A COMPARISON OF QUALITATIVE RESPONSE MODELS OF CONSUMER CHOICE

Vithala R. Rao and Hans S. Solgaard

#399

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign
A COMPARISON OF QUALITATIVE RESPONSE MODELS OF CONSUMER CHOICE

Vithala R. Rao and Hans S. Solgaard

#399
A COMPARISON OF QUALITATIVE RESPONSE MODELS OF CONSUMER CHOICE

Vithala R. Rao*
University of Illinois

Hans S. Solgaard
Cornell University

Working Paper
April, 1977

* Vithala R. Rao is Visiting Professor of Business Administration at the University of Illinois and Hans S. Solgaard is a doctoral candidate in marketing at the Graduate School of Business and Public Administration at Cornell University.
ABSTRACT

Several choice models applicable to qualitative response data collected in marketing research are reviewed in this paper. Following a discussion of a general model, four binary choice models are compared in terms of underlying choice processes and methods of estimation. Availability of computer algorithms for analysis and areas of application in marketing are also discussed.
I. Introduction

Development and testing of models to describe consumer choice has been a major concern in marketing and consumer behavior research [6, 16, 29]. The criterion variable—consumer choice—had been operationalized in many ways in the literature. Measures employed include amount bought or consumed of a product or brand, brand chosen, intention to buy a brand, preference toward a brand, and probability of brand switching. Usually, however, only one measure is used at a time for model construction.

From a technical viewpoint, measures of consumer choice belong to the three basic scales of measurement, namely, interval, ordinal or categorical. Methods of analysis associated with these measures have respectively been multiple regression, ordinal regression, and two-group or multiple discriminant analysis [8, 12].

The focus of this paper is on models when the consumer choice is measured on a categorical scale. This scale represents a variety of consumer choice situations such as buying or not buying a brand, viewing or not viewing television, buying a gift or not, an industrial buyer seeing a salesman or not, etc. In addition, the scale can also represent particular choices made within a set of alternatives such as brands of a product category, prime time television programs, television news programs, and suppliers of an industrial product.

Even when the measure of consumer choice is not categorical, it can easily be converted to that scale by a suitable regrouping (or collapsing) of the original scale. For example, consumers can be classified as heavy or light on the basis of amount of reported consumption measured on an interval scale. Such a conversion offers a potential advantage of reducing the errors associated with data collection. In addition, the concept of finally using a
categorical scale would make it easier to collect such data in the first place (as opposed to later conversion).

Despite the apparent niceties of the qualitative response variable (i.e., categorical scaled data), much of the model building of consumer choice has largely concentrated on measures of interval or ordinal scale. When the data are categorical, researchers usually utilize chisquare (contingency) analysis or multivariate discriminant analysis. An application of the multivariate probit model for purchasing decisions of farmers can be found in [17]. It is only recently that other models, namely, logit and log-linear, are proposed and used in marketing research [11]. The emphasis of this application is on contingency table analysis in contrast to model building of consumer choice, per se. These two models are relatively simple and quite well-known, but not much used in marketing prior to this application.

During the last five years or so, there has been a renewed interest in the analysis and modeling of qualitative dependent variables among econometricians. The interest apparently arose due to the need to look at consumption data such as the transportation mode choices and the inadequacy of using ordinary least squares analysis on a qualitative dependent variable owing to heteroskedasticity [10]. This recent effort gave rise to extensions of techniques such as the probit and logistic models and associated computer algorithms.

Against this background, the objective of this paper is to review some alternative models of consumer choice applicable to qualitative responses. Specifically, we will consider four models: (1) discriminant model, (2) linear probability model, (3) multivariate probit model, and (4) multivariate logit model. For sake of simplicity, we will only consider binary choice situations in some parts of the discussion.
The remainder of this paper is organized into five additional sections. In the next and second section, we present the notation and the general problem of modeling consumer choice using qualitative response data and one particular case leading to the above-mentioned four models. The third section describes briefly the methods of estimation for four binary choice models. The problem of measuring the effect of changes in the independent variables (e.g., characteristics of consumers or choice alternatives) on the probability of choosing an alternative for each model is considered in the fourth section. A brief review of the computer programs available for analysis of data according to these models is presented in the fifth section. We conclude the final section with a discussion of potential applications in marketing and some research issues with these models.

II. A General Model for Qualitative Responses

The problem of modeling qualitative responses of consumer choice from the econometric perspective has been reviewed by McFadden [20, 22]. When the choice is binary, the work by Cox [4] is relevant. Other references from a theoretical point of view include [14, 24, 26, 30]. While we do not wish to trace through the historical origins of the subject, mention should be made of the pioneering work on probit analysis by Finney [7]. Some applications in areas other than marketing are found in [5, 13, 27, 32, 33]. In the sequel, we will adapt much of this literature as it relates to the problem of modeling qualitative responses of consumer choice in marketing.

To see the relevance of the problem to marketing and consumer research, consider the following situation. Imagine observing a sample of consumers choosing one of many brands in a product category under a set of different choice situations or scenarios during a given period of time. Assume further
that it is possible to observe at least one choice for each consumer under each situation in this period. (Of course, many of these replications may be only one.) The data observed, namely, the brand chosen, are then the responses for each consumer and are qualitative. The response is related to the characteristics of consumers, characteristics of brands, and the characteristics of the situation. The problem of modeling the qualitative responses deals with the specification of the form of the function \( f \) relating the various characteristics to the probability of response for each brand. The methods of estimation are concerned with the determination of the parameters of \( f \) as dependent on the availability of replications of observations and the number of brands. We will assume that there exists no order in the responses (i.e., brands are not ordered).

**Notation.** We will adopt the following notation to present the model in a formal manner. For simplicity, we will consider the case of one choice situation.

\[
\begin{align*}
m &= \text{number of consumers} \\
n &= \text{number of brands} \\
R_i &= \text{number of replications observed under the situation for the } i\text{th consumer}; \ i=1,2, \ldots, m; \ (R_i \geq 1) \\
J &= \text{set of possible responses for any replication (i.e., set of } n \text{ brands)} \\
r &= \text{number of attributes of the brands} \\
s &= \text{number of characteristics of the consumer} \\
X_{B_j} &= \text{r-dimensional vector of attributes for the } j\text{th brand}; \ (j=1,2,\ldots,n) \\
X_{C_i} &= \text{s-dimensional vector of characteristics for the } i\text{th consumer}; \ (i=1,2,\ldots,m) \\
\beta &= \text{r-dimensional parameter vector associated with the brand attributes} \\
\gamma &= \text{s-dimensional parameter vector associated with consumer characteristics}
\end{align*}
\]
\( F_{ij} \) = observed frequency with which brand \( j \) is chosen across all replications by the \( i \)th consumer \((i=1,2,...,m; \ j=1,2,...,n)\)

\( P_{ij} \) = observed probability of choice of the \( j \)th brand \((P_{ij} = F_{ij}/R_{i}; \ i=1,2,...,m; \ j=1,2,...,n)\)

\( Y_i \) = \( n \)-dimensional vector of probabilities \((P_{i1},P_{i2},...,P_{in})\) of choice of the \( n \) brands for the \( i \)th consumer; \( i=1,2,...,m \). (If there is only one replication, then \( Y_i \) will contain \( n-1 \) zeroes and one unity.)

\( a \) = a constant parameter.

Models. The theory of qualitative responses postulates the existence of an indicator variable, denoted by \( I \), which takes on different values across various brands for a given consumer. In general, it is assumed to be a function of variables \( XB \) and \( XC \). Much of the modeling work involves specification of the functional form for \( I \). A convenient starting point is the linear model such as:

\[
I = \alpha + \beta \cdot XB + \gamma \cdot XC .
\] (1)

This form can be easily extended to include within-set interactions among the attributes of brands or characteristics of consumers as well as between set interactions. Generally, however, such specification should be guided by the substantive nature of the choice problem being modeled.

Further, the consumer is assumed to have threshold values on the indicator scale which lead to the choices of various brands. Assumption of a particular probability distribution for the threshold values would then generate a set of theoretical choice probabilities \((\pi_{i1},\pi_{i2},...,\pi_{in})\) for the \( i \)th person which are functions of parameters associated with brands and/or persons. The observed frequencies \((F_{i1},F_{i2},...,F_{in})\) can then be assumed to arise from a multinominal distribution with these theoretical probabilities. The parameters can be estimated using maximum likelihood methods or least squares methods.
This approach is indeed complicated for the general case. Several simplifications occur, however, for special cases. In particular, we will consider the binary choice case (i.e., \( n=2 \)) to compare the above-mentioned four models for the situation of one replication. This situation is highly appropriate in marketing where much of the analysis deals with cross-sectional data. Such data come closest to the case of one replication. Further, the binary choice analysis can be repeated to model the choices with respect to each brand in the choice set.

**Case \( n=2 \).** Here, there are only two choice alternatives. Therefore, we can reduce the vector variable \( Y \) to a scalar variable by considering only the probabilities for one of the two brands. Such reduction would preserve all of the information in the data for the case of one replication. Further, the reduced variable is either 1 or 0.

The index can be written simply in terms of the consumer specific variables. Thus, the model would become:

\[
I = \alpha + \gamma^\prime X .
\]

Let \( I^*_i \) denote the threshold value specific to the \( i \)th consumer. The four models—discriminant model, linear probability model, multivariate probit model, and multivariate logistic model—would result from different assumptions on the probability distributions for the threshold values, \( I^*_i \). These are shown in Table 1. The reader should note that different assumptions are also involved with respect to the threshold values across consumers. We have shown an extremely simplified conceptualization of the discriminant model in order to keep the assumptions to a minimum. The general two-group discriminant analysis model would follow when we assume multivariate normal distribution
for the consumer-specific variables (XB) and equal covariance matrix for the
two groups of consumers respectively choosing the two brands. See multivar-
iate texts by Anderson [1], Morrison [23] or Press [25] for a discussion of
these. In order to adapt this analysis to the modeling of qualitative re-
sponses, we indeed need additional knowledge on the prior belonging of the
consumer to the groups of buyers or nonbuyers of the brand.

III. Estimation of Parameters for Four Binary Choice Models

Table 2 reviews various methods of estimation appropriate to the case of
single replication (i.e., when y is either 0 or 1) for the four binary choice
models under comparison. It also shows the major problems with the procedure
and properties of estimates. The methods are based on variations of least

squares method or maximum likelihood procedure. Two additional comments may
be in order. First, the maximum likelihood method can be employed for any
probability distribution prespecified for the underlying choice process of
the consumer. Second, we have only covered one method of estimating the
parameters of the discriminant model; for others, see [1, 23, 25].

When the replications are more than one, several other methods could be
employed. One of these [3] involves converting the observed probability into
its logit, i.e., \( \log \left( \frac{P}{1-P} \right) \), expanding it as a Taylor series in terms of
the parameters to be estimated and using least squares method of estimation.
Some modifications to this method are possible in order to improve its accu-
tracy [4, 32]. Empirical comparisons of various methods discussed in this
section as applied to Monte Carlo and real data can be found in [5, 24].
IV. Response Effects of Changes in Independent Variables

We will briefly consider the effect of changes in consumer characteristics on the theoretical probability of choosing the brand according to each model. These measures are useful in forecasting the demand for a brand due to changes in consumer characteristics and also in the development of strategies to influence choice. Additionally, knowledge of the response coefficients could be valuable in testing the accuracy of alternative formulations of the choice process.

Simply stated, these are $\frac{\partial \pi_i}{\partial X_{C_k}}$ where $\pi_i$ is the theoretical probability of choosing one brand (in the binary situation) and $X_{C_k}$ is the kth measured characteristic of the ith consumer. It is computed using the relationship:

$$ \frac{\partial \pi_i}{\partial X_{C_k}} = \frac{\partial \pi_i}{\partial I_i} \cdot \frac{\partial I_i}{\partial X_{C_k}} $$

where $I_i$ is the indicator for the ith consumer. The response coefficients computed using equation (3) are summarized below for each model. Of course, to be correct, one needs to take into account the fact that probability cannot exceed unity for the discriminant model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Response Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discriminant Model</td>
<td>$\gamma_k$</td>
</tr>
<tr>
<td>Linear Probability Model</td>
<td>$(\gamma_k/(b-a))$ if $I_i \in [a,b]$</td>
</tr>
<tr>
<td></td>
<td>0 otherwise</td>
</tr>
<tr>
<td>Multivariate Probit Model</td>
<td>$\phi(I_i) \cdot \gamma_k$ where $\phi(\cdot)$ is the unit normal density function</td>
</tr>
<tr>
<td>Multivariate Logit Model</td>
<td>$\pi_i (1-\pi_i) \gamma_k$ where $\pi_i$ is the theoretical value of probability at $I_i$.</td>
</tr>
</tbody>
</table>
While the value of the response coefficient is uniformly the same for the discriminant model and the linear probability model (except for the end zones), it depends upon the location of the indicator variable for the probit and logit models. In the absence of the knowledge of the true underlying model, it is difficult to choose between these coefficients in practice. Empirical evidence and accuracy of predictive testing are some ways to resolve this issue. In fact, Haberman [14, p. 311] claims:

...that no empirical evidence exists than the normal distribution provides more accurate models than the logistic distribution. Theoretical arguments have been advanced which favor one or the other distribution, but none of them appears convincing, at least to the author.

V. Computational Algorithms

Several computer programs exist for implementing these models. We will briefly describe four of these: (a) Generalized Chi-square Analysis of Categorical data using a weighted least squares program which has the acronym GENCAT [18]; (b) Multiple Logistic Program due to Duncan and Walker [15, 32]; (c) Multivariate dichotomous variable program [24]; and (d) Conditional logit multinomial estimation program called XLOGIT [22, 34]. Our comments on these will be necessarily very brief.

(a) GENCAT Program: This program implements the analysis of multivariate categorical data. It enables estimation of functions to describe observed proportions in terms of several descriptor variables using a weighted least squares method. It also computes several statistics for testing hypotheses on the functional forms of the relationships.

(b) Multiple Logistic Program: This program implements the method developed by Duncan and Walker for estimating the probability of occurrence of an event from dichotomous or polychotomous data. A recursive technique is used
in estimating the multiple logistic risk function in accordance with maximum likelihood methods. The program also computes the linear discriminant function for obtaining initial estimates in the iterative process.

(c) Multivariate Dichotomous Variable Program: This program implements log-linear and logistic models for up to four jointly dependent dichotomous variables using maximum likelihood methods. Its special features include ability to study the bivariate interactions of the exogenous explanatory variables.

(d) XLOGIT Program: This program implements the estimation of the conditional logit multinomial model using maximum likelihood procedures. Estimation is carried out by standard unconstrained maximization procedures. While we have not described the theory of this procedure in this paper, the program can be employed for estimating the binary choice models.

VI. Conclusions

It should be clear from the foregoing discussion that there exists a significant body of knowledge on the qualitative response models and that it pertains almost exclusively to areas other than marketing. Researchers in marketing and consumer behavior could possibly benefit from a close scrutiny of the theory and analysis methods currently available in the literature.

While we have largely concentrated on the binary choice models, theory and estimation methodology extend to the polytomous qualitative variable. Multiple response variables can also be studied in this framework.

Obviously, these models need to be subjected to validation and testing. Opportunities exist for predictive testing using behavioral experimental techniques. The resolution as to which model to use and which method of estimation can only result from extensive application and research on the underlying choice process.
Nevertheless, various applications are possible in marketing and consumer areas. We will briefly touch upon three directions: (i) direct applications of the binary choice models reviewed; (ii) application to the decision processes of one consumer toward a set of brands or concepts; and (iii) study of longitudinal choice behavior.

Direct applications of binary choice models include a study of choice behavior toward brands, services, television programs, shopping centers, stores and the like. Emphasis here would be to fit models to cross-sectional data and estimate response coefficients to changes in characteristics of the population of consumers. Further, future demand can also be estimated. Differences among prespecified segments can be studied by fitting models to samples of consumers in each segment. Another application would be to study the response/nonresponse behavior in survey research.

The general model can be applied to describe the choice process of one consumer toward a set of brands or product concepts. This is the case when \( m=1 \). Such a situation is prototypical of the data collected in concept testing studies using such methods as conjoint measurement. The response here would be "no" or "yes" with respect to buying the brand represented by the concept (or some other criterion). In this case, the model would be \( I=\alpha+\beta X_B \). The model can be fitted to data for each consumer, thereby enabling an examination of individual differences in the response coefficients for changes in the brand attributes. Rao and Winter [28] present an application of this approach to the issue of product design and market segmentation. The general qualitative response models can be employed to extend current approaches to modeling comparative and categorical judgmental data [31].

The methodology can also be used to analyze panel data. The problem here would be to estimate the transition probabilities from one time period to the next using these models and compare them to known stochastic models [2, 19].
FOOTNOTES

1 While canonical correlation is an appropriate method for models with multiple measures, its use has been insignificant owing to difficulty of interpreting results.

2 Other data such as amount bought could also be treated using these models by appropriate discretization.

3 The problem of modeling responses that are either sequentially obtained or ordered in any manner is more complicated and is beyond the scope of this paper.

4 In fact, possibilities exist for combining a discriminant analysis model with logit analysis for purposes of estimation; see [21].

5 These do not include the quadratic programming algorithms applicable to the linear probability model.
TABLE 1

Some Assumptions and Probability of Choice for Four Binary Choice Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Assumed Probability Distribution for Threshold Value</th>
<th>Probability of Choosing Brand 1 for Consumer i</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discriminant Model*</td>
<td>Single point distribution with whole mass at I_i</td>
<td>0 if I_c ≥ I_i</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 if I_c &lt; I_i</td>
</tr>
<tr>
<td>Linear Probability Model</td>
<td>Uniform distribution in the interval (a,b)</td>
<td>Varies linearly with I_i in the interval (a,b)</td>
</tr>
</tbody>
</table>
|                        |                                                     | \[ \begin{cases} 
          0 & \text{if } I_i \leq a \\
          (I_i-a)/(b-a) & \text{if } a < I_i < b \\
          1 & \text{if } I_i \geq b 
\end{cases} \] |
| Multivariate Probit Model | Normal probability distribution; \( \phi(\cdot) \) is the cumulative density function | \( \phi(I_i) \) |
| Multivariate Logit Model | Logistic probability function; \( f(x) = \exp(-x)/\{1+\exp(-x)\}^2 \) | \( (1 + \exp(-I_i))^{-1} \) |

*See text for elaboration.
<table>
<thead>
<tr>
<th>Model</th>
<th>Estimation Method</th>
<th>Major Problems with the Method</th>
<th>Properties of Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Discriminant Model*</td>
<td>Weighted least squares, usually a two-step procedure.</td>
<td>1. Prediction of y could lie out (0,1) interval.</td>
<td>Unbiased, consistent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Extreme values of y predictions could be biased.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Estimates are sensitive to specification error.</td>
<td></td>
</tr>
<tr>
<td>(b) Linear Probability Model</td>
<td>Quadratic programming to minimize squared error subject to inequality constraints (e.g., Dantzig-Cottle Algorithm).</td>
<td>1. Very costly to implement; 2. Extreme value bias exists in prediction.</td>
<td>Consistent; not unbiased, but estimates tend to be distributed tightly about true values.</td>
</tr>
<tr>
<td>(c) Multivariate Probit Model</td>
<td>Maximum likelihood method; involves solution of nonlinear equations using iterative methods (e.g., Newton-Raphson method).</td>
<td>1. Very costly to implement; 2. Need fairly large samples</td>
<td>Consistent, not unbiased, efficient.</td>
</tr>
<tr>
<td>(d) Multivariate Logit Model</td>
<td>Same as for (c)</td>
<td>Same as for (c)</td>
<td>Same as for (c)</td>
</tr>
</tbody>
</table>

*This is not the same method used in standard packages for discriminant analysis.*
REFERENCES


