate of normality provide consistent estimates of the selection probability parameters. The prior distribution required by these conditions, a probability mixture of the posterior distributions, seems unlikely to be realized in applications. Hence, it is of interest to examine the robustness of the estimator of the selection probability parameters derived under the postulates above when alternative prior distributions prevail. We consider the alternative of a normal prior. Suppose binary choice and a single real explanatory variable, with

$$(41) \qquad \begin{aligned} p_i(B, z) &= e^{\gamma_1+\beta z_1}/(e^{\gamma_1+\beta z_1} + e^{\gamma_2+\beta z_2}) \\ &= 1/(1+e^{-\gamma-\beta z}) \end{aligned}$$

where $\gamma = \gamma_1 - \gamma_2$ and $z = z_1 - z_2$, and

(42) $$g(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

Then

(43) $$P_1 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{z^2/2} \frac{dz}{1+e^{-\gamma-\beta z}},$$

(44) $$\mu_1 = \frac{1}{P_1\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{z}{1+e^{-\gamma-\beta z}} e^{-z^2/2}\, dz$$

(45) $$\mu_2 = -P_1\mu_1/P_2$$

(46) $$\sigma_1^2 = \frac{1}{P_1\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{(z-\mu_1)^2}{1+e^{-\gamma-\beta z}} e^{-z^2/2}\, dz$$

(47) $$\sigma_2^2 = (1 - P_1\sigma_1^2 - P_1\mu_1^2 - P_2\mu_2^2)/P_2$$

(48) $$\sigma^2 = P_1\sigma_1^2 + P_2\sigma_2^2$$

(49) $$\hat{\beta} = (\mu_1 - \mu_2)/\sigma^2$$

(50) $$\hat{\gamma} = (\log P_1/P_2) - \tfrac{1}{2}(\mu_1^2 - \mu_2^2)/\sigma^2$$

where $P_i$, $\mu_i$, $\sigma_i^2$ are the mean selection probability, posterior mean, and posterior variance, respectively, for $i = 1,2$, $\sigma^2$ is the "pooled" variance, and $\hat{\beta}$, $\hat{\gamma}$ are the discriminant estimators of $\beta$, $\gamma$. As shown in Figure 1, the discriminant estimator $\hat{\beta}$ underestimates in magnitude the true parameter $\beta$. The percent of the selection probabilities lying between 0.1 and 0.9 is 73 percent at $\beta = 2$, $\gamma = 0$ and 19 percent

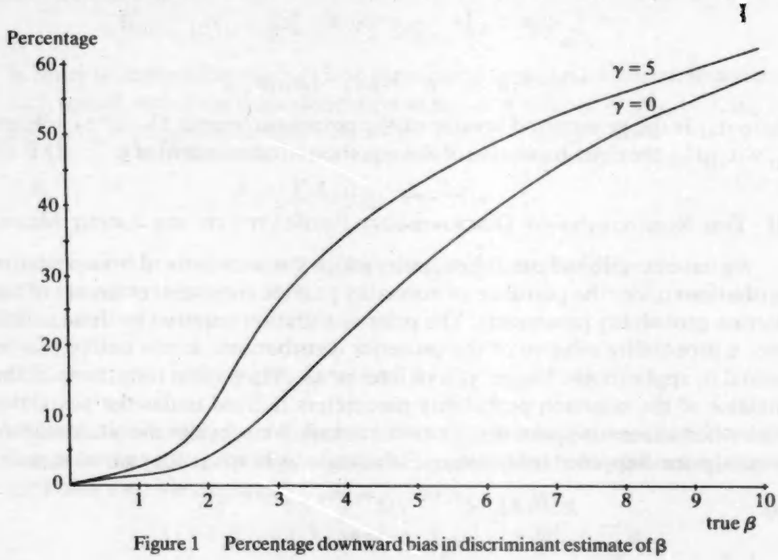


Figure 1 Percentage downward bias in discriminant estimate of β

at $\beta = 9$, $\gamma = 0$; these values would bracket the corresponding percentage in many applied studies. We conclude that for a typical prior distribution of the explanatory variables, multivariate normal, estimates of the selection probability parameters based on discriminant analysis will be substantially biased. Note that the discriminant estimator $\hat{\beta}$ coincides in finite samples with a linear probability model estimator; hence, this conclusion is consistent with results showing that the linear probability estimator applied to logistically generated responses leads to underestimates of the true parameters (McFadden (1973)).

## VIII. Conclusion

We conclude this comment with some observations on the experimental settings in which logit or discriminant analyses are appropriate. The first distinction to be made concerns the interpretation to be given to the response function $j = D(B, z, \xi)$ in Equation (3). On one hand, we may view this as a *causal* relationship, with $z$ and the unobserved vector $\xi$ determining $j$. On the other hand, we may view $(j, z)$ as being *conjoint*, or jointly distributed with no causal effect running from $z$ to $j$. In the first case, the function $D$ is of intrinsic methodological interest, while in the second case it is merely one of the ways of characterizing the joint distribution of $(j, z)$. Two examples will aid in exploring the implications of this distinction.

*Example* 1. (*Causal model*): Seeds are planted and observations $z$ are made on seed age, soil acidity, temperature, and time allowed for germination. Responses $j = 1$ (germination) and $j = 2$ (no germination) are observed.

*Example* 2. (*Conjoint model*): Eggs are candled, and observations $z$ are made on translucency. Responses $j = 1$ (high yolk = good egg) and $j = 2$ (spread yolk = bad egg) are observed.

In Example 1, theory suggests a causal relation between the explanatory variables and probability of germination. Then, the response function $D$ and selection probability will be of primary methodological interest. The selection probability would be used to forecast germination frequency for a new sample of seeds. It is not meaningful in this example to speak of two seed populations, "germinators" and "non-germinators," and attempt to classify seeds into one or the other. However, it is possible to classify seeds by probability of germination, and a binary classification into high and low probability germinators on the basis of selection probability is formally equivalent to a discriminant classification procedure.

In Example 2, translucency and yolk height can be viewed as jointly determined by unobserved variables, with no causal relation from translucency to yolk height. Then, the posterior distributions, or conditional distributions of $z$ given $j$, have the same status as the selection probabilities, or conditional distributions of $j$ given $z$. It is meaningful to speak of the populations of "good" and "bad" eggs, and attempt to classify an egg into one of these populations; this classification can be made using the selection probabilities.

We conclude from the comparison of these two examples that aside from the special causal interpretation given to the selection probabilities in causal models and the interpretation of the posterior populations in conjoint models, the problems of statistical analysis are identical, particularly with respect to the classification problem of forecasting response for new observations. Logit-type and discriminant-type statistical analysis could be used interchangeably, keeping in mind the logical interdependence of these models worked out earlier in this comment. *In any causal model, it becomes critical when the statistical formulation is of the discriminant type to check whether a consistent prior and selection probabilities exist, and whether the implied form of the selection probabilities is compatible with the underlying axioms of causality.*

An important distinction among quantal response models is whether it is meaningful to pose the question "If a policy is pursued which shifts a component of $z$, what is the effect on responses?". Clearly in a causal model this question is always meaningful, whether the component of $z$ is a characteristic of the subject or an attribute of an alternative. Thus, in Example 1, one may seek to determine the responsiveness of the germination probability to seed age or to time allowed for germination. What is important here is that the functional specification of the selection probabilities is assumed to not change when the policy changes, since it is determined by the underlying causal model. In a conjoint model, the question cannot be answered in general without specifying a causal relationship between underlying policy variables and $(j, z)$; there is no basis for assuming the functional specification of the selection probabilities remains unchanged when policy changes.

One distinction which has not been made in comparing causal and conjoint models is between characteristics of subjects and attributes of alternatives. It is often natural to associate with characteristics of subjects the notion of classifying the population into observable subpopulations according to response probabilities, and to associate with attributes of alternatives the notion of causal response. However, we have noted in discussing Example 1 that both types of variables, and the notion of classification, arise in causal models. Further, while conjoint models typically involve only characteristics of the subject, it is possible to give examples where attributes of alternatives enter, e.g., in Example 2 a dummy explanatory variable might appear indicating the method of measuring yolk height. We conclude that there is no logical relationship between causal or conjoint models on one hand and characteristics of subjects or attributes of alternatives on the other hand.

In summary, we see in causal models (1) that it is natural to specify problems in terms of selection probabilities, (2) that forecasting leads to classification problems within this model based on the selection probabilities, (3) that the model makes it meaningful to analyze the effects of policy affecting the explanatory variables, and (4) that the posterior distributions implied by the selection probabilities and prior distribution may provide a useful starting point for estimation of the selection probability parameters in a discriminant-type analysis, but this procedure does not tend to be robust with respect to misspecification of the prior. In conjoint models, (1) the posterior distributions and selection probabilities are alternative conditional distributions characterizing the joint distribution of $(j, z)$, and functional specifications can be made from either starting point, (2) classification procedures coincide with those of causal models despite the differing interpretation, and it (3) is generally not meaningful to pose questions about the effects of policies which shift the explanatory variables. In most social science applications, causal models are natural, suggesting that the models should be formulated in terms of selection probabilities, with discriminant-type methods applied to the posterior distributions only if there is considerable confidence in the validity of the implied specification of the prior.

*University of California, Berkeley*

## References

Johnson, N. L. (1949), "Systems of Frequency Curves Generated by Methods of Translations," *Biometrica*, 149–176.

Ladd, G. W. (1966), "Linear Probability Functions and Discriminant Functions," *Econometrica*, 873–885.

McFadden, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, editor, *Frontiers in Econometrics*, Academic Press.

McFadden, D. (1975), "Quantal Choice Analysis: A Survey," Department of Economics, University of California, Berkeley, this issue.

Westin, R. (1974), "Predictions from Binary Choice Models," *Journal of Econometrics*, April 1974.