Econometrica, Vol. 39, No. 5 (September, 1971)

SOME STATISTICAL MODELS FOR LIMITED DEPENDENT VARIABLES WITH APPLICATION TO THE DEMAND FOR DURABLE GOODS[1]

BY JOHN G. CRAGG

Several models for limited dependent variables (variables having a non-negligible probability of exactly equaling zero) are examined. Estimation in and discrimination among the various models are considered, followed by a small sampling experiment into the procedures and an example of their application.

1. INTRODUCTION

THIS PAPER DEVELOPS some models for limited dependent variables.[2] The distinguishing feature of these variables is that the range of values which they may assume has a lower bound and that this lowest value occurs in a fair number of observations. This feature should be taken into account in the statistical analysis of observations on such variables. In particular, it renders invalid use of the usual regression model.

The second section of this paper develops several models for such variables. Like Tobin's [**10**] model, they are extensions of the multiple probit analysis model.[3] They differ from that model by allowing the determination of the size of the variable when it is not zero to depend on different parameters or variables from those determining the probability of its being zero.

Estimation and discrimination in the models are considered in Section 3. The models, like their prototypes, seem particularly intractable to exact analysis and large sample approximations have to be used. The adequacy of inferences based on these procedures is explored in Section 4 through a small sampling experiment.

Limited dependent variables arise naturally in the study of consumer purchases, particularly purchases of durable goods. When a durable good is to be purchased, the amount spent may vary in fine gradations, but for many durables it is probably the case that most consumers in a particular period make no purchase at all. In Section 5 we apply the models to the demand for durable goods to provide an application of the techniques.

2. MODELS FOR LIMITED DEPENDENT VARIABLES

The basic situation being considered is as follows. There is an event which at each observation may or may not occur. If it does occur, associated with it will be a continuous, positive random variable. If it does not occur, this variable has a zero value.[4] A good example is provided by the purchase of a particular type of

[1] Research for this paper was supported by the National Science Foundation and the Canada Council.

[2] The term seems to be due to Tobin [**10**].

[3] Cf. Finney [**6**]. Another type of extension is found in Dagenais [**4**] and [**5**].

[4] Of course, for the variable of interest, a different (known) limit may be relevant, but transformation of the dependent variable would then produce the model discussed.

durable good by an individual in a particular period. Another might be an individual's deciding to move and the distance which he migrates if he does move.

All our models start from the probit analysis model where the probability that a particular event will occur at observation t, $p(E_t)$, is given by

$$p(E_t) = \int_{-\infty}^{X_t'\beta} (2\pi)^{-\frac{1}{2}} \exp\{-z^2/2\}\, dz, \tag{1}$$

where X_t is a $K \times 1$ vector of the values of the independent variables at observation t and β is a vector of coefficients.[5] In what follows, we designate the cumulative unit normal distribution by

$$C(z) = \int_{-\infty}^{z} (2\pi)^{-\frac{1}{2}} \exp\{-t^2/2\}\, dt.$$

Tobin's [**10**] model arises by assuming that a (latent) variable, q_t, is generated by

$$q_t = X_t'\gamma + \varepsilon_t \tag{2}$$

where γ is a vector of coefficients and ε_t is independently and normally distributed with mean zero and variance σ^2. If q_t should be negative, the variable that is actually observed, y_t, is zero. When q_t is positive, y_t is equal to q_t. For example, q_t might be thought of as the desired acquisition of a commodity and y_t as actual acquisition when circumstances prevent negative acquisition. The probability that y_t is zero is

$$f(y_t = 0|X_t) = C(-X_t'\gamma/\sigma) \tag{3}$$

and the density for positive values of y is

$$f(y_t|X_t) = (2\pi)^{-\frac{1}{2}}\sigma^{-1} \exp\{-(y_t - X_t'\gamma)^2/2\sigma^2\}. \tag{4}$$

It will be noted that the probit model is still a valid representation of the occurrence or non-occurrence of the event that y_t is greater than zero. The decisions on whether to acquire and on how much to acquire if acquisition occurs are basically the same in this model, in the sense that the same variables and parameters occur in (3) and (4).

In some situations the decision to acquire and the amount of the acquisition may not be so intimately related. In particular, even when the probability of a non-zero value is less than one half, one might not feel that values close to zero are more probable than ones near some larger value, given that a positive value will occur. This feature cannot be incorporated in the model (3) and (4).

While acquisition may occur only when desired acquisition is, in some sense, positive, there may be factors such as search, information, and transactions costs which inhibit the carrying out of desired plans. In such circumstances, failure for the variable to take on non-zero values may arise either because the desired change is not positive or because other factors inhibit carrying out changes which would be desired in their absence.

[5] The other model used frequently for binary choice, the multiple logit model: $p(E_t) = (1 + \exp\{-X_t\beta\})^{-1}$, may also be generalized with the use of the logistic distribution. Since normal distributions are so widely assumed, we shall concentrate our development on the probit model instead.

We may model this sort of situation in several ways. First, desired acquisition may be represented by (2). If q_t is non-positive, y_t will be zero. In addition, even if q_t is positive, y_t may still be zero because it has been decided not to carry out the adjustment. This aspect might be represented by a probit model. Then the probability that y_t is zero is the sum of the probability that q_t is negative plus the probability that the inhibition will be effective when q_t is positive:

$$f(y_t = 0|X_{1t}, X_{2t}) = C(-X'_{2t}\gamma/\sigma) + C(X'_{2t}\gamma/\sigma)C(-X'_{1t}\beta) \tag{5}$$

where X_{1t} and X_{2t} are vectors of independent variables at observation t (not necessarily distinct) and β and γ are vectors of coefficients. Correspondingly, the density for positive values of y_t is given by

$$f(y_t|X_{1t}, X_{2t}) = (2\pi)^{-\frac{1}{2}}\sigma^{-1} \exp\{-(y_t - X'_{2t}\gamma)^2/2\sigma^2\}C(X'_{1t}\beta). \tag{6}$$

In this model, two hurdles have to be overcome before positive values of y_t are observed. First, a positive amount has to be desired. Second, favorable circumstances have to arise for the positive desire to be carried out.

The position of these hurdles might be reversed. In this case, a decision first has to be made about whether to consider a change or not. Then a decision on the amount of the change is taken. The first decision might be represented by a probit model; the second, by a standard regression model:

$$f(y_t = 0|X_{1t}, X_{2t}) = C(-X'_{1t}\beta) \tag{7}$$

and

$$f(y_t|X_{1t}, X_{2t}) = (2\pi)^{-\frac{1}{2}}\sigma^{-1} \exp\{-(y_t - X'_{2t}\gamma)^2/2\sigma^2\}C(X'_{1t}\beta) \quad \text{for} \quad y_t \neq 0. \tag{8}$$

Equation (8), of course, allows for negative values of y_t. As such, it is appropriate when there are barriers to change, but change can occur in either positive or negative amounts. Non-negativity of y_t might be guaranteed by truncating the distribution at zero. Then (8) would become:

$$f(y_t|X_{1t}, X_{2t}) = (2\pi)^{-\frac{1}{2}}\sigma^{-1} \exp\{-(y_t - X'_{2t}\gamma)^2/2\sigma^2\} \; C(X'_{1t}\beta)/C(X'_{2t}\gamma/\sigma) \quad \text{for} \quad y_t > 0. \tag{9}$$

The truncation in (9) seems rather artificial, though the resulting distribution may not be a bad approximation to the distributions of some economic variables. Tobin's model (3) and (4) is a particular form of (7) and (9) with $X_{1t} = X_{2t}$ and $\beta = \gamma/\sigma$.

An alternative to the truncation in (9) would be to assume that y_t follows a different sort of distribution. For example, it might be presumed that

$$\log y_t = X'_{2t}\gamma + \varepsilon_t \tag{10}$$

where ε_t is normally and independently distributed, given that y_t is non-zero. That is, (8) would become

$$(11) \qquad f(y_t|X_{1t}, X_{2t}) = (y_t)^{-1}(2\pi)^{-\frac{1}{2}}\sigma^{-1} \exp\{-\log y_t - X'_{2t}\gamma)^2/2\sigma^2\}\, C(X'_{1t}\beta) \quad \text{for} \quad y_t > 0.$$

A model of the form of (7) and (11) might be particularly interesting when $X_{1t} = X_{2t}$, $\beta_i = \theta_{\gamma_i}/\sigma$ $(i, \ldots, K-1)$, and $\beta_K = \alpha + \gamma_K/\sigma$, where K refers to the constant term among the coefficients. In this model, the probability of y_t being zero is high when the likely amount (if y_t is positive) is low and conversely, while the two parts vary with the same linear combination of the independent variables. In a sense, this restriction thus produces a model which parallels a main feature of Tobin's model.[6]

Explicitly, the model now becomes

$$(12) \qquad f(y_t = 0|X_{1t}) = C[-(\alpha + \theta X'_{1t}\gamma)]$$

and

$$(13) \qquad f(y_t|X_{1t}) = (y_t)^{-1}(2\pi)^{-\frac{1}{2}}\sigma^{-1} \exp\{-(\log y_t - X'_{1t}\gamma)^2/2\sigma^2\}\, C[\alpha + \theta X'_{1t}\gamma] \quad \text{for} \quad y_t > 0.$$

A different way of dealing with the artificial truncation in (9) would be to assume, as in earlier models, that when q_t is negative in the model (7) and (8), a zero value of y_t will be observed. The probability that y_t will be zero is then given by

$$f(y_t = 0|X_{1t}, X_{2t}) = C(-X'_{1t}\beta) + C(X'_{1t}\beta)C(-X'_{2t}\gamma/\sigma).$$

Since $C(-X'_{1t}\beta) = 1 - C(X'_{1t}\beta)$, this is the same as (5). Thus, it will be noted that it is impossible to identify the sequence of the hurdles in the model (5) and (6).

3. STATISTICAL INFERENCE IN THE MODELS

Taking the separate versions of the models considered in the previous section as distinct, six models for limited dependent variables were considered: Tobin's model (3) and (4), and the five two-part models (5) and (6), (7) and (8), (7) and (9), (7) and (11), and (12) and (13). In each case, the first equation applies when the dependent variable assumes a zero value and the second when a non-zero value occurs. If in a sample of size T, m zero values occur, the likelihood function consists of the product of m terms from the first equation and $T - m$ terms from the second.

The models are computationally burdensome and seem intractable to exact analysis because of the repeated occurrence of integrals of the normal density in the terms of the likelihood function. In this respect, they resemble probit and logit

[6] A related, and possibly more appealing, development may be made using the logit analysis model mentioned in Footnote 5 and the model (10). The log of the odds in favor of a non-zero value is log $[p/(1-p)] = X'_{1t}\beta$. If we impose the constraints that $X_{1t} = X_{2t} = Z_t$ where $Z_{ti} = \log Q_{ti}$, say, and that $\beta_i = \gamma_i$ $(i = 1, \ldots, K-1)$, then the model imposes the constraint that the elasticities with respect to Q_i of the odds in favor of a positive value and of the expected value of y_t when it is positive, are the same. I am indebted to Professor Henri Theil for suggesting this model to me.

analysis. Maximum likelihood estimates may be calculated by iterative methods[7] since no direct way of solving the highly non-linear equations for the first-order conditions is available. The asymptotic variance-covariance matrix of the estimates may be obtained by evaluating the inverse of the negative of the matrix of second partial derivatives of the logarithm of the likelihood function at the maximum likelihood estimates. Similarly, tests of hypotheses within each model may be made by the likelihood-ratio test.

Two considerations may make it desirable to have quick, approximate methods for calculating estimates, especially ones which require only means and covariance matrices of the data. The first consideration is the large amounts of computer time required to maximize the likelihood function when a large number of observations is available, each of which has to be considered in each step of the maximization. The second consideration is that data for which the models are suitable are collected often by agencies which are not allowed to release individual observations but which may provide data in an aggregated form such as covariance matrices. Such problems may in part account for the use of regression analysis with these sorts of data, despite its apparent inappropriateness.

Only those terms in the likelihood function which involve the cumulative normal density require the availability and utilization of the individual observations. One way to simplify the calculations is to take the logarithms of these terms, expand them in Taylor series about $X_t = 0$, and truncate at the second term. This provides a quadratic approximation to the likelihood function which may then be maximized to yield estimates. Since the series expansion does not converge rapidly, the approximation may be poor.

To illustrate, we may consider Tobin's model (3) and (4). We may rearrange the data so that in the first m observations zero values of the dependent variable occur while in the last $T - m$, a positive value is observed. Denote by X_I and X_{II} the corresponding $m \times K$ and $(T - m) \times K$ matrices of observations on the independent variables and by y_I and y_{II} the corresponding vectors of observations on the dependent variable. Let $\phi = \gamma/\sigma$ and $h = 1/\sigma$. Then the logarithm of the likelihood function is

$$\begin{aligned} (14) \qquad L = {} & \sum_{t=1}^{m} \log C(-X_I'\phi) - \tfrac{1}{2}(hy_{II} - X_{II}\phi)'(hy_{II} - X_{II}\phi) \\ & + (T - m)\log h + K_1 \end{aligned}$$

where K_1 is a constant. Expanding the first term about $X_I = 0$,

$$\begin{aligned} (15) \qquad L \doteq {} & -(2/\pi)^{\frac{1}{2}} iX_I\phi - (\pi)^{-1}\phi'X_I'X_I\phi - \tfrac{1}{2}(hy_{II} - X_{II}\phi)'(hy_{II} - X_{II}\phi) \\ & + (T - m)\log h + K_2 = L_1 \end{aligned}$$

where i is a $1 \times m$ unit vector. Differentiating (15) with respect to the parameters

[7] Cf. Crockett and Chernoff [3] and Goldfeld, Quandt, and Trotter [7]. Fortran IV programs are available. Cf. Cragg [2].

and setting the derivatives equal to zero gives

$$(16) \qquad \partial L_1/\partial \phi = -(2/\pi)^{\frac{1}{2}} X_I' i' - (\pi)^{-1} X_I' X_I \phi + h X_{II}' y_{II} - h X_{II}' \phi = 0$$

and

$$(17) \qquad \partial L_1/\partial h = -h y_{II}' y_{II} + y_{II}' X_{II} \phi + (T - m)/h = 0.$$

Equations (16) and (17) are readily solved to give the estimates. A closely related procedure would be to express the independent variables as deviations from their averages, and expand about the means. Doing so requires using a non-zero constant term, and it seems sensible to choose the value which maximizes the likelihood function when the other coefficients are assumed to be zero.

A case of particular interest arises in the model (7) and (11). Proceeding as in (14)–(17), we obtain

$$(18) \qquad \hat{\beta}_1 = (2\pi)^{-\frac{1}{2}} (X_I' X_1)^{-1} (X_{II1}' i' - X_{I1}' i')$$

where

$$X_1 = \begin{bmatrix} X_{I1} \\ X_{II1} \end{bmatrix}.$$

Equation (18) is proportional to the least-squares estimates obtained by regressing a dummy variable, which assumes the values of -1 and 1 as the dependent variable is zero or positive, on the independent variables.

The array of models, among which it may be difficult to choose on a priori grounds, makes it of interest to be able to discriminate among them a posteriori. It is tempting to do so simply on the basis of the maximum values of the logarithms of the likelihood functions. Insight into such a procedure may be obtained from a Bayesian point of view.[8]

Consider L different models, $h_i, i = 1, \ldots, L$, with parameters θ_i in each. Let $p(h_i)$ be the prior probabilities that each model holds and $p(\theta_i|h_i)$ be the prior distribution for the parameters in the ith model ($p(\theta_i|h_j) = 0$ for $i \neq j$). Then the joint posterior distribution of the models and their parameters, given the data Z, is

$$(19) \qquad p(\theta_i, h_i|Z) \propto L(Z|\theta_i, h_i) p(\theta_i|h_i) p(h_i)$$

where L is the likelihood function. Using the usual logarithmic expansion and truncation of the likelihood function about the maximum likelihood estimates, θ_i^0, yields

$$(20) \qquad p(\theta_i, h_i|Z) \dot{\propto} \exp\{g(\theta_i^0|h_i, Z)\} \exp\{-\tfrac{1}{2}(\theta_i - \theta_i^0)' D_i^0 (\theta_i - \theta_i^0)\} = m(\theta_i, h|Z)$$

where g is the logarithm of the likelihood function and

$$(21) \qquad D_i^0 = \{(d_i^0)_{kj}\} = \{-\partial^2 g/\partial\theta_{ik}\partial\theta_{ij} | \theta_i = \theta_i^0\}.$$

Integrating θ_i out of (20) thus yields the (approximate) posterior probability of h_i.

[8] The procedure used is basically the one suggested by Jeffreys [**8**] for testing hypothesis. It is also the approach taken by Thornber [**9**], Box and Hill [**1**], and Zellner and Geisel [**12**].

Should

$$p(\theta_i|h_i)p(h_i) \propto (2\pi)^{-K_i/2}|D_i^0|^{\frac{1}{2}}, \tag{22}$$

where K_i is the number of elements in θ_i, comparison of the models under this procedure would be essentially the same as comparison of the logarithms of the likelihood functions at their maximum values. Equation (22) is, however, an unlikely and, indeed, unacceptable representation of prior knowledge since it embodies knowledge of the terms of the likelihood function at its maximum, and it has the unfortunate property that a wider version of a model would always be more probable than a narrow one. Instead we shall use priors of the form $p(\theta_i|h_i)p(h_i) \propto C_i$ a constant, so that based on (20),

$$p(h_i|Z) \dot{\propto} (2\pi)^{K_i/2}|D_i^0|^{-\frac{1}{2}} \exp\{g(\theta_i^0|h_i, Z)\}C_i. \tag{23}$$

Needless to say, if prior information about θ_i is available, it should be used.

4. A SAMPLING EXPERIMENT

The estimation and discrimination procedures outlined seem intractable to exact analysis. A sampling experiment was conducted to investigate the properties of the maximum-likelihood estimates and the suggested method of discrimination. The experiment examined all the models except the one which allows negative values for the dependent variable, (7) and (8), and the restricted log-normal one, (12) and (13).

The experiment consisted of drawing one hundred samples of observations from each of the models. Each sample contained 500 observations. It seemed more relevant to get a feel for the performances of the estimates by using samples with different independent variables and various true values for the parameters than to concentrate on a particular set for each. As a result, the parameters were drawn at random from a normal population. Needless to say, the same values of the parameters were used to generate each observation in any particular sample. This does, of course, mean that we are not studying the usual sampling distribution.

There was only one independent variable in the models and a constant term. The independent variable was log-normally distributed with mean zero and unit variance.[9] The population means and standard deviations of the parameters were:

Parameter	β_1	β_2	γ_1	γ_2	$\log\sigma$
Mean	$-.40$	.20	$-.30$	.30	1.0 .
Standard deviation	.10	.10	.10	.10	.10

The subscript 1 refers to the constant term. All parameters were independent of each other. This selection of the parameters and variables gave considerable diversity to the types of observations that were employed.[10]

[9] In generating and estimating the model (7) and (10), the logarithms of the (positive) values of the dependent variable were linearly associated with the logarithms of the independent variable for the part of the model in (10).

[10] More details are available in a mimeographed form of this paper available as Discussion Paper 8, Department of Economics, University of British Columbia.

Two possible dangers arise in calculating the estimates.[11] The first comes in the probit analysis model (1) which is embedded in the model (7) and (9) and in the model (7) and (10). If there should exist a linear combination of the independent variables such that this combination is always positive when the dependent variable is positive and negative when the dependent variable is zero, the probit analysis model gives a perfect fit to the data and the maximum likelihood estimates of the coefficients are not finite. This problem also arises in the double-hurdle model (5) and (6), in connection with the parameters β. For any set of independent variables and finite number of observations, there is always a positive probability that a sample of observations on the dependent variable will occur which gives rise to this difficulty. (This consideration makes dubious the appeal to the asymptotic normality of the estimates in actual empirical work.) It was decided that, should this problem arise in the experiments, the sample involved would be deleted from the study and a record kept of the number of such occurrences. In fact, it did not occur.

The second danger also arises in the double-hurdle model, again in connection with β. It will be noted that this model becomes the same as Tobin's model, (3) and (4), when the constant term approaches infinity. Stated differently, there is a possibility that the best fit in this model will be obtained when the probability of overcoming the hurdle represented by the part of the model involving β is unity. Even if this extreme case is not met, there is a danger of inaccurate estimates occurring if the probability is always very high, since the calculation of the cumulative normal distribution and the normal density at extreme values is not particularly accurate and rounding error may become severe. In view of this, it was decided to delete from the experiment any sample in which it was calculated that the smallest probability of overcoming this hurdle was greater than .9999, or where for less than two per cent of the observations it was calculated that this probability was less than .99999. This situation arose several times in the course of the experiment: thirty-one times using (3) and (4) to generate the data; no times with (5) and (6); twice with (7) and (9); and no times with (7) and (10). Thus in almost one quarter of the instances where trouble from this source might be expected it was encountered. The two cases using (7) and (9) which gave rise to the problem were instances where the parameters in the two parts of the model were almost equal so that the Tobin model (3) and (4) fitted the data well.

The experiment investigated three questions. First, did the asymptotic approximate distribution of the estimates seem to be adequate? Second, how adequate were the approximate estimates based only on the means and variance-covariance matrices of the observations? Third, how well did the Bayesian procedures succeed in discriminating among the models?

Under the hypothesis that the maximum likelihood estimates, $\hat{\theta}$, are approximately normally distributed with mean equal to the population value, θ, and variance-covariance matrix $V = [D^0]^{-1}$ (where D^0 is defined in (21)), the marginal distribution of the ratio $(\hat{\theta}_i - \theta_i)/v_{ii}^{\frac{1}{2}}$ would be normal with mean zero and unit variance. Table I gives the means and standard deviations of the values of these

[11] Computations were performed on an IBM 360/67 computer in double precision.

TABLE I

SUMMARY OF ESTIMATES OF $(\hat{\theta}_i - \theta_i)/v_{ii}^{\frac{1}{2}}$

I. Model (3) and (4)				
Parameter	γ_1			γ_2
Mean	$-.073$			.204
Standard deviation	.913			.970
Kolmogorov-Smirnov test statistic	.117			.130[a]
II. Model (5) and (6)				
Parameter	β_1	β_2	γ_1	γ_2
Mean	$-.392$	.120	.151	.302
Standard deviation	1.068	.932	1.291	.930
Kolmogorov-Smirnov test statistic	.146[b]	.098	.101	.206[c]
III. Model (7) and (9)				
Parameter	β_1	β_2	γ_1	γ_2
Mean	$-.249$	.193	.177	.053
Standard deviation	.956	.987	1.058	1.024
Kolmogorov-Smirnov test statistic	.116	.101	.906	.080
IV. Model (7) and (10)				
Parameter	β_1	β_2	γ_1	γ_2
Mean	$-.064$	.057	.051	$-.014$
Standard deviation	1.059	1.047	.926	.926
Kolmogorov-Smirnov test statistic	.084	.096	.067	.076

[a] Significant at the .10 level.
[b] Significant at the .05 level.
[c] Significant at the .01 level.

ratios for each of the coefficients based on estimates made from the data generated by the corresponding model. Needless to say, these ratios are not independent among the various coefficients in any particular model so that the different columns do not contain separate information about the distributions. Also reported in the table are the results of using the Kolmogorov-Smirnov test of the hypothesis that the estimates did have a unit normal distribution. Only in three cases were these statistics significant even at the .10 level. Two of these, and the most significant ones, occurred in the double-hurdle model (5) and (6). The other one occurred in the Tobin model (3) and (4) and may be the result of dropping observations when the double-hurdle model ran into trouble. When these observations are included, the means for γ_1 and γ_2 are $-.069$ and .124 while the Kolmogorov-Smirnov test statistics drop to .096 and .102 respectively. In all cases, the problem seemed to be bias rather than dispersion.

Estimates based on the approximation discussed in Section 3 were calculated for all coefficients other than the regression parameters in (10). In making the calculation, the likelihood function was expanded about the average of the independent variables, an initial constant term being based on the model with only a constant. This procedure was not used in the truncated-normal model for γ—where it can lead to a complex solution for the variance—and in that model the

TABLE II

PERFORMANCES OF APPROXIMATIONS[a]

(Means of Various Measures)

I. $(|\tilde{\theta}_i - \theta_i| - |\hat{\theta}_i - \theta_i|)/\theta_i$

Model \ Parameter	β_1	β_2	γ_1	γ_2
(3) and (4)	—	—	.260	−.001
(5) and (6)	6.235	.883	2.131	.149
(7) and (9)	.073	.107	3.791	.813
(7) and (10)	.088	.119	—	—

II. $\log (|\tilde{\theta}_i - \theta_i|)/(|\hat{\theta}_i - \theta_i|)$

Model \ Parameter	β_1	β_2	γ_1	γ_2
(3) and (4)	—	—	1.072	.095
(5) and (6)	2.346	1.296	1.442	.472
(7) and (9)	.428	.714	1.795	2.492
(7) and (10)	.426	.761	—	—

III. $(|\tilde{\theta}_i - \theta_i| - |\hat{\theta}_i - \theta_i|)/v_{ii}^{\frac{1}{2}}$

Model \ Parameter	β_1	β_2	γ_1	γ_2
(3) and (4)	—	—	1.053	−.004
(5) and (6)	8.610	2.020	2.044	.823
(7) and (9)	.368	.745	4.056	8.234
(7) and (10)	.405	.826	—	—

[a] $\tilde{\theta}_i$: approximation; $\hat{\theta}_i$: maximum-likelihood estimate; θ_i: true value; $v_{ii}^{\frac{1}{2}}$: standard error of $\hat{\theta}_i$.

expansion was about the origin. The approximations, $\tilde{\theta}_i$, were compared with the maximum-likelihood estimates, $\hat{\theta}_i$, in three ways: first by examining $(|\tilde{\theta}_i - \theta_i| - |\hat{\theta}_i - \theta_i|)/\theta_i$; second by evaluating $\log [(|\tilde{\theta}_i - \theta_i|)/(|\hat{\theta}_i - \theta_i|)]$; and third by calculating $(|\tilde{\theta}_i - \theta_i| - |\hat{\theta}_i - \theta_i|)/v_{ii}^{\frac{1}{2}}$.

Averages of these measures are shown in Table II.

The approximations fare rather poorly in comparison with the maximum likelihood estimates. Only for γ_2 in the Tobin model were they about as accurate as the maximum likelihood estimates. In other cases they were decidedly poorer, the worst cases arising in the more complicated models (5) and (6) and for the γ coefficients in (9).

Three prior densities were tried in investigating the discrimination procedure:

$$(24) \qquad p(\theta_i|h_i) \propto (2\pi)^{-K_i/2}$$

(where K_i is the number of parameters in model i);

$$(25) \qquad p(\beta_1, \beta_2) \propto (2\pi)^{-1} \left[\sum_{t=1}^{T} (x_t - \bar{x})^2\right]^{\frac{1}{2}}$$

and

$$p(\gamma_1, \gamma_2, \sigma) \propto (2\pi)^{-3/2}\left[\sum_{t=1}^{T_\gamma} (x_t^\gamma - \bar{x}^\gamma)^2\right]^{\frac{1}{2}};$$

(26) $$p(\beta_1, \beta_2) \propto (2\pi)^{-1}\left[\sum_{t=1}^{T} (x_t - \bar{x})^2/(T-1)\right]^{\frac{1}{2}}$$

and

$$p(\gamma_1, \gamma_2, \sigma) \propto (2\pi)^{-3/2}\left[\sum_{t=1}^{T_\gamma} (x_t^\gamma - \bar{x}^\gamma)^2/(T_\gamma - 1)\right]^{\frac{1}{2}}$$

where x_t^γ are the values of the independent variables used in calculating γ—the whole sample for models (3) and (4), and (5) and (6), and the values when positive values of the dependent variable occurred in models (7) and (9), and (7) and (10).

A summary of the results is found in Table III. Each row of that table records the number of samples in which the model that actually generated the data was judged more probable than each of the others. It will be noted that discrimination between the model (7) and (10) and the others was the most clear cut. The greatest sensitivity to the prior distribution occurred with the comparison of Tobin's model (3) and (4) with the models (5) and (6), and (7) and (9). This was especially

TABLE III

DISCRIMINATION PROCEDURE

(Number of times true model judged more probable a posteriori than other models)

		I. Prior Distribution (24)		
Model for Comparison	(3) and (4)	(5) and (6)	(7) and (9)	(7) and (10)
True model				
(3) and (4)	—	81	100	100
(5) and (6)	67	—	80	93
(7) and (9)	52	44	—	99
(7) and (10)	100	100	100	—
		II. Prior Distribution (25)		
Model for Comparison	(3) and (4)	(5) and (6)	(7) and (9)	(7) and (10)
True model				
(3) and (4)	—	0	58	100
(5) and (6)	100	—	87	92
(7) and (9)	84	43	—	99
(7) and (10)	100	100	100	—
		III. Prior Distribution (26)		
Model for Comparison	(3) and (4)	(5) and (6)	(7) and (9)	(7) and (10)
True model				
(3) and (4)	—	73	99	100
(5) and (6)	73	—	60	90
(7) and (9)	58	49	—	99
(7) and (10)	100	100	100	—

marked when the Tobin model generated the data and was compared with (5) and (6), the correct model usually being found more probable when (24) and (26) were used and never being more probable with (25). The one case where the incorrect model was judged more probable than the correct one more than half the time occurred when (7) and (9) generated the data and was compared[12] with (5) and (6). The root of the problem may lie in one of three places. First, the prior densities apparently give too much weight to (5) and (6). It is worth noting that the maximum likelihood in the model (7) and (9) was greater than that in (5) and (6)—83 times when it generated the data while it was greater 48 times when (5) and (6) generated the data. Second, since the approximation used is asymptotic, 500 observations may not render it valid. In addition, in estimating the second part of (7) and (9), only a fraction of the observations is used. Third, the result may be caused by rounding error, since the logarithms of the numbers being compared are much larger in absolute value than their difference. The extent to which the wrong model was judged more probable was often not great and, as Table IV illustrates, the average of the logarithms of the odds in favor of the correct model was positive with prior distributions (26). This was also the case, though the magnitude was smaller, when (24) or (25) was used. A feature of the results revealed by Table IV is the fairly modest magnitudes of the posterior odds in the comparisons of the first three models with each other.

To investigate whether the incorrect discrimination would tend to be corrected with a larger sample, the sampling experiment, in limited form, was also conducted using (7) and (9) to generate the data and with one thousand rather than five hundred observations. The problem did indeed seem to right itself somewhat. Model (7) and (9) were judged more probable than (5) and (6) in 56, 54, and 58 of the observations using (24), (25), and (26) respectively.[13] Thus it seems that reliable discrimination in this model does require a large sample size.

5. AN APPLICATION TO PURCHASES OF CONSUMER GOODS

The models developed in Section 2 were applied to the purchase of consumer durable goods to provide an example of their performances. The data used were from the 1964 Survey of Consumer Finances of the Survey Research Center of the University of Michigan. This body of data contains a wealth of information on the position of consumers, though it does not contain information on the existing stock of consumer durables. Although in deterministic models it is possible to get around this problem,[14] it is a serious drawback in stochastic models.

The model used is a simplified version of one developed by Wu [**11**]. The simplification consisted mainly in reducing the number of explanatory variables. The independent variables were:

[12] It is worth noting that when the Tobin model generated the data, (5) and (6) was judged more probable than (7) and (9) in 99 per cent of the samples using any of the three prior densities.

[13] The correct model was more probable than (3) and (4) in 59, 89, and 60 of the observations and always more probable than (7) and (10).

[14] Cf. Wu [**11**].

TABLE IV

SUMMARY OF POSTERIOR DISTRIBUTION USING (26)
(Logarithms of odds in favor of true model)

I. True Model (3) and (4)

Alternative model	(5) and (6)	(7) and (9)	(7) and (10)
Average	.58	2.73	44.40
Lowest	−4.95	−.42	7.77
First quartile	.14	2.31	30.79
Median	1.03	2.99	42.35
Third quartile	1.59	3.36	56.55
Highest	2.12	4.00	96.68

II. True Model (5) and (6)

Alternative model	(3) and (4)	(7) and (9)	(7) and (10)
Average	2.82	3.21	15.60
Lowest	−1.83	−1.57	−8.40
First quartile	−.14	−.29	5.91
Median	2.17	.35	12.39
Third quartile	5.12	.78	21.13
Highest	19.65	5.00	54.40

III. True Model (7) and (9)

Alternative model	(3) and (4)	(5) and (6)	(7) and (10)
Average	3.71	.61	30.27
Lowest	−3.71	−5.61	−5.25
First quartile	−2.14	−.97	19.14
Median	1.03	−.02	29.78
Third quartile	6.71	1.97	41.59
Highest	42.35	10.21	70.19

IV. True Model (7) and (10)

Alternative model	(3) and (4)	(5) and (6)	(7) and (9)
Average	123.3	71.9	72.4
Lowest	61.1	39.1	38.9
First quartile	109.1	61.0	60.9
Median	123.1	70.8	71.5
Third quartile	136.7	85.1	84.6
Highest	171.5	105.8	133.0

Y_t: disposable income (\$1,000);
Y_{t-1}: income in previous year (\$1,000);
LY_t: log Y_t;
LY_{t-1}: log Y_{t-1};
A_t: age of head of spending unit (years);
N_t: number of children present in spending unit;
H_{1t}: 1 if owns home, 0 otherwise;
H_{2t}: 1 if moved into house in the last year, 0 otherwise;
M_{1t}: 1 if married, 0 otherwise;
M_{2t}: 1 if became married in last year, 0 otherwise; and
M_{3t}: 1 if became married in preceding year, 0 otherwise.

The dependent variable was total outlay on purchases of durable goods net of any trade-ins during the year prior to the survey or the logarithm of this variable. The survey contained information for 1540 households, but data on all the variables were available from only 1445. Among this latter group, 598 made some purchases of durable goods.

Table V records the maximum likihood estimates of the parameters in the models and some summary statistics. The posterior probabilities found in the last line are based on the prior distribution:

$$p(\beta) \propto |X'X/T|^{\frac{1}{2}}(2\pi)^{-K/2},$$

$$p(\gamma) \propto |X'X/T|^{\frac{1}{2}}(2\pi)^{-K/2},$$

$$p(\sigma) \propto (2\pi)^{-\frac{1}{2}}/\sigma,$$

$$p(\alpha) \propto (2\pi)^{-\frac{1}{2}},$$

$$p(\theta) \propto (2\pi)^{-\frac{1}{2}},$$

where K is the number of coefficients.

As judged by the posterior probabilities or by the values of the logarithms of the likelihood functions at their maxima, the best model appears to be the one using equations (7) and (10). This is the model which combines a probit analysis model for the purchase of durables with a regression model for the logarithms of the (positive) amounts spent. This model is followed by the one given in (11) and (12), which restricts the coefficients in the two parts of (7) and (10) to be proportional. It should be pointed out, however, that the specification of these models does not allow negative or zero values for income and so is not really fully adequate.[15]

Among the remaining models, (7) and (9) appears to be the best. It is followed by the double-hurdle model, (5) and (6); and Tobin's model, (3) and (4), seems to fit these data most poorly. However, there is reason to doubt the accuracy of the estimates for the double-hurdle model. The large sizes of the constant term and the coefficients for the dummy variables suggest that we may have encountered the danger mentioned earlier which can arise when this model attempts to approximate the Tobin model. Indeed, these coefficients do mean that the probability of surpassing the first hurdle is always virtually unity so that there is great danger of computational error dominating the results. This problem, of course, does not apply to the other two-decision model, (7) and (9). Thus, the hypothesis that the different variables have the same relative influences in the two parts of the models is supported neither in the models for the amount spent nor in those for the logarithms of these quantities. It is worth mentioning that all these models are clearly better than the corresponding versions which restrict all coefficients except the constants to be zero—whether the likelihood ratio test or the posterior probabilities is used in making this judgment.

The estimates of the coefficients seem quite reasonable, and are similar in nature to those found in Wu [**11**], as well as those given by the least-squares equations.

[15] This problem was solved here by simply dropping such observations. There were sixteen cases of non-positive incomes in the data. For purposes of comparability of the models, these observations were also deleted in fitting the models using income rather than its logarithm.

TABLE V

MODELS FOR THE PURCHASE OF CONSUMER DURABLES—MAXIMUM LIKELIHOOD ESTIMATES[a]

Model	(3) and (4)	(5) and (6) estimates of β	(5) and (6) estimates of γ	(2)–a estimates of β	(2)–b estimates of β	(7) and (9) estimates of γ	(7) and (10) estimates of γ	(11) and (12)	Regression dependent 0–1	Regression for amount
Variable										
Y_t	49.56	−.0721	53.27	.7562		122.81			.2721	22.41
	(11.97)	(.1195)	(11.81)	(.2299)		(54.79)			(.0735)	(9.64)
LY_t					.4161		.3483	.4208		
					(.1114)		(.1180)	(.0915)		
Y_{t-1}	−2.57	−.2930	6.57	−.1661		50.58			−.0688	10.86
	(9.83)	(.1586)	(9.82)	(.1879)		(44.74)			(.0624)	(7.94)
LY_{t-1}					−.0834		.1487	.0195		
					(.0986)		(.1038)	(.0793)		
A_t	−7.30	.0196	−6.76	−.0142	−.0124	−15.14	−.0028	−.0094	−.0047	−.82
	(1.57)	(.0329)	(1.55)	(.0028)	(.0029)	(10.82)	(.0031)	(.0024)	(.0010)	(1.47)
N_t	7.59	2.7838	12.21	.0627	.0637	−132.18	−.0251	.0239	.0240	−14.68
	(13.62)	(1.4874)	(13.44)	(.0257)	(.0256)	(86.01)	(.0241)	(.0199)	(.0091)	(11.36)
H_{1t}	126.31	4.5907	78.59	.2028	.1976	300.98	.0067	.1395	.0712	42.87
	(43.53)	(2.4011)	(46.84)	(.0843)	(.0846)	(290.26)	(.0902)	(.0687)	(.0286)	(42.34)
H_{2t}	293.36	7.1276	280.27	.2511	.2693	1698.85	.4263	.3767	.0842	277.69
	(53.80)	(16.4589)	(52.95)	(.1008)	(.1013)	(307.34)	(.0979)	(.0790)	(.0351)	(46.17)
M_{1t}	112.35	−7.1793	105.98	.1690	.1231	760.97	.0433	.1061	.0519	42.88
	(53.80)	(15.0064)	(52.68)	(.0939)	(.0965)	(435.07)	(.1128)	(.0792)	(.0315)	(51.65)
M_{2t}	552.98	3.1757	536.91	.6174	.6264	868.34	.5224	.6734	.2357	320.36
	(124.69)	(20.8213)	(121.97)	(.2657)	(.2652)	(450.70)	(.2068)	(.1945)	(.0864)	(95.20)
M_{3t}	289.24	−.2531	286.86	.7727	.7826	81.48	.1754	.4646	.2882	21.65
	(92.10)	(2.4216)	(90.87)	(.1983)	(.1997)	(464.95)	(.1521)	(.1450)	(.0649)	(70.95)
Constant	−373.18	9.1743	−414.86	−.3250	−2.8478	−3787.92	1.3485	−3.8127	.3795	121.01
	(98.78)	(15.2024)	(98.29)	(.1783)	(.5123)	(884.79)	(.5462)	(.4221)	(.6114)	(90.51)
σ	613.46		599.29			1067.4	.7790	.8978	.4572	418.51
	(19.38)		(18.97)			(146.2)	(n.c.)	(.0294)	(n.c.)	(n.c.)
α								6.3398		
								(.1951)		
θ								1.1117		
								(.1416)		
Log of likelihood function at maximum	−5125.	−5107.		−876	−877.	−4190.	−4167.	−5060.		
Log of likelihood function at maximum for model	−5125.	−5107.				−5066.[b]	−5044.[c]	−5060.		
Log of posterior probability (+ constant)	−5142.4	−5126.1				−5107.3[b]	−5074.9[c]	−5076.7		

[a] Standard errors are in parentheses.
[b] Includes estimates from (2)-a.
[c] Includes estimates from (2)-b.

Current income plays a strong role in both parts of the models. Lagged income is much weaker, being not significant, and enters the two parts of the models with different signs. The variables categorizing the households tend to have seemingly sensible coefficients. They play a larger role in explaining the decision to make durable goods purchases than in accounting for the amount spent. The large effects ascribed to recent marriage and to recently moving into a house are worth noting. These variables might be interpreted as being proxies for differences between desired and actual stock of consumer goods.

The results obtained in this section, together with those of Section 4, are quite promising for the usefulness of the models developed. The particular outcome regarding which model is best is, of course, specific to the data and specifications used. It should be stressed that the dependent variable—being the sum of all durable goods purchases rather than the purchases of a particular durable—is probably too aggregated to fit the decision-making processes envisaged in the formulation of the models. The specification was largely determined by the availability of observations. It is hoped, however, that this exercise will suggest the possible usefulness of the approach and formulations developed in this paper and may lead to more fruitful applications.

University of British Columbia

Manuscript received May, 1969; revision received January, 1970.

REFERENCES

[1] Box, G. E. P., and W. J. Hill: "Discrimination Among Mechanistic Models," *Technometrics*, 9 (1967), 57–72.
[2] Cragg, John G.: "Programs for Multiple Probit and Logit Analysis and Extensions to Them," mimeographed, 1968.
[3] Crockett, Jean B., and Herman Chernoff: "Gradient Methods of Maximization," *Pacific Journal of Mathematics*, 5 (1955), 33–59.
[4] Dagenais, M. G.: "Note sur le Traitement de Certains Problems de Discontinuité dans la Régression Multiple," *Bulletin of the International Statistical Institute*, Proceedings of the 34th Session, XL, 2 (1963), 767–9.
[5] ———: "Modèle de Régression Multiple Comportant l'Estimation de Senils de Réaction," *Bulletin of the International Statistical Institute*, XLI, 1 (1965), 385–8.
[6] Finney, D. J.: *Probit Analysis*. Cambridge: Cambridge University Press, 1952.
[7] Goldfield, Stephen M., Richard E. Quandt, and Hale F. Trotter: "Maximization by Quadratic Hill Climbing," *Econometrica*, 34 (1966), 541–51.
[8] Jeffreys, H.: *Theory of Probability*, Third edition. Oxford: Clarendon Press, 1961.
[9] Thornber, H.: "Applications of Decision Theory to Econometrics," unpublished doctoral dissertation, University of Chicago, 1966.
[10] Tobin, James: "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26 (1958), 24–36.
[11] Wu, De-Min: "An Empirical Analysis of Household Durable Goods Expenditures," *Econometrica*, 33 (1965), 761–80.
[12] Zellner, Arnold, and Martin S. Geisel: "Analysis of Distributed Lag Models with Application to Consumption Function Estimation," mimeographed, 1968.