NBER WORKING PAPER SERIES

ANALYSIS OF QUALITATIVE VARIABLES

G. S. Maddala*

Forrest D. Nelson**

Working Paper No. 70

COMPUTER RESEARCH CENTER FOR ECONOMICS AND MANAGEMENT SCIENCE
National Bureau of Economic Research, Inc.
575 Technology Square
Cambridge, Massachusetts 02139

October 1974

*Rochester University. Research supported in part by National Science Foundation Grant GJ-1154X3 to the National Bureau of Economic Research, Inc.

**NBER Computer Research Center. Research supported in part by National Science Foundation Grant GJ-1154X3 to the National Bureau of Economic Research, Inc.

Abstract

A variety of qualitative dependent variable models are surveyed with attention focused on the computational aspects of their analysis. The models covered include single equation dichotomous models; single equation polychotomous models with unordered, ordered, and sequential variables; and simultaneous equation models. Care is taken to illucidate the nature of the suggested "full information" and "limited information" approaches to the simultaneous equation models and the formulation of recursive and causal chain models.

Contents

1. Introduction

Very often economic variables observed are qualitative rather than quantitative, e.g., whether or not a person buys a car, whether or not a person buys a house, whether or not a person goes to a college, what mode of travel a person chooses, what occupation a person chooses, etc. If these variables are exogenous, there is no problem with the usual analysis. If these variables are endogenous then we have a problem. We will first review some of these models and then discuss the computational aspects. Earlier reviews can be found for example in Amemiya [2], Cox [8], McFadden [17] and Nerlove and Press [18]. In section 2 we outline the several functional forms that have been suggested for binary choice or dichotomous models. In section 3 we outline the polychotomous models and discuss the difference between the McFadden logit model and the usual multinomial logit model. In section 4 we discuss multivariate and simultaneous equations models and some problems of identification. In section 5 we discuss some estimation procedures and computational considerations that would be useful in the preparation of a package for the TROLL system.

2. Different Functional Forms

Consider the case where there are only two possible choices, e.g., buy a car or not, go to college or not, etc. The dependent variable y takes on the values 1 or 0 corresponding to the two choices. The easiest method is to ignore the nature of y and estimate the usual regression model

$$y = \beta'x + u \tag{1}$$

But since y takes on the values 1 or 0, u can take on only two values $(1-\beta'x)$ and $-\beta'x$. Hence we cannot assume homoscedasticity of the residuals (nor can we assume normality). Goldberger [10] discusses a two step procedure to tackle this problem of heteroscedasticity. Further discussion of the inadequacies of this model can be found in Nerlove and Press [18].

To solve the different complications arising from the direct estimation of (1), it has been suggested that a reasonable way of proceeding is to change the specification (1) so that we say

$$P = \text{Prob}\ (y = 1) = \text{Prob}\ (u < \beta'x) = F(\beta'x) \tag{2}$$

$$\text{and } 1-P = \text{Prob}\ (y = 0) = 1 - F(\beta'x)$$

and treat the observed values of y as a realization of a binomial process with these probabilities.

There have been several suggestions in the literature for the distribution of u.

(1) If u is Normal, $F(\beta'x)$ is cumulative normal and we have what is known as Normit or Probit analysis. $\beta'x$ is called Normit P and $\beta'x+5$ is called probit P. The term 'probit' is due to Bliss.

(2) If u has the sech^2 distribution

$$f(u) = \frac{e^u}{(1+e^u)^2}\ du \qquad -\infty < u < \infty$$

then $F(\beta'x)$ is the logistic $\dfrac{e^{\beta'x}}{1+e^{\beta'x}}$ (3)

$\beta'x = \text{Log}\,\frac{P}{1-P}$ is called logit P. It is also called log odds. Fisher and Yates defined logit P as $\frac{1}{2}\,\text{Log}\,\frac{P}{1-P}$. The term logit is due to Berkson.

(3) If u has the Cauchy distribution we have

$$F(\beta'x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} (\beta'x)$$

This is known as Urban's curve.

(4) Another function suggested by Knudsen and Curtis [4] is

$$F(\beta'x) = \frac{1}{2} [1 + \text{Sin}(\beta'x)] \qquad -\frac{\pi}{2} \leq \beta'x \leq \frac{\pi}{2}$$

The range for $\beta'x$ is restricted in this model.

(5) If $F(\beta'x) = e^{-e^{\beta'x}}$ the Gompertz curve we have what is known as Gompit analysis (see Zellner and Lee [24]. In this case $\ln \ln (\frac{1}{P}) = \beta'x$ is called Gompit P.

(6) If u follows the Burr distribution [6] then

$$F(\beta'x) = 1 - \frac{1}{[1+(\beta'x)^c]^k} \qquad c,k > 0, \quad \beta'x > 0$$

and we have Burrit analysis.

From the computational point of view there is possibly nothing to choose from. Except the normal distribution, all the other functions have closed forms for the cumulative distributions. However, this is not a major consideration in the choice between the above mentioned models because computationally fast and accurate approximations are available for the cumulative normal. What is of relevance is whether the tails of the distribution of u are expected to be thicker than those of the normal. If so then the other distributions have to be preferred. More importantly it is the symmetry aspect of the distribution of u that might be disturbing. If so one can use other distributions of u like the gamma or lognormal.

Though there is a wide variety in the underlying distributions for a binary choice or dichotomous model, this is not true for the polychotomous model except under special circumstances (when the variables are ordered or sequential).

3. Polychotomous Models

Here, as outlined in Cox ([8], chapter 7) and Amemiya [2], we have to distinguish between unordered, sequential and ordered variables. Examples of unordered variables are: choice of mode of transport - car, bus, train, choice of occupation - teacher, lawyer, doctor, plumber, etc. Suppose there are k categories. Let P_1 P_2 ... P_k be the probabilities associated with these k categories. Then the idea is to express these probabilities in binary form. Let

$$\frac{P_1}{P_1+P_k} = F(\beta_1'x)$$

$$\frac{P_2}{P_2+P_k} = F(\beta_2'x) \tag{4}$$

$$\vdots$$

$$\frac{P_{k-1}}{P_{k-1}+P_k} = F(\beta_{k-1}'x)$$

These imply

$$\frac{P_j}{P_k} = \frac{F(\beta_j'x)}{1-F(\beta_j'x)} = G(\beta_j'x) \tag{5}$$

for j = 1,2 ... k-1

Since $$\sum_{j=1}^{k-1} \frac{P_j}{P_k} = \frac{1-P_k}{P_k} = \frac{1}{P_k} - 1$$

we have $$P_k = [1 + \sum_{j=1}^{k-1} G(\beta_j'x)]^{-1} \tag{6}$$

and hence from (5) $$P_j = \frac{G(\beta_j'x)}{[1 + \sum_{j=1}^{k-1} G(\beta_j'x)]} \tag{7}$$

for j = 1,2 ... k-1

One can consider the observations as arising from a multinomial distribution with probabilities given by (6) and (7). Though, in principle, any of the previously mentioned underlying distributions of u can be used, from the computational point of view the logistic is the easiest to handle. In this case $G(\beta_j'x)$ in (5) is nothing but $e^{\beta_j'x}$. There are also other reasons for preferring the logistic which we will discuss later on. Since the logistic form will be the one that will be used in further analysis of the multinomial model, we will write down equations (6) and (7) explicitly. They are

$$P_j = e^{\beta_j'x}/D \qquad j = 1,2 \,..\, (k-1)$$

and

$$P_k = 1/D$$

where

$$D = 1 + \sum_{j=1}^{k-1} e^{\beta_j'x}$$

For an ordered response model Cox ([8], chapter 7) and Amemiya [2] mention the following model (originally analyzed by Aitchison and Silvey [1], Ashford [4], Gurtland, Lee and Dolan [12]). Here we assume

$$P_k = F(\beta'x)$$

$$P_k + P_{k-1} = F(\beta'x + \alpha_1)$$

$$P_k + P_{k-1} + P_{k-2} = F(\beta'x + \alpha_1 + \alpha_2) \qquad (8)$$

$$- - - -$$

$$P_k + P_{k-1} + \,..\, + P_2 = F(\beta'x + \alpha_1 + \alpha_2 + \,..\, + \alpha_{k-2})$$

and

$$P_1 = 1 - F(\beta'x + \alpha_1 + \alpha_2 + \,..\, + \alpha_{k-2})$$

where

$$\alpha_1, \alpha_2, \,..\, \alpha_{k-2} > 0$$

These equations imply

$$P_k = F(\beta'x)$$

$$P_{k-1} = F(\beta'x + \alpha_1) - F(\beta'x) \tag{9}$$

$$P_{k-2} = F(\beta'x + \alpha_1 + \alpha_2) - F(\beta'x + \alpha_1) \quad \text{etc.}$$

An example of ordered response variable is the following. Suppose we group individuals by the amount of expenditures they make on a car. We define the variable y as follows:

$y = 1$ if the individual spends < \$1,000

$y = 2$ if the individual spends > \$1,000 but < \$2,000

$y = 3$ if the individual spends > \$2,000 but < \$4,000

$y = 4$ if the individual spends > \$4,000

Finally, we have sequential models. An example of this is the following:

$y = 1$ if the individual has not completed high school

$y = 2$ if the individual has completed high school but not college

$y = 3$ if the individual has completed college but not a higher degree

$y = 4$ if the individual has a professional degree

Then the probabilities can be written as (see Amemiya [2])

$$P_1 = F(\beta_1'x)$$

$$P_2 = [1 - F(\beta_1'x)]\, F(\beta_2'x)$$

$$P_3 = [1 - F(\beta_1'x)][1 - F(\beta_2'x)]\, F(\beta_3'x) \tag{10}$$

$$P_4 = [1 - F(\beta_1'x)][1 - F(\beta_2'x)][1 - F(\beta_3'x)]$$

As Amemiya [2] points out, in such models the likelihood function can be maximized by maximizing the likelihood function of dichotomous models repeatedly.

For the case of ordered and sequential models any of the underlying distributions mentioned in section 2 can be used. For the ordered response model, there is only one underlying random variable u. For the sequential model with k categories, there are (k-1) underlying variables but we assume these to be independent. For instance, equations (10) are:

$$\begin{aligned} P_1 &= Pr(u_1 < \beta_1'x) \\ P_2 &= Pr(u_1 > \beta_1'x,\ u_2 < \beta_2'x) \\ P_3 &= Pr(u_1 > \beta_1'x,\ u_2 > \beta_2'x,\ u_3 < \beta_3'x) \\ P_4 &= Pr(u_1 > \beta_1'x,\ u_2 > \beta_2'x,\ u_3 > \beta_3'x) \end{aligned} \qquad (11)$$

One question we have to ask is what is the underlying distribution for the multinomial model given by (6) and (7)? Suppose there are k random variables $u_1\ u_2\ ..\ u_k$ and category j is chosen if

$$\beta_j'x+u_j < \beta_i'x+u_i \qquad \text{for all } i \neq j$$

i.e., $$P_j = Pr(u_j-u_i < \beta_j'x-\beta_i'x) \qquad \text{for all } i \neq j \qquad (12)$$

McFadden [16] shows that if u_i are independently and identically distributed with the distribution function

$$F(u) = e^{-e^{-u}}$$

then P_j in (12) can be written as $\dfrac{e^{\beta_j'x}}{\sum_{i=1}^{k} e^{\beta_i'x}}$

It is often customary to write the multinomial logit model without any discussion of the underlying probability distribution just by analogy to the binomial model

(Theil [21], Cox [7]). McFadden gave some justification to it in terms of a stochastic choice theory.

McFadden's derivation of the multinomial logit is very general and some of the economic applications that have been made are different from those of what is referred to as the multinomial logit in the statistical literature. He assumes (12) to be of the more general form.

$$P_j = \Pr\,[u_j - u_i < V_j(x_j) - V_i(x_i)]$$

Thus the function $V_j(x_j)$ can be of the form $\beta_j' x$ or $\beta' x_j$ or $\beta_j' x_j$ and the interpretation of the models is different.

As an illustration consider the two studies on occupational choice done by Boskin [5] and Schmidt and Strauss [19]. In the study by Boskin there are a number of occupations and each is characterized by three variables: present value of potential earnings, training cost/net worth, and present value of time unemployed. Let x_j denote the vector of the values of these characteristics for occupation j. The probability that an individual chooses occupation j is

$$P_j = \frac{e^{\beta' x_j}}{\sum_i e^{\beta' x_i}} \qquad (13)$$

Note the difference between this model and the multinomial model given by (6) and (7) where the P_j have different coefficient vectors β_j. In the model (13) the vector β gives the vector of implicit prices for the characteristics. Thus, the problem analyzed here is similar to that analyzed in the hedonic price index problem. Boskin obtains a different set of "implicit prices" for the characteristics for white males, black males, white females and black females. These coefficients tell us the relative valuation of the three characteristics mentioned above by these different groups. Also, if we are given a new occupation not considered in the estimation procedure and the characteristics of this new occupation, then we can use the estimated coefficients to predict the probability that a member chosen at random from each of the four groups would join this occupation. Thus, the main use of the model (13) is to predict the probability of choice for a category not considered in the estimation procedure but for which we are given the vector of characteristics x_j.

By contrast, the model on occupational choice considered by Schmidt and Strauss [19] is the usual multinomial model given by equations (6) and (7). Suppose there are k occupations and y_i is the vector of individual characteristics for individual i(age, sex, years of schooling, experience, etc.). Then the probability that the j^{th} occupation is chosen by the individual with characteristics y_i is

$$P_j = e^{\alpha_j' y_i} / \sum_m e^{\alpha_m' y_i} \tag{14}$$

with some normalization like $\alpha_k = 0$.

In (13) the number of parameters to be estimated is equal to the number of characteristics of the occupations. In (14) the number of parameters to be estimated is equal to the number of individual characteristics multiplied by (k-1) where k is the number of occupations. The questions answered by the models are different. In (14) we estimate the parameters and then given a new individual and his characteristics, we can predict the probabilities that he (or she) will choose one of k occupations considered.

Of course one can combine both the models and write

$$P_j = \frac{e^{\beta' x_j + \alpha_j' y_i}}{\sum_m e^{\beta' x_m + \alpha_m' y_i}} \tag{15}$$

In McFadden [16], he considers an example of shopping mode choice where he takes into account both the mode characteristics x_j (transit walk time, transit wait plus transfer time, auto vehicle time, etc.) and individual characteristics y_i (ratio of number of autos to number of workers in the household, race, occupation - blue collar or white collar).

In his discussion on choice of modes of transport Theil [22] specifies the following model: Let D_1 be bus, D_2 train, D_3 car, x_1, x_2, x_3 the respective travel times, and y income. Then he specifies the model as

$$\text{Log } \frac{P(D_r | x_1 \; x_2 \; x_3 \; y)}{P(D_s | x_1 \; x_2 \; x_3 \; y)} = \alpha_r - \alpha_s + (\beta_r - \beta_s) \log y + \gamma \log \left(\frac{x_r}{x_s}\right) \quad (16)$$

The implied probabilities P_j are

$$P_j = \frac{e^{\alpha_j + \beta_j \log y + \gamma \log x_j}}{\sum_{k=1}^{3} e^{\alpha_k + \beta_k \log y + \gamma \log x_k}}$$

which are of the same form as (15) except that the explanatory variables are in the log form.

4. <u>Multivariate Models</u>

Suppose there are two jointly dependent variables each taking the values 0 and 1. Then this can be reduced to a multinomial model with 4 categories and the methods of the previous section hold good. If there are too many parameters involved in this procedure then we can set some of the parameters equal to zero and simplify the models. This will be discussed later on.

This method is applicable if the variables are unordered. Amemiya [2] gives the extension of the sequential model to the bivariate case. It consists of specifying equations like (11) for each of the variables and assuming the corresponding residuals correlated with correlations ρ_1 ρ_2 ρ_3 respectively. The bivariate extension of the ordered response case follows along the same lines. We specify equations like (8) or (9) for each of the variables and since there is only one underlying random variable u (viz., u_1 for the first variable and u_2 for the second variable) we have to estimate also a correlation coefficient ρ. This model is computationally cumbersome to handle. Hence it may be better, from the practical point of view, to ignore the ordering. Mantel ([15], p. 91) and Nerlove and Press ([18], p. 22) suggest that the analysis be not constrained by ordering.

Thus, from the practical point of view it is enough to consider the unordered case. For illustrative purposes we will consider the case of three dichotomous variables y_1 y_2 and y_3.

Let $P_{ijk} = \Pr\ (Y_1 = i,\ Y_2 = j,\ Y_3 = k) \qquad i,j,k = 0 \text{ or } 1.$

One question that we can ask is why not just treat this case as a multinomial model with 8 catogories? We will here show what happens if we do that. Since, as mentioned earlier, the multinomial model is most easily handled for the logistic case, we will make that assumption. We can then write

$$\begin{aligned} P_{000} &= 1/D \\ P_{100} &= e^{\beta_1' x}/D \\ P_{010} &= e^{\beta_2' x}/D \\ P_{001} &= e^{\beta_3' x}/D \\ P_{110} &= e^{\beta_4' x}/D \\ P_{101} &= e^{\beta_5' x}/D \\ P_{011} &= e^{\beta_6' x}/D \\ P_{111} &= e^{\beta_7' x}/D \end{aligned} \tag{17}$$

where $$D = 1 + \sum_{i=1}^{7} e^{\beta_i' x}$$

These equations imply the following relations:

$$\frac{P_{100}}{P_{000}} = e^{\beta_1' x} \qquad \frac{P_{110}}{P_{010}} = e^{(\beta_4-\beta_2)'x} \qquad \frac{P_{101}}{P_{001}} = e^{(\beta_5-\beta_3)'x} \qquad \frac{P_{111}}{P_{011}} = e^{(\beta_7-\beta_6)'x}$$

$$\frac{P_{010}}{P_{000}} = e^{\beta_2' x} \qquad \frac{P_{110}}{P_{100}} = e^{(\beta_4-\beta_1)'x} \qquad \frac{P_{011}}{P_{001}} = e^{(\beta_6-\beta_3)'x} \qquad \frac{P_{111}}{P_{101}} = e^{(\beta_7-\beta_5)'x}$$

$$\frac{P_{001}}{P_{000}} = e^{\beta_3' x} \qquad \frac{P_{101}}{P_{100}} = e^{(\beta_5-\beta_1)'x} \qquad \frac{P_{011}}{P_{010}} = e^{(\beta_6-\beta_2)'x} \qquad \frac{P_{111}}{P_{110}} = e^{(\beta_7-\beta_4)'x}$$

These relations can be written as

$$\begin{aligned}
\text{Log}\,\frac{P(y_1 = 1|y_2\ y_3)}{P(y_1 = 0|y_2\ y_3)} &= \beta_1'x + (\beta_4-\beta_2-\beta_1)'x\ y_2 + (\beta_5-\beta_3-\beta_1)'x\ y_3 \\
&\quad + (\beta_7-\beta_6-\beta_5-\beta_4+\beta_3+\beta_2+\beta_1)'x\ y_2\ y_3 \\
\text{Log}\,\frac{P(y_2 = 1|y_1\ y_3)}{P(y_2 = 0|y_1\ y_3)} &= \beta_2'x + (\beta_4-\beta_2-\beta_1)'x\ y_1 + (\beta_6-\beta_3-\beta_2)'x\ y_3 \\
&\quad + (\beta_7-\beta_6-\beta_5-\beta_4+\beta_3+\beta_2+\beta_1)'x\ y_1\ y_3 \\
\text{Log}\,\frac{P(y_3 = 1|y_1\ y_3)}{P(y_3 = 0|y_1\ y_2)} &= \beta_3'x + (\beta_5-\beta_3-\beta_1)'x\ y_1 + (\beta_6-\beta_3-\beta_2)'x\ y_2 \\
&\quad + (\beta_7-\beta_6-\beta_5-\beta_4+\beta_3+\beta_2+\beta_1)'x\ y_1\ y_2
\end{aligned} \tag{18}$$

Note the symmetry in the coefficients of the equations (18). This symmetry was discussed first by Nerlove and Press [18] and later by Amemiya [2] and Schmidt and Strauss [20]. The model considered by them is a special case of (18) with the following conditions:

$$
\begin{aligned}
(\beta_4 - \beta_2 - \beta_1)'x &= \beta_{12} \\
(\beta_5 - \beta_3 - \beta_1)'x &= \beta_{13} \\
(\beta_6 - \beta_3 - \beta_2)'x &= \beta_{23} \\
(\beta_7 - \beta_6 - \beta_5 - \beta_4 + \beta_3 + \beta_2 + \beta_1)'x &= \gamma
\end{aligned} \tag{19}
$$

We can get this model if the first element of x is 1, all but the first elements of the vector β_4 are equal to the sum of the corresponding elements of β_2 and β_1, with similar conditions holding for β_5 and β_6, and for β_7 all but the first element are equal to the sum of the corresponding elements of β_1, β_2 and β_3.

Thus, an important consequence of the multinomial logistic model (17) is that we get the well defined conditional distributions (18). In actual practice, if there are a number of categories, the complete multinomial model (17) involves too many parameters. That is why Nerlove and Press suggest estimating equations (18) arguing that one can get consistent estimators for the parameters (though these are not fully efficient because they ignore the cross equation constraints). This procedure reduces the number of parameters to be estimated considerably. Further reduction can be achieved by making some simplifying assumptions like (19). If we further impose the restriction $\beta_7-\beta_6-\beta_5-\beta_4+\beta_3+\beta_2+\beta_1 = 0$ we can also eliminate the product terms involving $y_1\, y_2$, $y_2\, y_3$, $y_3\, y_1$ in equations (18).

Unlike the usual simultaneous equations model where it is not possible to interpret each equation as a conditional expectation (except in a recursive system) the specification (17) permits of well defined conditional probabilities (18). Also, it looks as if we cannot have causal chains in simultaneous equation logit models. This is indeed not so. Consider a situation where the causal relations between $y_1\, y_2\, y_3$ are as shown in figures 1 and 2.

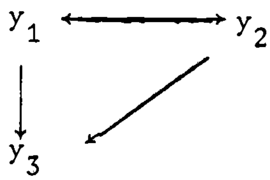

Figure 1

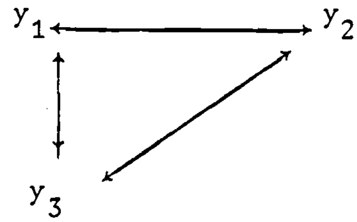

Figure 2

Suppose that y_1 and y_2 are variables that do precede (in time or in some other sense) variable y_3. Then a relationship as in Figure 2 obviously does not make sense and it is a relationship as in Figure 1 that we should be considering. It might be thought that the symmetry conditions in equations (18) imply that if y_3 depends on y_1, then the reverse must be true with the same effect. This is of course not true. What the symmetry conditions imply is that if y_1 depends on y_3 and y_3 depends on y_1 then the two effects should be equal. We have to interpret the conditional probability equations (18) as depicting the nature of the causal relationships between the variables. For the model in Figure 1 these causal relationships can be written in the following form

$$\text{Log}\,\frac{\Pr(y_1 = 1|y_2,x)}{\Pr(y_1 = 0|y_2,x)} = \delta y_2 + \alpha_1'x$$

$$\text{Log}\,\frac{\Pr(y_2 = 1|y_1,x)}{\Pr(y_2 = 0|y_1,x)} = \delta y_1 + \alpha_2'x \tag{20}$$

$$\text{Log}\,\frac{\Pr(y_3 = 1|y_1,y_2x)}{\Pr(y_3 = 0|y_1,y_2x)} = \beta_1 y_1 + \beta_2 y_2 + \alpha_3'x$$

Note that the symmetry conditions have been imposed only for the first two equations in (20) since y_1 and y_2 are jointly determined. One can estimate δ, α_1, α_2 from the joint probability distribution of y_1 and y_2. These joint probabilities are:

$$P_{11} = e^{(\alpha_1+\alpha_2)'x + \delta}/\Delta$$
$$P_{01} = e^{\alpha_2'x}/\Delta$$
$$P_{10} = e^{\alpha_1'x}/\Delta \qquad (21)$$
$$P_{00} = 1/\Delta$$

where $\Delta = 1 + e^{\alpha_1'x} + e^{\alpha_2'x} + e^{(\alpha_1+\alpha_2)'x + \delta}$

As for the third equation in (20) its parameters are estimated separately. This equation implies

$$\text{Log}\,\frac{P_{111}}{P_{110}} = \beta_1 + \beta_2 + \alpha_3'x$$
$$\text{Log}\,\frac{P_{011}}{P_{010}} = \beta_2 + \alpha_3'x$$
$$\text{Log}\,\frac{P_{101}}{P_{100}} = \beta_1 + \alpha_3'x \qquad (22)$$
$$\text{Log}\,\frac{P_{001}}{P_{000}} = \alpha_3'x$$

and equations (22) in conjunction with (21) will enable us to estimate the joint probabilities P_{ijk} for any goodness of fit tests. If we assume the causal relationship in Figure 2, the conditional probabilities will be given by equations (18), with any appropriate zero restrictions, and the joint probabilities will be given by (17), again with the appropriate zero restrictions.

Given any specification of the conditional odds ratios as in (18) one can deduce the joint probabilities (17). The ML estimation procedure based on the implied joint probabilities (17) has been called the full information ML procedure by Nerlove and Press [18]. They argue that it is computationally less cumbersome to estimate the conditional equations (18), that these estimates though not fully efficient are consistent (see Amemiya [2] for a heuristic proof) and that in practice these should be adequate.

In the case of a recursive model, of course, as in the usual simultaneous equations context, the estimates from the conditional equations (18) would be fully efficient. As an illustration consider the causal model

$$y_1 = f(x)$$

$$y_2 = f(x, y_1)$$

where y_1 and y_2 are binary.

$$\Pr(y_1 = 1) = \frac{e^{\beta_1' x}}{1 + e^{\beta_1' x}}$$

$$\Pr(y_2 = 1 | y_1) = \frac{e^{\beta_2' x + \gamma y_1}}{1 + e^{\beta_2' x + \gamma y_1}} \tag{23}$$

These give the joint probabilities

$$P_{11} = F(\beta_1' x)\ F(\beta_2' x + \gamma)$$

$$P_{10} = F(\beta_2' x)\ [1 - F(\beta_1' x)]$$

$$P_{01} = F(\beta_1' x)\ [1 - F(\beta_2' x + \gamma)] \tag{24}$$

$$P_{00} = [1 - F(\beta_1' x)]\ [1 - F(\beta_2' x)]$$

where $$F(z) = \frac{e^z}{1 + e^z}$$

The separate estimation of equations (23) and the joint estimation of equations (24) are the same.

5. Methods of Estimation and Computational Considerations

Though a variety of models have been enumerated, and all these models can, in principle, be analyzed with any of the functional forms mentioned in section 2, the most commonly used functional form is the logistic and the most commonly used models are

(a) the McFadden logit model given by equation (13)

(b) the multinomial logit model given by equation (14)

(c) the general logit model given by equation (15)

and (d) the simultaneous equations models of the Nerlove-Press type.

The ordered response models may be just treated as unordered multinomial models following the suggestion of Mantel [15]. For instance Cragg and Uhler [9] use the multinomial model to study the decisions:

(i) sell a car, replace a car, purchase additional car

and (ii) number of cars owned 1, 2, 3 or more.

Strictly speaking neither of these can be treated as a standard multinomial model but possibly introducing additional complications is not worthwhile.

Of the four models mentioned above, the likelihood function for (a) seems to be better behaved than for (b). McFadden [16] reports that the Newton-Raphson method converged slowly and that Davidson's variable metric method gave faster convergence. He says that the ML method proved practical up to 20 variables and 2000 observations. It is obvious that for model (b) it is not possible to handle 20 variables. The number of parameters to be estimated for model (b) increases with the number of categories, unlike the case for model (a). Schmidt and Strauss [19] use five occupational categories and have five variables (including a constant term). Thus they estimate $5 \times 4 = 20$ parameters. This seems to be the number of parameters that can be conveniently handled in these models.

Jones [13] reports in connection with the estimation of model (b) that initial estimates obtained from the linear discriminant functions were often far from the ML solution and that simple Newton-Raphson procedures often diverged. He uses the Fletcher-Powell method and Davidson's variable metric method. On the other hand there have been others (like Haggstrom at Rand) who have argued that their experience is that the ML estimates are not much different from those given by the discriminant function and hence it is not worthwhile bothering to use the

ML method. In any case, the initial values for method (b) are almost always obtained from the discriminant function. The same is not true for method (a). The multinomial model (b) has indeed been derived using the discriminant function approach (e.g., see Cox [7], Jones [13]) and there is a close relationship between the two. The same is not the case with McFadden's model - model (a).

Walker and Duncan [23] suggest a weighted least squares method for the multinomial logit model. The procedure is as follows:

$$\text{Write } p_n = f(x_n,\beta) + \varepsilon_n \qquad n = 1,2, \ldots N$$

where $f(x_n,\beta) = P_n = [1 + \exp(- x_n'\beta)]^{-1}$

$$E(\varepsilon_n) = 0 \qquad \text{Var}(\varepsilon_n) = P_n Q_n \tag{25}$$

Expand p_n in a Taylor series around an initial guess or estimate $\bar{\beta}$.

$$p_n \simeq f(x_n,\bar{\beta}) + f'(x_n,\bar{\beta})(\beta-\bar{\beta}) + \varepsilon_n$$

Let $y_n^* = p_n - f(x_n,\bar{\beta}) + f'(s_n,\bar{\beta})\bar{\beta}$

Then $y_n^* \simeq f'(x_n,\bar{\beta})\beta + \varepsilon_n$ (26)

Walker and Duncan weighted iterative non-linear least squares based on equations (25) and (26). Amemiya [3] shows that this method is equivalent to the method of scoring. Thus, we need not consider this method as an alternative to the ML method.

Broadly speaking we need two sets of programs to handle the two types of data we encounter: grouped and ungrouped data. For the former we can use weighted least squares methods. For the latter the ML methods. If there is only one explanatory variable x, then even if we have detailed observations, we might want to group the data and use the weighted least squares methods because they are much simpler to use. This is what was done by Zellner and Lee [24] where there was only one explanatory variable, income, and the spending units were grouped into 12 income classes.

For the binomial model, the weighted least squares method (also known as minimum logit X^2 method) is as follows:

Let $L_i = \text{Log} \frac{P_i}{1 - P_i}$ be the true logits.

$l_i = \text{Log} \frac{f_i}{1 - f_i}$ be the observed logits

f_i is the sample proportion

write $l_i = \beta' x + \varepsilon_i$

where $\varepsilon_i = \text{Log} \frac{f_i}{1 - f_i} - \text{Log} \frac{P_i}{1 - P_i}$

Now $\text{Var}(f_i) = \frac{P_i(1 - P_i)}{n_i}$

Thus, expanding Log $\frac{f_i}{1-f_i}$ around P_i we get

$$\text{Log} \frac{f_i}{1 - f_i} - \text{Log} \frac{P_i}{1 - P_i} \simeq \frac{f_i - P_i}{P_i(1 - P_i)}$$

Hence $E(\varepsilon_i) = 0 \qquad \text{Var}(\varepsilon_i) = \frac{\text{Var}\ (f_i)}{[P_i(1 - P_i)]^2} = \frac{1}{n_i P_i (1 - P_i)}$

This variance will be estimated by $\frac{1}{n_i f_i (1 - f_i)}$

Thus, the weighted least squares procedure would be to regress $l_i \sqrt{n_i\ f_i\ (1 - f_i)}$ on the x's to get estimates of the β's.

Theil extends this method to the multinomial model (see Theil [22]).

McFadden's program reduces the general logit model (15) with any extra interaction terms, into a model of the type (13) with an appropriate definition of dummy variables corresponding to the individual characteristics y_i. In practice this would involve the creation of a very large number of variables many of which will assume zero values for a large number of cases and it is doubtful if this is the most appropriate way of handling the model on the Troll system.

Goodman has a program for the analysis of contingency tables. In econometric work we anyhow need to incorporate a number of exogenous variables. Thus the Goodman programs would not be of much use. The Nerlove-Press program does allow for exogenous explanatory variables. However, the program imposes the conditions (19) and also assumes all second and higher order interactions to be zero (i.e., coefficients of y_1y_2, y_2y_3, etc., in (18) are zero). Also, as discussed earlier one might not always want to consider a model in which all the endogenous variables are determined jointly. One might sometimes want to formulate causal chain models. How much flexibility one should allow will depend on what sort of problems one would be solving on the Troll system.

References

[1] Aitchison, J. and S. Silvey - "The Generalization of Probit Analysis to the Case of Multiple Responses" - Biometrika (1957) pp. 131-140

[2] Amemiya, T. - "Qualitative Response Models" - Technical Report No. 135, June 1974, Institute for Mathematical Studies in the Social Sciences, Stanford University

[3] Amemiya, T. - "The Equivalence of the Non-linear Weighted Least Squares Method and the Method of scoring in the General Qualitative Response Model" - Technical Report No. 137, June 1974, Stanford University

[4] Ashford, J. R. - "An Approach to the Analysis of Data for Semi-Quantral Responses in Biological Assay" - Biometrics (1959) pp. 573-81

[5] Boskin, M. J. - "A Conditional Logit Model of Occupational Choice" - Journal of Political Economy April 1974

[6] Burr, I. W. - "Cumulative Frequency Functions" - Annals of Mathematical Statistics (1942) pp. 215-232

[7] Cox, D. R. - "Some Procedures Connected with the Logistic Qualitative Response Curve" - Research Papers in Statistics (F. N. David ed.), (Wiley, New York) 1966, pp. 55-71

[8] Cox, D. R. - "Analysis of Binary Data (Metheun, London) 1970

[9] Cragg, J. G. and R. S. Uhler - "The Demand for Automobiles" - Canadian Journal of Economics August 1970, pp. 386-406

[10] Goldberger, A. S. - Economic Theory (Wiley, New York) 1964

[11] Goodman, L. A. - "Causal Analysis of Date from Panel Studies and Other Kinds of Surveys" - American Journal of Sociology March 1973, pp. 1135-1191

[12] Gurland, J., I. Lee and P. A. Dolan - "Polychotomous Quantal Response in Biological Assay" - Biometrics (1960) pp. 382-98

[13] Jones, R. H. - "Probability Estimation Using a Multinomial Logistic Function

[14] Knudsen, L. F. and J. M. Curtis - "The Use of Angular Transformation in Biological Assays" - J.A.S.A. (1947) pp. 282-296

[15] Mantel, N. - "Models for Complex Contingency Tables and Polychotomous Response Curves" - Biometrics (1966) pp. 83-110

[16] McFadden, D. - "Conditional Logid Analysis of Qualitative Choice Behaviour" - in Frontiers in Econometrics (ed. by P. Zasembka), Academic Press, New York, 1973, pp. 105-142

[17] McFadden, D. - "Quantal Choice Analysis: A Survey" - March 1974

[18] Nerlove, M. and S. J. Press - Univariate and Multivariate Log-linear and Logistic Models, Rand Memorandum R-1306-EDA/NIH, December 1973

[19] Schmidt, P. and R. P. Strauss - "The Prediction of Occupational Level Using Multiple Logid Models: A Study of Labor Market Discrimination" - (University of North Carolina, Chapel Hill)

[20] Schmidt, P. and R. P. Strauss - "Estimation of Models with Jointly Dependent Qualitative Variables: A Simultaneous Logit Approach" - (University of North Carolina, Chapel Hill)

[21] Theil, H. - "A Multinomial Extension of the Linear Logit Model" - International Economic Review October 1969, pp. 251-259

[22] Theil, H. - Statistical Decomposition Analysis - (North Holland Publishing Company, Amsterdam) 1972

[23] Walker, S. H. and D. B. Duncan - "Estimation of the Probability of an Event as a Function of Several Independent Variables" - Biometrika (1967) pp. 167-179

[24] Zellner, A. and T. H. Lee - "Joint Estimation of Relationships Involving Discrete Random Variables" - Econometrica April 1965, pp. 382-394