

## A MAXIMUM LIKELIHOOD METHOD FOR LATENT CLASS REGRESSION INVOLVING A CENSORED DEPENDENT VARIABLE

KAMEL JEDIDI

MARKETING DEPARTMENT  
GRADUATE SCHOOL OF BUSINESS  
COLUMBIA UNIVERSITY

VENKATRAM RAMASWAMY

MARKETING DEPARTMENT  
SCHOOL OF BUSINESS ADMINISTRATION  
UNIVERSITY OF MICHIGAN

WAYNE S. DESARBO

MARKETING AND STATISTICS DEPARTMENTS  
SCHOOL OF BUSINESS ADMINISTRATION  
UNIVERSITY OF MICHIGAN

The standard tobit or censored regression model is typically utilized for regression analysis when the dependent variable is censored. This model is generalized by developing a conditional mixture, maximum likelihood method for latent class censored regression. The proposed method simultaneously estimates separate regression functions and subject membership in  $K$  latent classes or groups given a censored dependent variable for a cross-section of subjects. Maximum likelihood estimates are obtained using an EM algorithm. The proposed method is illustrated via a consumer psychology application.

Key words: censored regression, latent class analysis, maximum likelihood estimation, consumer psychology.

### 1. Introduction

Tobit models are part of a general class of models for analyzing truncated and censored data where the range of the dependent variable is constrained (see Amemiya, 1984, for a survey). A typical case of censoring occurs when the dependent variable has a number of its values concentrated at a limiting value, say zero. For instance, in a large-scale study of the number of hours worked by married women (Greene & Quester, 1982), about 66% of the over 10,000 wives surveyed reported zero hours. This makes the use of the classical regression model inappropriate, as to be discussed below.

The standard tobit or censored regression model (Tobin, 1958) in which the dependent variable is censored (at zero, without loss of generality) can be expressed as

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + u_i, \quad i = 1, \dots, I \text{ subjects}, \quad (1)$$

where the random variable  $y_i^*$  may be viewed as an index or a partially latent variable whose observed value,  $y_i$ , is concentrated at zero when it is nonpositive. Hence,

We thank the Editor and referees for their insightful comments and suggestions. Requests for reprints should be sent to Kamel Jedidi, Marketing Department, Graduate School of Business, Columbia University, New York, NY 10027.

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0, \\ 0 & \text{if } y_i^* \leq 0, \end{cases}$$

where  $y_i$  is the value of the observed censored dependent variable for subject  $i$ . In (1),  $\mathbf{x}'_i$  is the  $i$ -th row vector of  $\mathbf{x} = ((x_{ij}))$  containing the values of  $J$  explanatory variables ( $j = 1, \dots, J$ ) for subject  $i$ . The regression coefficients for these explanatory variables are contained in  $\boldsymbol{\beta} = \{\beta_j\}$ , a  $J \times 1$  column vector, and the error terms  $\{u_i\}$  are assumed to be iid drawings from a normal distribution,  $N(0, \sigma^2)$ . Note that  $y_i$  and  $\mathbf{x}_i$  are known for each of the  $I$  subjects ( $i = 1, \dots, I$ ), but  $y_i^*$  is unobserved if it is nonpositive (i.e.,  $y_i = 0$ ) and is therefore partially latent.

The expected value of  $y_i = E(y_i) = \text{Prob}(y_i^* > 0) \cdot E(y_i | y_i^* > 0)$ , and the conditional expectation is given by (see Amemiya, 1985, p. 367)

$$\begin{aligned} E(y_i | y_i^* > 0) &= \mathbf{x}'_i \boldsymbol{\beta} + E(u_i | u_i > -\mathbf{x}'_i \boldsymbol{\beta}) \\ &= \mathbf{x}'_i \boldsymbol{\beta} + \sigma \frac{\left( \frac{\phi(\mathbf{x}'_i \boldsymbol{\beta})}{\sigma} \right)}{\left( \frac{\Phi(\mathbf{x}'_i \boldsymbol{\beta})}{\sigma} \right)}, \end{aligned} \quad (2)$$

where  $\phi(\cdot)$  and  $\Phi(\cdot)$  are the density and distribution functions, respectively, of a standard normal variable. Note that (2) implies that the classical regression estimator  $\hat{\boldsymbol{\beta}}$  is biased and inconsistent regardless of whether one uses all the observations or just the positive observations. Thus, the classical regression model is inappropriate when the dependent variable is censored.

Numerous applications of standard tobit models for censored data have appeared in the social science literature, in part due to the increasing availability of computational resources as well as micro-level survey and panel data. These include, inter alia, psychological, economic, and social research studies concerning family relations and attachment behavior (Fisher & Tesler, 1986), number of extramarital affairs (Fair, 1978), job absenteeism (Baba, 1990), personal wealth transfers (Adams, 1980), ratio of unemployed hours to employed hours (Ashenfelter & Ham, 1979), household purchases of durable goods (Tobin, 1958), number of credit card accounts possessed by consumers (Kinsey, 1981), number of criminal arrests (Witte, 1980), performance on achievement tests (Gross, 1980), and household purchases of grocery products (Elrod & Winer, 1982; Tellis, 1988).

In most of these applications, a single set of coefficients  $\boldsymbol{\beta}$  is estimated from the censored data. While this may be justified if one is only interested in aggregate-level estimates, it may be inadequate and potentially misleading if there is considerable heterogeneity in subjects' responses. Consider the study of the number of hours worked by married women (Greene & Quester, 1982) that investigates the impact of explanatory variables such as second marriage, divorce probability, presence of small kids, as well as education and wage differences between husband and wife. Aspects of these tobit analyses are reproduced in Table 1 (see Greene, 1990, pp. 728–729).

Separate tobit estimates are reported for black versus white wives. The coefficients in Table 1 suggest differences in the impact of the explanatory variables on the average number of hours worked. For instance, second marriage has a much larger positive impact for white wives than black wives, while education difference has a positive effect for black wives and a negative effect for white wives. Hence, if differences in responses exist across subjects, disaggregate analyses are necessary to reveal such differential effects. However, *a priori* bases for performing such disaggregate analyses

TABLE 1  
Tobit Coefficients For Labor Study of Number of Hours Worked  
(Adapted from Greene, 1990)

Variable	White Wives	Black Wives
Second marriage	175.85 (3.47)	25.33 (0.41)
Mean divorce probability	417.39 (6.52)	481.02 (5.28)
Education difference	-48.08 (-4.77)	22.59 (1.96)
Relative wage	312.07 (5.71)	286.39 (3.32)
High divorce probability	670.22 (8.40)	578.66 (5.33)
Small kids	-1324.84 (-19.78)	-824.19 (-10.14)
Intercept	-1803.13 (-8.64)	-2753.87 (-9.68)
$\hat{\sigma}$	1559.00	1511.00
Sample size	7459.00	2798.00
% reporting zero hours	71%	54%

Note: values in parentheses are asymptotic t-statistics of the respective coefficients.

may not always be obvious or known to the researcher, and even if these are known, the relevant data may be unavailable. Further, typically collected background data such as demographics may not usefully describe such response differences in other instances (see Moore, 1980).

This paper presents the development of a maximum likelihood method for conducting disaggregate tobit analyses that does not require a priori bases to be prespecified. Alternative approaches for modeling structural heterogeneity such as switching regression models (see Maddala, 1983, for a review) often entail the specification of a switching mechanism and/or require exogenous sample separation information. In ad-

dition, the "latent" switching regression models (see Quandt & Ramsey, 1978) typically assume that subjects switch among different regimes in a discrete manner, and more importantly, do not accommodate censored data. Our proposed method allows for probabilistic membership of subjects into  $K$  latent classes, given a censored dependent variable,  $y$ , and explanatory variables,  $x$ , for a sample of  $I$  subjects. The latent class membership probabilities for the  $I$  subjects and the class-specific regression parameters are estimated *simultaneously*.

The next section presents the technical structure of the model. We assume that the latent random variable  $y_i^*$  is distributed as a mixture of univariate conditional normal densities with the observations being censored if  $y_i^* \leq 0$  (i.e., the observed value of the dependent variable,  $y_i$ , is zero if  $y_i^* \leq 0$ ). An estimation algorithm based on an EM framework (Dempster, Laird, & Rubin, 1977) is devised for estimating the model's parameters in a maximum likelihood fashion. (We note by way of background that the classical regression latent class model proposed by DeSarbo & Cron, 1988, is a special case of the proposed method when the dependent variable is uncensored for all subjects). To illustrate the proposed method, we provide a consumer psychology application with censored data concerning yogurt consumption.

## 2. Method

### The Model

We assume that the probability density function for the latent random variable,  $y_i^*$ , is distributed as a finite mixture of conditional univariate normal densities,  $f(\cdot)$ :

$$y_i^* \sim \sum_{k=1}^K \lambda_k f(y_i^* | \mathbf{x}_i, \sigma_k^2, \boldsymbol{\beta}_k) = \sum_{k=1}^K \lambda_k \frac{1}{\sigma_k} \phi\left(\frac{y_i^* - \mathbf{x}_i' \boldsymbol{\beta}_k}{\sigma_k}\right), \quad (3)$$

where

- $k = 1, \dots, K$  latent classes;
- $\boldsymbol{\beta}_k$  = the vector of regression coefficients  $\{\beta_{jk}\}$  for  $J$  explanatory variables ( $j = 1, \dots, J$ ) for latent class  $k$ ;
- $\sigma_k^2$  = the variance parameter for latent class  $k$ ;
- $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_K)$ , a vector of the  $K - 1$  independent mixing proportions of the finite mixture such that  $\lambda_k > 0$  and  $\sum_{k=1}^K \lambda_k = 1$ ;
- $\phi(\cdot)$  = the standard normal density.

The distribution of the censored dependent variable  $y_i$  is (see Maddala, 1983, p. 152, for the aggregate  $K = 1$  case)

$$h(y_i | \mathbf{B}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = \sum_{k=1}^K \lambda_k \left[ \left( 1 - \Phi\left(\frac{\mathbf{x}_i' \boldsymbol{\beta}_k}{\sigma_k}\right) \right) \right]^{1-w_i} \cdot \left[ \frac{1}{\sigma_k} \phi\left(\frac{y_i - \mathbf{x}_i' \boldsymbol{\beta}_k}{\sigma_k}\right) \right]^{w_i}, \quad (4)$$

where

$$w_i = \begin{cases} 1 & \text{iff } y_i > 0, \\ 0 & \text{iff } y_i = 0, \end{cases} \quad (5)$$

$\Phi(\cdot)$  is the distribution function of the standard normal,  $\mathbf{B} = ((\beta_{jk}))$ , and  $\boldsymbol{\Sigma} = (\sigma_1^2, \dots, \sigma_K^2)$ . Hence, assuming a sample  $\mathbf{y} = (y_1, y_2, \dots, y_I)$  drawn from a

mixture of censored conditional normal densities,  $h(y_i|\mathbf{B}, \mathbf{\Sigma}, \boldsymbol{\lambda})$ , the likelihood function, is given by

$$L = \prod_{i=1}^I h(y_i|\mathbf{B}, \mathbf{\Sigma}, \boldsymbol{\lambda}). \tag{6}$$

We estimate  $\mathbf{B}$ ,  $\mathbf{\Sigma}$ , and  $\boldsymbol{\lambda}$  given observed  $\mathbf{x}$  and  $\mathbf{y}$  and a prespecified value of  $K$  by maximizing  $L$  or  $\ln L$  in (6), given the constraints imposed on  $\boldsymbol{\lambda}$  above and all elements of the vector  $\mathbf{\Sigma} > \mathbf{0}$ . The latter condition is necessary since the likelihood function is unbounded when any  $\sigma_k^2 = 0$ , and maximum likelihood estimators are not guaranteed to be consistent in this case.

Given maximum likelihood parameter estimates,  $\hat{\boldsymbol{\beta}}_k$ ,  $\hat{\sigma}_k$ , and  $\hat{\lambda}_k$ , one can then estimate the posterior probabilities of membership,  $\hat{P}_{ik}$ , of each subject  $i$  into each of the  $K$  latent classes using Bayes' rule:

$$\hat{P}_{ik} = \frac{\left[ \hat{\lambda}_k \left( 1 - \Phi \left( \frac{\mathbf{x}'_i \hat{\boldsymbol{\beta}}_k}{\hat{\sigma}_k} \right) \right) \right]^{1-w_i} \cdot \left[ \frac{\hat{\lambda}_k}{\hat{\sigma}_k} \phi \left( \frac{y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}}_k}{\hat{\sigma}_k} \right) \right]^{w_i}}{h(y_i|\hat{\mathbf{B}}, \hat{\mathbf{\Sigma}}, \hat{\boldsymbol{\lambda}})}, \tag{7}$$

such that  $\sum_{k=1}^K \hat{P}_{ik} = 1$ . These posterior probabilities provide a probabilistic clustering of the subjects into the  $K$  derived latent classes and may be utilized to form discrete partitions by assigning each subject  $i$  to the latent class whose  $\hat{P}_{ik}$  is highest.

*The EM Algorithm*

We introduce nonobserved indicator variables to allow for the formulation of an EM algorithm for parameter estimation (see De Soete & DeSarbo, 1990, for a similar basic framework with respect to a latent class probit model for "pick any/ $N$ " binary data). Let

$$z_{ik} = \begin{cases} 1 & \text{iff subject } i \text{ belongs to latent class } k, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

We assume that, for a particular subject  $i$ , the nonobserved data,  $\mathbf{z}_i = (z_{i1}, z_{i2}, \dots, z_{iK})'$ , are iid multinomially distributed with probabilities  $\boldsymbol{\lambda}$ ,

$$(\mathbf{z}_i|\boldsymbol{\lambda}) \sim \prod_{k=1}^K \lambda_k^{z_{ik}}. \tag{9}$$

The distribution of  $y_i^*$  given  $\mathbf{z}_i$  is therefore

$$(y_i^*|\mathbf{z}_i) = \prod_{k=1}^K [f(y_i^*|\mathbf{x}_i, \sigma_k^2, \boldsymbol{\beta}_k)]^{z_{ik}}. \tag{10}$$

The complete data ln likelihood function to be maximized is given by

$$\ln L_c(\mathbf{B}, \mathbf{\Sigma}, \boldsymbol{\lambda}|\mathbf{y}^*, \mathbf{z}, \mathbf{x}) = \sum_{i=1}^I \sum_{k=1}^K z_{ik} \ln f(y_i^*|\boldsymbol{\beta}_k, \sigma_k^2, \mathbf{x}_i) + \sum_{i=1}^I \sum_{k=1}^K z_{ik} \ln (\lambda_k)$$

$$\begin{aligned}
&= -\frac{1}{2} \sum_{i=1}^I \sum_{k=1}^K z_{ik} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^I \sum_{k=1}^K z_{ik} \ln(\sigma_k^2) \\
&\quad - \frac{1}{2} \sum_{i=1}^I \sum_{k=1}^K \frac{z_{ik}}{\sigma_k^2} (y_i^* - \mathbf{x}_i' \boldsymbol{\beta}_k)^2 + \sum_{i=1}^I \sum_{k=1}^K z_{ik} \ln(\lambda_k).
\end{aligned} \tag{11}$$

With  $\mathbf{z} = ((z_{ik}))$  considered as missing data and  $\mathbf{y}^*$  considered as a vector with unobservable values, we formulate a nested pair of EM procedures for maximum likelihood estimation—a “major” E-step (expectation step), and a “major” M-step (maximization step) with a “minor” EM part nested within. In the major E-step, we compute the expected value of  $\mathbf{z}_i$  given provisional estimates for  $\mathbf{B}$ ,  $\boldsymbol{\Sigma}$ ,  $\boldsymbol{\lambda}$  and the latent part of  $y_i^*$ . In the major M-step, we use a minor EM algorithm for estimating the nonpositive values of  $y_i^*$ ,  $\mathbf{B}$ , and  $\boldsymbol{\Sigma}$ , conditional upon the newly estimated values of  $\mathbf{z}$ ; subsequently, we update the latent class proportions  $\boldsymbol{\lambda}$  which completes the major M-step. These E- and M-steps are successively applied until no further improvement of the likelihood function in (11) is possible. We now discuss these steps in detail and then provide a summary of the estimation scheme.

#### The “Major” E-Step

Using Bayes’ rule, the conditional distribution of  $\mathbf{z}_i$  is

$$(\mathbf{z}_i | y_i^*, \mathbf{B}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) \sim \prod_{k=1}^K \left[ \frac{\lambda_k f(y_i^* | \boldsymbol{\beta}_k, \sigma_k^2, \mathbf{x}_i)}{\sum_{k=1}^K \lambda_k f(y_i^* | \mathbf{x}_i, \boldsymbol{\beta}_k, \sigma_k^2)} \right]^{z_{ik}}. \tag{12}$$

This implies that

$$E(z_{ik} | y_i^*, \mathbf{B}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}) = \frac{\lambda_k f(y_i^* | \mathbf{x}_i, \boldsymbol{\beta}_k, \sigma_k^2)}{\sum_{k=1}^K \lambda_k f(y_i^* | \mathbf{x}_i, \boldsymbol{\beta}_k, \sigma_k^2)}. \tag{13}$$

Given  $\mathbf{y}^*$  and provisional estimates  $\mathbf{B}^{(s)}$ ,  $\boldsymbol{\Sigma}^{(s)}$ ,  $\boldsymbol{\lambda}^{(s)}$  of the parameters  $\mathbf{B}$ ,  $\boldsymbol{\Sigma}$ ,  $\boldsymbol{\lambda}$ , respectively, the expectation of  $\ln L_c$  (ignoring the constant term without loss of generality) at the  $s$ -th major iteration is

$$\begin{aligned}
E_z[\ln L_c | \mathbf{y}^*, \mathbf{B}^{(s)}, \boldsymbol{\Sigma}^{(s)}, \boldsymbol{\lambda}^{(s)}] &\equiv Q_1(\mathbf{B}^{(s)}, \boldsymbol{\Sigma}^{(s)}, \boldsymbol{\lambda}^{(s)}, \mathbf{z}^{(s)} | \mathbf{y}) \\
&= -\frac{1}{2} \sum_{i=1}^I \sum_{k=1}^K z_{ik}^{(s)} \ln(\sigma_k^{2(s)}) \\
&\quad - \frac{1}{2} \sum_{i=1}^I \sum_{k=1}^K \frac{z_{ik}^{(s)}}{\sigma_k^{2(s)}} (y_i^* - \mathbf{x}_i' \boldsymbol{\beta}_k^{(s)})^2 \\
&\quad + \sum_{i=1}^I \sum_{k=1}^K z_{ik}^{(s)} \ln(\lambda_k^{(s)}),
\end{aligned} \tag{14}$$

where  $z_{ik}^{(s)}$  is approximated by the conditional expected value in (13) evaluated at  $\mathbf{B}^{(s)}$ ,  $\Sigma^{(s)}$ , and  $\lambda^{(s)}$ . In computing the expected value of  $z_{ik}$ , the unobservable part of  $y_i^*$  is replaced by its conditional expectation when  $y_i = 0$ . These estimates are obtained via a minor EM algorithm nested within the major M-step, discussed next.

*The ‘‘Major’’ M-Step*

In the E-step of the minor EM algorithm (nested within the major M-step), the expectation of (14) is evaluated over the conditional distribution of the partially nonobserved variable  $\mathbf{y}^*$ , given the observed data  $\mathbf{y}$  and provisional estimates  $\mathbf{B}^{(s)}$ ,  $\Sigma^{(s)}$ ,  $\lambda^{(s)}$ , and  $\mathbf{z}^{(s)}$  of the parameters  $\mathbf{B}$ ,  $\Sigma$ ,  $\lambda$ , and  $\mathbf{z}$ , respectively. This expectation is

$$\begin{aligned}
 E_{y_i^*}[E_{\mathbf{z}}(\ln L_c | \mathbf{y}^*, \mathbf{B}^{(s)}, \Sigma^{(s)}, \lambda^{(s)}) | \mathbf{y}] &= -\frac{1}{2} \sum_{i=1}^I \sum_{k=1}^K z_{ik}^{(s)} \ln(\sigma_k^{2(s)}) \\
 &- \frac{1}{2} \sum_{i=1}^{I_1} \sum_{k=1}^K \frac{z_{ik}^{(s)}}{\sigma_k^{2(s)}} (y_i - \mathbf{x}'_i \boldsymbol{\beta}_k^{(s)})^2 + \sum_{i=1}^I \sum_{k=1}^K z_{ik}^{(s)} \ln(\lambda_k^{(s)}) \\
 &- \frac{1}{2} \sum_{i=I_1+1}^I \sum_{k=1}^K \frac{z_{ik}^{(s)}}{\sigma_k^{2(s)}} E((y_i^* - \mathbf{x}'_i \boldsymbol{\beta}_k^{(s)})^2 | \mathbf{z}_i^{(s)}, \mathbf{B}^{(s)}, \Sigma^{(s)}, \lambda^{(s)}, y_i = 0), \tag{15}
 \end{aligned}$$

where we assume, without loss of generality, that the data are arranged in a manner such that the first  $I_1$  subjects correspond to those having  $y_i > 0$ .

From (10), the conditional distribution of  $y_i^*$  is

$$(y_i^* | y_i = 0, \mathbf{z}_i) \sim \sum_{k=1}^K \frac{z_{ik}}{\sigma_k} \frac{\phi\left(\frac{y_i^* - \mathbf{x}'_i \boldsymbol{\beta}_k}{\sigma_k}\right)}{1 - \Phi\left(\frac{\mathbf{x}'_i \boldsymbol{\beta}_k}{\sigma_k}\right)}. \tag{16}$$

This implies

$$E(y_i^* | y_i = 0, \mathbf{z}_i) = \sum_{k=1}^K z_{ik} \left[ \mathbf{x}'_i \boldsymbol{\beta}_k - \frac{\sigma_k \phi\left(\frac{\mathbf{x}'_i \boldsymbol{\beta}_k}{\sigma_k}\right)}{1 - \Phi\left(\frac{\mathbf{x}'_i \boldsymbol{\beta}_k}{\sigma_k}\right)} \right], \tag{17}$$

and

$$\begin{aligned}
 \text{Var}(y_i^* | y_i = 0, \mathbf{z}_i) &= \sum_{k=1}^K z_{ik} \left[ \sigma_k^2 + \left( \mathbf{x}'_i \boldsymbol{\beta}_k \frac{\sigma_k^2 \phi\left(\frac{\mathbf{x}'_i \boldsymbol{\beta}_k}{\sigma_k}\right)}{1 - \Phi\left(\frac{\mathbf{x}'_i \boldsymbol{\beta}_k}{\sigma_k}\right)} - \left( \frac{\sigma_k \phi\left(\frac{\mathbf{x}'_i \boldsymbol{\beta}_k}{\sigma_k}\right)}{1 - \Phi\left(\frac{\mathbf{x}'_i \boldsymbol{\beta}_k}{\sigma_k}\right)} \right)^2 \right]. \tag{18}
 \end{aligned}$$

Using (17) and (18), evaluated with estimates  $\mathbf{B}^{(s)}$ ,  $\Sigma^{(s)}$ , and  $\mathbf{z}^{(s)}$ , it can be shown that

$$\begin{aligned}
 E_{y_i^*} [E_z(\ln L_c | y^*, \mathbf{B}^{(s)}, \boldsymbol{\Sigma}^{(s)}, \boldsymbol{\lambda}^{(s)} | \mathbf{y})] &\equiv Q_2(\mathbf{B}, \boldsymbol{\Sigma} | \mathbf{y}, \mathbf{z}^{(s)}, \boldsymbol{\lambda}^{(s)}) \\
 &= -\frac{1}{2} \sum_{i=1}^I \sum_{k=1}^K z_{ik}^{(s)} \ln(\sigma_k^{2(s)}) - \frac{1}{2} \sum_{i=1}^{I_1} \sum_{k=1}^K \frac{z_{ik}^{(s)}}{\sigma_k^{2(s)}} (y_i - \mathbf{x}'_i \boldsymbol{\beta}_k^{(s)})^2 \\
 &\quad - \frac{1}{2} \sum_{i=I_1+1}^I \sum_{k=1}^K \frac{z_{ik}^{(s)}}{\sigma_k^{2(s)}} [E(y_i^* | y_i = 0, \mathbf{z}_i^{(s)}) - \mathbf{x}'_i \boldsymbol{\beta}_k^{(s)}]^2 \\
 &\quad - \frac{1}{2} \sum_{i=I_1+1}^I \sum_{k=1}^K \frac{z_{ik}^{(s)}}{\sigma_k^{2(s)}} \text{Var}(y_i^* | \mathbf{z}_i^{(s)}, y_i = 0) + \sum_{i=1}^I \sum_{k=1}^K z_{ik}^{(s)} \ln(\lambda_k^{(s)}). \tag{19}
 \end{aligned}$$

In the M-phase of the minor EM algorithm, (19) is maximized with respect to  $\mathbf{B}$  and  $\boldsymbol{\Sigma}$ . This maximization problem can be converted into  $K$  optimization problems of the form

$$\begin{aligned}
 \max_{\boldsymbol{\beta}_k, \sigma_k^2} &= \left\{ -\frac{1}{2} \sum_{i=1}^I z_{ik}^{(s)} \ln(\sigma_k^{2(r)}) - \frac{1}{2} \sum_{i=1}^{I_1} \frac{z_{ik}^{(s)}}{\sigma_k^{2(r)}} (y_i - \mathbf{x}'_i \boldsymbol{\beta}_k^{(r)})^2 \right. \\
 &\quad - \frac{1}{2} \sum_{i=I_1+1}^I \frac{z_{ik}^{(s)}}{\sigma_k^{2(r)}} [E(y_i^* | y_i = 0, \mathbf{z}_i^{(s)}) - \mathbf{x}'_i \boldsymbol{\beta}_k^{(r)}]^2 \\
 &\quad \left. - \frac{1}{2} \sum_{i=I_1+1}^I \frac{z_{ik}^{(s)}}{\sigma_k^{2(r)}} \text{Var}(y_i^* | y_i = 0, \mathbf{z}_i^{(s)}) \right\}, \tag{20}
 \end{aligned}$$

where  $r$  denotes the minor iteration index. This is a weighted least-squares maximization problem whose  $\boldsymbol{\beta}_k$  and  $\sigma_k^2$  estimators are

$$\hat{\boldsymbol{\beta}}_k^{(r)} = (\mathbf{x}' \mathbf{D}(z_{ik}^{(s)}) \mathbf{x})^{-1} \mathbf{x}' \mathbf{D}(z_{ik}^{(s)}) \hat{\mathbf{y}}^*, \tag{21}$$

$$\hat{\sigma}_k^{2(r)} = \left[ \sum_{i=1}^{I_1} z_{ik}^{(s)} (y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}}_k^{(r)})^2 + \sum_{i=I_1+1}^I z_{ik}^{(s)} (E(y_i^* | y_i = 0, \mathbf{z}_i) - \mathbf{x}'_i \hat{\boldsymbol{\beta}}_k^{(r)})^2 \right. \\
 \left. + \frac{\sum_{i=I_1+1}^I z_{ik}^{(s)} \text{Var}(y_i^* | \mathbf{z}_i^{(s)}, y_i = 0)}{\sum_{i=1}^I z_{ik}^{(s)}} \right], \tag{22}$$



where  $\mathbf{D}(z_{ik}^{(s)})$  denotes the  $I \times I$  diagonal matrix with  $i$ -th diagonal element  $z_{ik}^{(s)}$ , and  $\hat{\mathbf{y}}^*$  is the original  $\mathbf{y}$  vector whose  $y_i = 0$  elements are replaced by the expected values from (17).

These new estimators  $\hat{\mathbf{B}}^{(r)}$  and  $\hat{\mathbf{\Sigma}}^{(r)}$  become the new provisional estimates for the next E-step of the minor EM algorithm to compute new estimates for the unobserved values of  $\mathbf{y}^*$ . The latter are then used in the subsequent M-step of the minor EM to arrive at new estimates of  $\mathbf{B}$  and  $\mathbf{\Sigma}$ . We continue this iterative process until no further improvement in the function in (19) is possible. Thus, the minor EM algorithm amounts to predicting all the unobservable values of  $\mathbf{y}^*$  based on their conditional expectations. To complete the major M-step, (14) must be maximized with respect to  $\boldsymbol{\lambda}$ , subject to the constraints  $\sum_{k=1}^K \lambda_k = 1$  and  $\lambda_k \geq 0$ , to obtain updated estimates for the latent class proportions,  $\hat{\boldsymbol{\lambda}}^{(s)}$ . To achieve this, it suffices to maximize the augmented function

$$\sum_{i=1}^I \sum_{k=1}^K z_{ik}^{(s)} \ln(\lambda_k) - \mu \left[ \sum_{k=1}^K \lambda_k - 1 \right], \tag{23}$$

where  $\mu$  denotes a Lagrange multiplier. It can be easily shown that (23) is maximized when

$$\hat{\lambda}_k^{(s)} = \frac{\sum_{i=1}^I \hat{z}_{ik}^{(s)}}{I} = \frac{\sum_{i=1}^I \hat{P}_{ik}^{(s)}}{I}. \tag{24}$$

Hence, upon completion of the major M-step, we obtain new provisional estimates for  $\mathbf{B}$ ,  $\mathbf{\Sigma}$ , and  $\boldsymbol{\lambda}$ , which are used in the next major E-step to compute new estimates for  $\mathbf{z}$ . We continue alternating between the major E- and M-steps until no further improvement in the likelihood function in (14) is possible. Convergence to at least a local optimum solution can be established using Jensen’s inequality and/or a limiting sums argument.

*Estimation Scheme*

Schematically, the EM algorithm utilized for parameter estimation can be described as follows:

1. Initialize the “major” iteration index  $s$ :  $s \leftarrow 0$ .  
 Generate initial estimates  $\mathbf{B}^{(0)}$ ,  $\mathbf{\Sigma}^{(0)}$ , and  $\boldsymbol{\lambda}^{(0)}$ , of  $\mathbf{B}$ ,  $\mathbf{\Sigma}$ , and  $\boldsymbol{\lambda}$ , respectively.
2. Compute initial estimates of  $\mathbf{z}^{(0)}$  using (13).
3. a. Set the minor iteration index  $r$ :  $r \leftarrow 0$  and initialize  $\mathbf{B}^{(r)} = \mathbf{B}^{(s)}$ ,  $\mathbf{\Sigma}^{(r)} = \mathbf{\Sigma}^{(s)}$ .  
 b. Compute  $\hat{\mathbf{y}}^*$  using (17) for  $y_i = 0$ .  
 c. For each latent class  $k = 1, \dots, K$ , obtain estimates  $\beta_k^{(r+1)}$  and  $\sigma_k^{2(r+1)}$  using (21) and (22).  
 d. Minor convergence test:  
 if  $Q_2(\mathbf{B}^{(r+1)}, \mathbf{\Sigma}^{(r+1)} | \mathbf{y}, \mathbf{z}^{(s)}, \boldsymbol{\lambda}^{(s)}) - Q_2(\mathbf{B}^{(r)}, \mathbf{\Sigma}^{(r)} | \mathbf{y}, \mathbf{z}^{(s)}, \boldsymbol{\lambda}^{(s)})$   
 is smaller than some user-specified positive constant ( $\epsilon$ ), set  $\mathbf{B}^{(s+1)} = \mathbf{B}^{(r+1)}$ ,  $\mathbf{\Sigma}^{(s+1)} = \mathbf{\Sigma}^{(r+1)}$ , and go to Step 4.  
 e. Increment  $r$ :  $r \leftarrow r + 1$ . Return to Step 3b.
4. Compute a new estimate  $\boldsymbol{\lambda}^{(s+1)}$  of  $\boldsymbol{\lambda}$  using (24).
5. Major convergence test:  
 if  $Q_1(\mathbf{B}^{(s+1)}, \mathbf{\Sigma}^{(s+1)}, \boldsymbol{\lambda}^{(s+1)}, \mathbf{z}^{(s)} | \mathbf{y}) - Q_1(\mathbf{B}^{(s)}, \mathbf{\Sigma}^{(s)}, \boldsymbol{\lambda}^{(s)}, \mathbf{z}^{(s)} | \mathbf{y})$   
 is smaller than some user-specified positive constant ( $\epsilon$ ), stop.

6. Increment  $s$ :  $s \leftarrow s + 1$  and go back to step 2.

Thus, each minor iteration conditionally maximizes the maximum likelihood function given provisional estimates of the latent class subject memberships into the  $K$  classes. The major iteration updates the class memberships. Upon convergence of the proposed EM algorithm, we obtain estimates of  $\mathbf{B}$ ,  $\mathbf{\Sigma}$ ,  $\boldsymbol{\lambda}$ , and  $\mathbf{z}$ . The final estimates of  $\mathbf{z}$  are expectations of the nonobserved membership indicator vector, and are equivalent to the posterior probabilities ( $\hat{P}_{ik}$ ). Hence, the proposed method simultaneously estimates  $K$  tobit-like regression functions with parameters  $\mathbf{B}$ ,  $\mathbf{\Sigma}$ , the mixing proportions  $\boldsymbol{\lambda}$ , and posterior probabilities of membership ( $\hat{P}_{ik}$ ) of each subject  $i$  into each of the  $K$  latent classes.

Note that the parameters of finite mixture of univariate normal densities are identified (see Teicher, 1961, 1963; Yakowitz, 1970; Yakowitz & Spragins, 1968), although there exist no sufficient estimators for such parameters (Dynkin, 1961). Maximum likelihood estimation is probably the most commonly utilized method for estimating the standard tobit model parameters (see Olsen, 1978, and Amemiya, 1984, for a discussion of different estimation methods). Iterative procedures such as the Newton-Raphson or the EM algorithm seem to converge somewhat rapidly in the standard tobit model as revealed from the simulation study performed by Schmee and Hahn (1979). We compute the asymptotic variance-covariance matrix of the regression parameters using the asymptotic efficiency property of maximum likelihood estimators. The variance-covariance matrix,  $\mathbf{F}$ , of the regression coefficients is estimated analytically as (Berndt, Hall, & Hausman, 1974)

$$\mathbf{F} = \frac{1}{I} \left[ \sum_{i=1}^I \left( \frac{\partial \ln L_i}{\partial \boldsymbol{\theta}} \right) \left( \frac{\partial \ln L_i}{\partial \boldsymbol{\theta}} \right)' \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}}^{-1}, \quad (25)$$

where  $L_i = \ln h(y_i | \mathbf{B}, \mathbf{\Sigma}, \boldsymbol{\lambda})$ ,  $\boldsymbol{\theta}$  is a vector stacking all the free parameters, and  $\hat{\boldsymbol{\theta}}$  contains maximum likelihood sample estimates at the optimum.

#### Tests for $K$

As mentioned earlier, the number of latent classes  $K$  must be specified when analyzing data by the proposed method. However, in most applications,  $K$  is unknown and therefore needs to be inferred from the data. Since the regularity conditions necessary for the traditional likelihood ratio test for  $K$  do not hold in the case of our model, we use alternative heuristic measures typically utilized for most problems involving mixture distributions (see McLachlan & Basford, 1988) to identify the appropriate number of classes. Sclove (1977, 1983) and Bozdogan and Sclove (1984) have proposed the use of Akaike's (1974) information criterion,

$$\text{AIC}_K = -2 \ln L_K + 2N_K, \quad (26)$$

where  $N_K$  is the effective number of parameters estimated in a  $K$ -class solution,

$$N_K = JK + 2K - 1. \quad (27)$$

According to this criterion, the optimal number of latent classes is the  $K$  that produces minimum  $\text{AIC}_K$ . An alternative measure to the AIC is the consistent AIC (CAIC) heuristic that corrects for the overestimation bias of the AIC by penalizing overparameterization (see Bozdogan, 1987). This measure is defined as

$$\text{CAIC}_K = -2 \ln L_K + N_K[\ln I + 1]. \quad (28)$$

Bozdogan (1987) recommends using the CAIC rather than the AIC in the case of large samples. Because of its correction for overspecification difficulties associated with the use of the AIC, the CAIC is our measure of choice here. However, it is important to note that both AIC and CAIC rely on the same regularity conditions needed for the validity of the likelihood ratio test (see Bozdogan, 1983; Sclove, 1987; Titterington, Smith, & Makov, 1985), and hence we regard these criteria as hueristics. Further, to assess the separation of the latent classes, we utilize an entropy-based measure,  $E_K$ , using the posterior probabilities (analogous to measures of population diversity suggested by Teachman, 1980),

$$E_K = 1 - \frac{\sum_i \sum_k -\hat{P}_{ik} \ln \hat{P}_{ik}}{I \ln K}. \quad (29)$$

$E_K$  is a relative measure of "fuzziness" of the derived latent classes. It takes a minimum value of 0 when all the posterior probabilities are equal for each subject and a maximum value of 1 when the  $K$  classes are discrete partitions. Therefore, a value of  $E_K$  close to zero is cause for concern as it implies that the centroids of the  $K$  latent classes are not sufficiently separated. Note, this measure should be used in conjunction with one of the aforementioned tests for  $K$  to choose an appropriate solution.

### 3. An Application in Consumer Psychology

Understanding the behavior of consumers is an active area of research in consumer psychology and the social sciences. Very often, researchers are faced with situations in which observations on a dependent variable of interest are censored. For instance, in modeling household purchases of products during an observational period as a function of certain explanatory variables (e.g., price), a significant proportion of households may have zero purchases. Censored regression models are typically utilized (see Elrod & Winer, 1982; Tellis, 1988) since the use of a classical regression model is inappropriate as discussed in the introductory section. However, if households vary in their responsiveness to these explanatory variables, estimating a single, aggregate censored regression model may be inadequate.

We investigate this issue with respect to purchases of yogurt which is an inexpensive frequently-bought consumer product. Purchase records were obtained for a panel of households maintained by a major consumer research firm in a city in the central United States. This city is a small insulated market that has been determined by this firm to be nationally representative of the U.S. in demographic composition and consumption patterns. All the major supermarkets in this city are equipped with electronic scanning equipment to record purchase transactions. Actual purchase records indicating the brand and quantity bought are available at the household level since each panelist presents an identification card during checkout. In addition, information on potential explanatory variables characterizing the transaction and the purchasing environment such as price, use of coupons, and store advertising are available for each household.

Prior research investigating consumer purchase behavior (e.g., Elrod & Winer, 1982; Narasimhan, 1984; Tellis, 1988) suggests that consumers evaluate and choose brands based on inferences made using information inherent in "cues" such as price, advertising, and sales promotional devices such as cents-off coupons. However, such purveyors of information may have a differential impact upon consumers' purchase

behavior depending upon their sensitivities to these cues. We therefore utilize our proposed method to analyze such heterogeneity in consumer sensitivities.

The differential impact of explanatory variables on consumers' purchase behavior is assessed by deriving different sets of parameter estimates for different groups of consumers corresponding to  $K$  latent classes. Here, we evaluate the impact of three explanatory variables on the purchase behavior for each household (see Blattberg & Neslin, 1990): price ( $x_1$ ), coupon use ( $x_2$ ), and store advertising ( $x_3$ ). The price variable is measured in mean cents per ounce paid over the observational time period, coupon use is the proportion of household transactions made with a coupon, and store advertising is the proportion of household transactions made when the brand bought was featured in store advertising flyers. In addition to these three explanatory variables, we include an intercept term in the model specification.

We analyze the impact of these explanatory variables on households' total purchases of the leading brand of yogurt in the market (denoted as Brand L). The observations span the time period September 1986 to August 1987. Of the 2097 households in the sample that purchased one or more brands of yogurt in this observational period, 1013 had nonzero purchases for Brand L. For these households, the dependent variable ( $y$ ) is the log of the total quantity (ounces) of yogurt (Brand L) purchased by each household (see Narasimhan, 1984). The remaining 1084 households made zero purchases of Brand L, although each of these households did purchase other brands of yogurt. Hence, for these households, the dependent variable is censored at zero since they did not purchase Brand L. Consistent with the estimation procedure described in the previous section, the observations for the entire sample of 2097 households are arranged in a manner such that the first  $I_1 = 1013$  households correspond to those having nonzero purchases (i.e.,  $y_i > 0$ ), while the remaining households correspond to those for whom the dependent variable is censored at zero. The mean value of the dependent variable is 32.83 ounces across all households in the sample, and 67.97 ounces among purchasers of Brand L.

For the 1013 households reporting nonzero purchases of Brand L, the explanatory variables reflect the price, coupon use, and advertising information pertaining to occasions when Brand L is purchased during the observational period. For the remaining 1084 households reporting zero purchases of Brand L, values of the explanatory variables are obtained based on their "purchase opportunities" for Brand L during the observational period (see Bucklin & Lattin, 1991). We define "purchase opportunities" as purchase occasions wherein these households bought some brand of yogurt other than Brand L. To obtain the values of the explanatory variables for a household reporting zero purchases of Brand L, we first identified its purchase opportunities and the corresponding calendar weeks, and then computed the mean values of price, coupon, and advertising variables across those households that purchased Brand L in these weeks. Hence, while the dependent variable is zero for households reporting zero purchases of Brand L, the values of the respective explanatory variables are known, based on purchase opportunities for these households when Brand L could have been bought instead.

### *Analyses and Results*

We first compare the results of classical (OLS) regression and aggregate tobit ( $K = 1$ ) analyses, wherein a single set of regression coefficients  $\beta$  is estimated. We then compare the aggregate tobit model with the disaggregate results obtained using our proposed method. The inadequacy of aggregate censored regression models in capturing heterogeneous response to explanatory variables is discussed subsequently.

The aggregate classical (OLS) regression and aggregate tobit results are shown in

TABLE 2  
Aggregate Regression Coefficients

Variable	OLS	Tobit (K=1)
Price ( $\hat{\beta}_1$ )	-0.171* (0.037)	-0.278* (0.025)
Coupons ( $\hat{\beta}_2$ )	0.290* (0.027)	0.553* (0.027)
Advertising ( $\hat{\beta}_3$ )	1.159* (0.266)	2.326* (0.233)
Intercept	3.084* (0.386)	2.499* (0.257)

Note: \* denotes the ratio of the parameter estimate / standard error is greater than 2; the value in parentheses is the standard error of the parameter estimate.

Table 2. From this table, note that all the coefficients for the OLS and tobit analyses are significant at the 0.01 level. The tobit estimates indicate that households are sensitive to price ( $\hat{\beta}_1 = -0.278$ ), as well as promotional activities, such as coupons ( $\hat{\beta}_2 = 0.553$ ) and advertising ( $\hat{\beta}_3 = 2.326$ ). Households increase their consumption when prices are lower, when yogurt is featured in store advertising, and when cents-off coupons are available to obtain price discounts. However, note that the OLS regression model under-estimates the impact of the explanatory variables. While similar in sign, the coefficients for the price, coupon, and advertising variables are lower in absolute magnitude than the aggregate tobit estimates. This is because of the inherent assumptions of the classical regression model which, as discussed in the introduction section, do not accommodate zero (censored) values of the dependent variable explicitly.

However, the aggregate tobit estimates in Table 2 may mask any underlying heterogeneity in household response to price, coupons, and advertising. To investigate this further, we use our proposed method to analyze the data for a varying number of latent classes  $K$ . The values of the ln likelihood ( $\ln L$ ), number of major iterations required for convergence (IT), number of estimated parameters ( $N_K$ ), the entropy-based measures ( $E_K$ ), and information criterion values ( $CAIC_K$ ) are shown in Table 3 as  $K$  varies from 1 to 4. Note that the ln likelihood increases considerably from the aggregate case ( $\ln L = -3407.09$ ) to the two-class case ( $\ln L = -2914.12$ ), indicating the presence of considerable heterogeneity in the data. As  $K$  increases further however, the ln likelihood still increases as expected due to the increased number of parameters, although the amount of increase is much lower beyond  $K = 2$  classes. The CAIC values have a minimum at  $K = 2$ , and hence, we choose the two-class solution. The entropy measure  $E_K$  is 0.91 for the two-class solution indicating that the latent class centroids are well-separated for  $K = 2$ .

Table 4 presents the parameter estimates for each of the two latent classes. From

TABLE 3  
Model Selection

K	IT	$N_K$	Ln L	$E_K$	CAIC
1	28	5	-3407.09	0.00	6857.41
2	34	11	-2914.12	0.91	5923.37*
3	25	17	-2904.45	0.65	5955.92
4	17	23	-2884.51	0.59	5967.94

Note: \* denotes minimum CAIC value.

this table, note that there are significant differences between the two latent classes with respect to the magnitudes of the regression coefficients  $\hat{\beta}$  for the explanatory variables. Further, the regression coefficients for the aggregate case ( $K = 1$ ) in Table 2 are in a middle range and do not distinguish between the differential responsiveness of consumers to the explanatory variables. The estimate of  $\hat{\sigma}^2$  in the  $K = 1$  case ( $\hat{\sigma}^2 = 3.337$ ) is higher than either of the within-class estimates ( $\hat{\sigma}_1^2 = 1.131$ ,  $\hat{\sigma}_2^2 = 1.325$ ) indicating that the error variance is higher for the aggregate ( $K = 1$ ) solution. From Table 4, Class 1 consists of 44.9% of consumers who are not sensitive to coupons and advertising, and with a very low sensitivity to price. In contrast, Class 2 consists of 55.1% of consumers who are very sensitive to price, coupons, and advertising. Hence, these findings suggest the existence of two types of demand functions for Brand L, corresponding to two groups of consumers with one group (class 2) having larger absolute magnitudes of sensitivities than the other. (We also inspected the  $K = 3$  class solution for additional potential insights into the structure of the data. In the  $K = 3$  class solution, the coefficients for the third class closely resembled those of the second class. In addition, the size of the third latent class was much too small to be managerially useful.)

Prior consumer research suggests the potential existence of two types of shoppers: those who do not respond to information-based cues, and those who respond to store advertising and promotional signals (Bucklin & Lattin 1991; Cobb & Hoyer 1986; Hoyer 1984; Inman, McAlister, & Hoyer 1990; Park, Iyer, & Smith 1989). These two types of shoppers are seemingly reflected in the results in Table 4. Latent Class 1 seems to consist of "habitual" consumers who may plan their purchases and are not influenced by coupons and advertising. Given that yogurt is a relatively inexpensive, frequently-bought perishable item, these consumers may have built loyalty towards the leading brand and make their purchases in a habitual manner (Hoyer, 1984). Hence, consumers in Class 1 are perhaps more loyal to Brand L. Latent class 2 seems to consist of "opportunistic" consumers who are more sensitive to price and promotion, and may be inclined to purchase lesser-priced brands. They may be less loyal to Brand L and are perhaps heavy coupon users who are more prone to purchasing products featured in store advertising, given their high sensitivity to information-based cues. The next section explores these hypotheses in more depth.

TABLE 4  
Parameter Estimates For The Two-Class Solution

Variable	Latent class 1	Latent class 2
Class size ( $\hat{\lambda}_k$ )	0.449	0.551
Price ( $\hat{\beta}_{1k}$ )	-0.067* (0.029)	-1.872* (0.255)
Coupons ( $\hat{\beta}_{2k}$ )	-0.036 (0.030)	2.643* (0.302)
Advertising ( $\hat{\beta}_{3k}$ )	0.049 (0.285)	4.559* (0.824)
Intercept	4.191* (0.302)	11.786* (1.655)
$\hat{\sigma}_k^2$	1.131* (0.028)	1.325* (0.216)

Note: \* denotes the ratio of the parameter estimate / standard error is greater than 2; the value in parentheses is the standard error of the parameter estimate.

#### Latent Class Diagnosis

To assess the impact of loyalty, coupon usage, and advertising proneness upon class membership, we conduct a posterior regression analysis. We regress the log odds of a household belonging to Class 2 versus Class 1 (i.e.,  $\ln(\hat{P}_{i2}/\hat{P}_{i1})$ ) against six descriptor variables. In addition to the three consumer behavior characteristics (i.e., loyalty, overall coupon usage, and advertising proneness), we include three other variables capturing total household (yogurt) category consumption, income, and family size that potentially relate to differences in consumer sensitivities (Blattberg & Neslin, 1990) as evinced for the two derived classes. (Of these six descriptor variables, income and family size are demographic variables obtained via a household survey, while the remaining variables are based on purchases made from January 1986 to August 1986,

**TABLE 5**  
**Analysis of Posterior Probabilities**

Variable	Impact Coefficient
Loyalty	-0.375* (0.011)
Coupon usage	0.486* (0.220)
Advertising proneness	0.368* (0.074)
Category consumption	-0.020* (0.002)
Income	-0.613* (0.124)
Family Size	0.185 (0.252)
Intercept	5.758
F-statistic	252.123
$R^2$	0.420

Note: \* denotes the ratio of the parameter estimate / standard error is greater than 2; the value in parentheses is the standard error of the parameter estimate.

prior to the time frame utilized for deriving the two clusters). These descriptor variables are measured as follows: loyalty is measured as the proportion of purchases made towards the leading Brand L, coupon usage is the proportion of purchases made with a coupon, advertising proneness is the proportion of purchases made when yogurt was featured in store advertisements, income is a fourteen-point scale with "1" denoting less than \$5000 and "14" denoting \$100,000 or more, and household size is the number of members in the household including children. The results of the posterior regression analysis are displayed in Table 5.



TABLE 6  
Ten K=2 Class Solutions Using Different  
Random Starts for the Yogurt Data

	Run Number:									
	1	2*	3	4	5	6*	7	8	9	10
$\hat{\beta}_{01}$	4.292	4.227	4.315	4.292	4.297	4.257	4.271	4.289	4.288	4.283
$\hat{\beta}_{11}$	-0.076	-0.071	-0.078	-0.076	-0.076	-0.073	-0.074	-0.076	-0.076	-0.075
$\hat{\beta}_{21}$	-0.037	-0.044	-0.036	-0.037	-0.039	-0.044	-0.035	-0.037	-0.038	-0.036
$\hat{\beta}_{31}$	-0.002	-0.181	-0.012	-0.006	-0.035	-0.159	0.038	0.002	-0.004	0.016
$\hat{\beta}_{02}$	11.932	40.964	10.797	12.097	13.396	29.054	10.834	11.848	12.198	11.397
$\hat{\beta}_{12}$	-1.826	-7.798	-1.687	-1.852	-2.068	-4.922	-1.643	-1.810	-1.865	-1.736
$\hat{\beta}_{22}$	2.465	11.488	2.270	2.501	2.793	6.735	2.225	2.446	2.524	2.347
$\hat{\beta}_{32}$	3.964	18.687	3.865	3.998	4.324	9.782	3.701	3.935	4.009	3.830
$\hat{\sigma}_1^2$	1.135	1.121	1.136	1.134	1.133	1.124	1.135	1.135	1.134	1.135
$\hat{\sigma}_2^2$	1.840	6.915	2.012	1.852	2.052	4.563	1.600	1.801	1.828	1.722
$\hat{\lambda}_1$	0.441	0.469	0.438	0.441	0.443	0.460	0.441	0.441	0.442	0.441
$\hat{\lambda}_2$	0.559	0.531	0.561	0.559	0.557	0.540	0.559	0.559	0.558	0.559
IT	32	19	28	44	31	21	39	36	29	41
Ln L	-2912.17	-2966.51	-2914.43	-2912.35	-2914.98	-2949.09	-2910.11	-2911.82	-2912.22	-2911.05

Note: \* denotes a severe local optimum solution.

From Table 5, the regression analysis is significant overall ( $F = 252.123$ ) with about 42% of the variation in the log-odds of posterior class membership being explained by the chosen predictor variables. Of the six variables, loyalty, advertising proneness, category consumption, and income are significant at the 0.01 level, while coupon usage is significant at the 0.05 level. The family size variable has a positive impact coefficient, albeit insignificant. Hence, as expected, consumers in latent Class 1 are more brand loyal, while consumers in latent Class 2 are heavy coupon users and are very prone to advertised store specials. Interestingly, latent class 1 consists of households who consume relatively more yogurt. It might be that these habitual consumers buy on a somewhat regular basis given the perishable nature of yogurt. In contrast, consumers in latent class 2 may be more occasional users who are influenced by information-based cues and take advantage of promotional opportunities; moreover, they seem to have relatively lower income consistent with their high price sensitivity. Future research could explore such issues further to provide a better understanding of consumer behavior.

### *Investigation of Local Optima*

As mentioned earlier, the proposed EM estimation procedure converges to at least a local optimum solution. To examine the severity of the potential problem of local optimality, we recomputed the  $K = 2$  solution some ten times using different random starts for each of these analyses. In each case, the convergence tolerance  $\epsilon$  was set to 0.001. Table 6 presents the parameter estimates, iterations required for convergence, and  $\ln$  likelihood values for each of the ten analyses. As shown in the table, two of the ten analyses (#2 and #6) result in rather noticeable local optimal solutions. Otherwise, the estimates are quite stable, especially those concerning the mixing proportions and significant coefficients. Note, all ten analyses converged within 50 major iterations indicating somewhat rapid convergence. (CPU time was not available for these analyses).

### 4. Conclusion

We have presented a conditional mixture, maximum likelihood method for latent class regression involving a censored dependent variable. This method models sample heterogeneity by simultaneously estimating separate regression functions and subject membership in  $K$  latent classes or clusters. Maximum likelihood estimates are obtained using an EM algorithm. We illustrate the proposed method via a consumer psychology application concerning the consumption of an inexpensive, frequently-bought perishable item. Overall, our proposed method renders meaningful results by separating "habitual" consumers from "opportunistic" consumers via the derived latent classes. These two types of consumers possess different demand functions and exhibit varying patterns of responsiveness to information-based cues. Estimating a single set of aggregate tobit estimates masks this heterogeneity in consumers' response patterns. The posterior regression analysis provides further insights into the derived latent classes and corroborates the interpretation of the class-specific results.

Given the increasing use of tobit models for censored data, our proposed method can be especially useful to researchers conducting psychological and social studies where sound theoretical bases and/or relevant descriptors are not available to form a priori groups of subjects. Some directions for future research include accommodating prior information and/or subject-specific variables on the mixing parameters (see Aitkin & Rubin, 1985; Dayton & MacReady, 1988) and investigating issues such as the identification of outliers (see Aitkin & Wilson, 1980) and other stochastic model selection criteria (see Bozdogan, 1991; Rissanen 1989; Windham & Cutler 1991).

### References

- Adams, J. D. (1980). Personal wealth transfers. *Quarterly Journal of Economics*, *95*, 159–179.
- Aitkin, M., & Wilson, J. (1980). Mixture models, outliers, and the EM algorithm. *Technometrics*, *22*, 325–331.
- Aitkin, M., & Rubin, D. B. (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society, Series B*, *47*, 67–75.
- Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control*, *6*, 716–723.
- Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics*, *24*, 3–61.
- Amemiya, T. (1985). *Advanced econometrics*. Cambridge, MA: Harvard University Press.
- Ashenfelter, O., & Ham, J. (1979). Education, unemployment, and earnings. *Journal of Political Economy*, *87*, S99–S116.
- Baba, V. (1990). Methodological issues in modeling absence: A comparison of least squares and tobit analyses. *Journal of Applied Psychology*, *75*, 428–432.
- Berndt, E. K., Hall, B. H., & Hausman, J. A. (1974). Estimation and inference in non-linear structural models. *Annals of Economic and Social Measurement*, *3*, 653–665.

- Blattberg, R. C., & Neslin, S. A. (1990). *Sales promotion: Concepts, methods, and strategies*. Englewood Cliffs, NJ: Prentice Hall.
- Bozdogan, H. (1983). *Determining the number of component clusters in standard multivariate normal mixture models using model-selection criterion* (Technical Report # VIC/DQM/A83-1). Washington, DC: Army Research Office.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extension. *Psychometrika*, *52*, 345-370.
- Bozdogan, H. (1991, June). *Choosing the number of component clusters in the mixture model using a new information theoretic complexity criterion of the inverse Fisher information matrix*. Conference paper presented at the 1991 joint meeting of the Classification and Psychometric Societies, Rutgers University, New Brunswick, New Jersey.
- Bozdogan, H., & Sclove, S. L. (1984). Multi-sample cluster analysis using Akaike's information criterion. *Annals of the Institute of Statistical Mathematics*, *36*, 163-180.
- Bucklin, R. E., & J. M. Lattin (1991). A two-state model of purchase incidence and branch choice. *Marketing Science*, *10*, 24-39.
- Cobb, C. J., & Hoyer, W. D. (1986). Planned versus impulse purchase behavior. *Journal of Retailing*, *62*, 384-409.
- Dayton, C. M., & MacReady, G. B. (1988). Concomitant-variable latent class models. *Journal of the American Statistical Association*, *83*, 173-178.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the E-M algorithm. *Journal of Royal Statistical Society, Series B*, *39*, 1-38.
- DeSarbo, W. S., & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, *5*, 249-282.
- De Soete, G., & DeSarbo, W. S. (1990). A latent class probit model for analyzing pick any/N data. *Journal of Classification*, *8*, 45-64.
- Dynkin, E. B. (1961). Necessary and sufficient statistics for a family of probability distributions. *Selected translations in mathematical statistics and probability* (pp. 17-40) Providence, RI: American Mathematical Society.
- Elrod, T., & Winer, R. S. (1982). An empirical evaluation of aggregation approaches for developing market segments. *Journal of Marketing*, *46*, 65-74.
- Fair, R. (1978). A theory of extramarital affairs. *Journal of Political Economy*, *86*, 45-61.
- Fisher, G. A., & Tesler, R. C. (1986). Family bonding of the mentally ill: An analysis of family visits with residences of board and care homes. *Journal of Health and Social Behavior*, *27*, 236-249.
- Greene, W. H. (1990). *Econometric analysis*. New York: Macmillan.
- Greene, W. H., & Quester, A. (1982). Divorce risk and wives' labor supply behavior. *Social Science Quarterly*, *63*, 16-27.
- Gross, A. L. (1980). A maximum likelihood approach to test validation with missing and censored dependent variables. *Psychometrika*, *55*, 553-549.
- Hoyer, W. D. (1984). An examination of consumer decision-making for a common repeat purchase product. *Journal of Consumer Research*, *11*, 822-829.
- Inman, J. J., McAlister, L., & Hoyer, W. D. (1990). Promotion signal: Proxy for a price cut? *Journal of Consumer Research*, *17*, 74-81.
- Kinsey, J. (1981). Determinants of credit card accounts: An application of tobit analysis. *Journal of Consumer Research*, *8*, 172-182.
- Maddala, G. S. (1983). *Limited dependent and qualitative variables in econometrics*. New York: Cambridge University Press.
- McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering*. New York: Marcel Dekker.
- Moore, W. L. (1980). Levels of aggregation in conjoint analysis: An empirical comparison. *Journal of Marketing Research*, *17*, 516-523.
- Narasimhan, C. (1984). A price discrimination theory of coupons. *Marketing Science*, *3*, 128-146.
- Olsen, R. J. (1978). Note on the uniqueness of the maximum likelihood estimator for the tobit model. *Econometrica*, *46*, 1211-1215.
- Park, W. C., Iyer, E. S., & Smith, D. C. (1989). The effects of situational factors on in-store grocery shopping behavior: The role of store environment and time available for shopping. *Journal of Consumer Research*, *15*, 422-433.
- Quandt, R., & Ramsey, J. (1978). Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association*, *73*, 730-752.
- Rissanen, J. (1989). *Stochastic complexity in statistical inquiry*. Teaneck, NJ: World Scientific Publications.
- Schmee, J., & Hahn, G. J. (1979). A simple method for regression analysis with censored data. *Technometrics*, *21*, 417-432.

- Sclove, S. L. (1977). Population mixture models and clustering algorithms. *Communications in Statistics, Series A*, 6, 417-434.
- Sclove, S. L. (1983). Application of the conditional population-mixture model to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-5*, 428-433.
- Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 333-343.
- Teachman, J. D. (1980). Analysis of population diversity. *Sociological Methods and Research*, 8, 341-362.
- Teicher, H. (1961). Identifiability of mixtures. *Annals of Mathematical Statistics*, 32, 244-248.
- Teicher, H. (1963). Identifiability of finite mixtures. *Annals of Mathematical Statistics*, 34, 1265-1269.
- Tellis, G. J. (1988). Advertising exposure, loyalty, and brand purchase: A two-stage model of choice. *Journal of Marketing Research*, 25, 134-144.
- Titterton, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. New York: Wiley & Sons.
- Tobin, J. (1958). Estimation of relationship for limited dependent variables. *Econometrica*, 26, 24-36.
- Windham, M. P., & Cutler, A. (1991, June). *Information ratios for validating cluster analyses*. Conference paper presented at the 1991 joint meeting of the Classification and Psychometric Societies, Rutgers University, New Brunswick, New Jersey.
- Witte, A. D. (1980). Estimating the economic model of crime with individual data. *Quarterly Journal of Econometrics*, 94, 57-84.
- Yakowitz, S. J. (1970). Unsupervised learning and the identification of finite mixtures. *IEEE Transactions Information Theory and Control, IT-16*, 330-338.
- Yakowitz, S. J., & Spragins, J. D. (1968). On the identifiability of finite mixtures. *Annals of Mathematical Statistics*, 39, 209-214.

*Manuscript received 2/19/92*

*Final version received 7/14/92*