An Application of Markov Latent Class Analysis for Evaluating Reporting Error in CE Screening Questions

Paul P. Biemer
Research Triangle Institute

1. Introduction

In the Consumer Expenditure Interview Survey, respondents are asked a number of details about the expenditures they incurred during the past three months. These detail questions are asked for a consumer item only it the respondent responds positively to a screening question that essentially determines if any purchases of the item have been made over the past three months. If the response is negative, the detailed questions are skipped.

Due to the large number of items in the survey and the extensive information requested for some purchases, interviews can last between one and two hours or longer and the burden on respondents is considerable. Although the potential for under-reporting expenditures as a result of false negative responses to the screening questions is apparent, the error in CE screening questions has never been formally investigated. One reason for the lack of information on this important source of error is the difficult of evaluating the error.

Traditional methods for evaluating survey error require the use of a gold standard measurement that can serve as the truth for purposes of estimating reporting error. In the CE Interview Survey such gold standard measurements are often very difficult to obtain or unavailable. For example to verify that the respondent had no automobile maintenance expense, a reinterview audit survey could be conducted where the respondent is asked to locate all receipts for automobile expenditures over the last three months. The interviewer could then check for the existence of auto maintenance receipts in this census of receipts. However, even then, the absence of automobile maintenance documentation is no assurance that expenditures of this type were not incurred over the reference period. Thus, information on expenditure under-reporting in the CE due to screening question error has heretofore not been available.

Latent class analysis (LCA) is a fairly new methodology for evaluating the error in survey reports without the use of gold standard measures. In lieu of measurements which are accurate, the method assumes an error model for the available measurements and uses maximum likelihood estimation techniques to estimate the parameters of the error model. Thus, the validity of the LCA estimates hinges on the ability of the model to accurately represent the error-generating process. LCA usually requires at least two but preferably three replicate measurements of the same item (at the same point in time) as a condition of estimability of the error parameters.

For panel surveys such as the CE Interview Survey, a related statistical method referred to as Markov latent class analysis (MLCA) is available which essentially relaxes the requirement that the replicate measurements pertain to the same point in time in order to use the panel reinterviews in the estimation process. MLCA requires a minimum of three measurements of the same units over different time periods as would be the case for a panel survey where units are interviewed on three occasions. The MLCA model then specifies parameters for both the period to period changes in the status of the item as well as the measurement error associated with measuring those changes. In the automobile maintenance item example, the MLCA model would specify parameters associated with both the month to month change in the true status of the expenditure (i.e., transitions from no expense to some expense and vice versus) as well as the error in reporting this status in the CE.

Recently, the MLCA methodology was successfully applied to the Current Population Survey (CPS) to estimate the classification error in reports of labor force participation (see Biemer and Bushery, in press). This study also provided evidence of the validity of the MLCA methodology for estimating labor force status classification error. The present paper applies similar models to the CE screening question data in order to determine whether useful information on the magnitudes and correlates of screening question reporting error can be extracted directly from the CE panel data.

The purpose of this investigation is twofold. First, we provide some background on the MLCA modeling approach and illustrate how it can be applied to the CE Interview Survey. Second, using the estimates of screening question reporting error obtain from the MLCA, we further illustrate how MLCA can be used to conduct exploratory data analysis of the correlates of screening error.

2. The Data

The data used in this study comes from the 1997 Consumer Expenditure Interview Survey. This survey was designed to collect information on data on up to 95 percent of total household expenditures. Consumer units (CU's) are interviewed once every three months for five consecutive quarters to obtain the expenditures for 12 consecutive months. The initial interview for a CU is used as a bounding interview and these data are not used in the estimation. The survey is designed to collect data on major items of expense which respondents can be expected to recall for three months or longer. New panels are initiated every month of the year so that each month, 20 percent of the CU's are being interviewed for the first time.

Approximately 7,000 sample units are contacted for an interview each quarter. Allowing for bounding interviews and nonresponse (including vacancies), the number of participating sample units per quarter is targeted at approximately 5,000.

Our analysis was confined to CU's that were interviewed three consecutive quarters beginning in the first quarter of 1997. Therefore, three subsamples of CU's can be defined as shown in Figure 1. As we will see in the discussion of the Markov latent class approach, a minimum of three observations, equally spaced in time on all the CU's in the analysis is required for the identifiability of the MLC models. CU's where one or more interviews were missing from the three consecutive quarters were deleted from the analysis. Thus, after deleting CU's interviewed once, four, or five times in January, February, and March, as well as CU's not completing all three interviews, the total sample size for the study across all three subsamples is 2,189 CU's.

As previously stated, the objective of our study is to assess measurement errors and its causes in the data collection operations. To this end, only unweighted data is used in the analysis for to used weighted data might distort data on the error processes operating during the data collection operations. One drawback of using unweighted data is that inferences regarding the overall quality of the published CE estimates of expenditures cannot be made; however, this is not an important objective in the study. For purposes of this analysis, the sample will be treated essentially as a simple random sample from a superpopulation which is the CE data series for the current survey design.

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.
Subsample 1	2 & 3			3 & 4			4 & 5		
Subsample 2		2 & 3			3 & 4			4 & 5	
Subsample 3			2 & 3			3 & 4			4 & 5

Figure 1. Definition of the Three Subsamples for the Analysis. Subsample 1 is composed of CU's interviewed for the second or third times in January, the third or fourth times in April, and the fourth or fifth times in July. The other two subsamples are defined analogously for February and March interviews.

3. The Markov Latent Class Model

Markov latent class (MLC) models were first proposed by Wiggins (1973) and refined by Poulsen (1982). Van de Pol and de Leeuw (1986) established conditions under which the model is identifiable and gave other conditions of estimability of the model parameters. In this section, we develop the MLC model in the context of the CE Interview Survey and suggest other applications and its generalizations.

MLC models assume that all the variables in the analysis are classification variables. For example, in our analysis we will consider the screening questions in the CE where outcome variable is a dichotomous response taking the value 1 if the CU reports a purchase for a particular consumer item for the month and 2 if not.

Let the CE target population be divided into L groups or domains and let the variable G be the indicator for group membership. For example, G may be related to the administration of the survey - such as interview length, use of records, number of times previously interviewed, etc. - or may describe CU characteristics such as size, income, age of CU members, etc. Let, $G_i = 1$ if the ith population member is in group 1, $G_i = 2$ for group 2 and so on.

In the preliminary stages of our analysis, we considered models for describing the error in a dichotomous variable, say D_m, defined for a single consumer item (such as pet supplies) and a particular month, m, of the CE where $D_m = 1$ if the CU purchased the item during the month and $D_m = 2$, otherwise. Thus, for the nine months of data collection shown in Figure 1, we would define D_m, $m = 1, \ldots, 9$ for the expenditure pattern for the item over the entire nine-month period. However, there were a number of difficulties with this modeling approach. First, and most important, the MLC models provided a very poor fit to these data due primary, we believe, to the failure of the Markov assumption (described below) to hold. Second, the models were quite complex with many 0-cells that caused convergence problems in the EM algorithm used for maximum likelihood estimation. In addition, the model fitting process was quite tedious and time consuming since a single model run could require one hour on a Pentium III 450. Therefore, this modeling approach was abandoned in favor of the following simpler modeling approach.

Rather than specifying a variable for a monthly purchase, we define a summary variable for the frequency of purchases of the item for all three months of a quarter. Such a model would be much more likely to satisfy the Markov assumption and would run much more efficiently as a result of the reduction in the dimensionality of the problem.

Let the subscript combination (g, i) denote the i-th CU in group $G = g$ for $g = 1, \ldots, L$ and $i = 1, \ldots, n_g$. For the first quarter of the year, we define A_{gi} as follows.

$$A_{gi} = \begin{cases} 1 & \text{if CU } (g,i) \text{ reports the item was purchased in all three months of the quarter} \\ 2 & \text{if CU } (g,i) \text{ reports the item not purchased in any month of the quarter} \\ 3 & \text{if CU } (g,i) \text{ reports the item was purchased in one or two months of the quarter} \end{cases}$$

For analyzing three consecutive quarters (for example, quarters beginning in January, April, and July), let B_{gj}, and C_{gk} denote the analogously defined purchase status variable for the second and third quarters, respectively

Associated with each of the three observed variables is a latent variable for the *true* quarterly purchase status of the CU. For quarters 1, 2, and 3 let X_{gi}, Y_{gj}, and Z_{gk} denote trichotomous variables with categories defined analogously to A_{gi}, B_{gi}, and C_{gi}, respectively, except that they represent the true rather than observe statuses of the CU's. For notational convenience, we will drop the subscripts (g,i), but retain the relationship of the unsubscripted variable to an individual unit within a group. Further, the term "true purchaser" will be used to describe CU's who purchase the item in all three months of a quarter (denote for quarter 1 by $X = 1$), "true non-purchaser" for CU's not purchasing the item in any quarter (denote for quarter 1 by $X = 2$), and "true mixed consumer" for CU's who purchase in the item in some but not all quarters (denoted for quarter 1 by $X = 3$).

Let $\pi_{x,y,z|g}$ denote $\Pr(X=x, Y=y, Z=z \mid G=g)$, let $\pi_{y|g,x}$ denote $\Pr(Y = y \mid X = x, G = g)$ and let $\pi_{z|g,y,x}$ denote $\Pr(Z = z \mid Y = y, X = x, G = g)$. Then, the probability that an individual in group g is has purchase status x in quarter 1, y in quarter 2, and z in quarter 3 is

$$\pi_{x,y,z|g} = \pi_{x|g}\,\pi_{y|g,x}\,\pi_{z|g,x,y} \tag{3.1}$$

Finally, under the Markov assumption (which is required for model identifiability; see Van de Pol and de Leeuw, 1986), we assume

$$\pi_{z|g,x,y} = \pi_{z|g,y}, \tag{3.2}$$

I.e., at quarter 3, the true status of an individual does not depend on the quarter 1 status once the quarter 2 status is known. An alternate interpretation is that the quarter 3 purchase status given the quarter 2 status does not depend upon the quarter 1 to quarter 2 transition.

One can conceive of some situations where the Markov assumption may not hold for the CE interview survey. For example, for large and costly purchases such as automobiles, CU's who are non-purchasers in quarter 2 may be much more likely to be non-purchasers in quarter 3 if they were mixed consumers in quarter 1 than if they were non-purchasers in quarter 1. This is because buying an automobile occurs infrequently and CU's purchasing a car in the last six months are less likely to purchase another car in the next three months than CU's who have not purchased a car in the last six months. For other items, such as cable TV subscription, the Markov assumption may hold quite well since the purchases statuses are more stable over a three month period Thus, the fit of the MLCM's and the validity of the resulting estimates are expected to vary considerably by consumer item.

At least two methods for assessing the validity of the Markov assumption for panel data are available. One method suggested by Van de Pol and de Leeuw (1986) is based upon four waves of panel data. Another method requires test-retest reinterview data for each quarter. These analyses are beyond the scope of this present paper. However, a description of the method based upon reinterview data can be found in Van de Pol and Langeheine (1997) who conducted such an analysis for labor force data in The Netherlands.

Using an extension of the notation established above, we denote the response probabilities in each of these classifications as follows:

$$\begin{aligned} \pi_{a|g,x} &= \Pr(\mathrm{A} = a \,|\, \mathrm{X} = x) \\ \pi_{b|g,y} &= \Pr(\mathrm{B} = b \,|\, \mathrm{Y} = y) \\ \pi_{c|g,z} &= \Pr(\mathrm{C} = c \,|\, \mathrm{Z} = z) \end{aligned} \tag{3.3}$$

Thus, $\pi_{a=1|g,x=2}$ is the probability that the CE classifies a person in group g as a purchaser (A = 1) when the true status is non-purchaser (X = 2). Likewise, $\pi_{a=2|g,x=2}$ is the probability that the CE correctly classifies a person in group g as a non-purchaser.

Finally, we assume that

$$\begin{aligned} &\Pr(A=a, B=b, C=c \mid G=g, X=x, Y=y, Z=z) \\ &\quad = \Pr(A=a \mid G=g, X=x)\Pr(B=b \mid G=g, Y=y)\Pr(C=c \mid G=g, Z=z) \end{aligned}$$

and write

$$\pi_{a,b,c\,|\,x,y,z} = \pi_{a\,|\,gx}\,\pi_{b\,|\,gy}\,\pi_{c\,|\,gz}. \tag{3.4}$$

This latter assumption has been examined in the literature for labor force survey data, but never for expenditure survey data. For labor force data, Meyers (1988) and Singh and Rao (1995) investigated the assumption and concluded that it was a reasonable approximation. Van de Pol and Langeheine (1997) used latent class analysis that involved both panel data and reinterview data to estimate the classification error for various types of month to month labor force transitions. They found only weak evidence that respondents who change labor force status have lower reliability than those who do not. However, more work is needed to determine whether the assumption is reasonable for expenditure survey data.

In the analysis that follows, less emphasis will be placed on the MLCA estimates for a particular item. Rather, we will be looking for trends in error rates across all the items in the study. Further, our analysis is exploratory in that we will be using MLCA to generate hypotheses regarding the causes of error that can be followed up and tested in other settings. For example, a finding that some consumer items are less subject to under-reporting than others could suggest a study that might be conducted with a small number of subjects in a laboratory setting to determine if the result can be verified and, if so, the reasons for the differential data quality. Thus, the manner in which MLCA will be used in the following is somewhat robust to failures of the model assumptions to hold.

With these assumptions, we can write the probability of classifying a CE sample member in cell (g,a,b,c) of the *GABC* table as follows:

$$\pi_{g,a,b,c} = \sum_{x,y,z} \pi_{g} \pi_{x|g} \pi_{a|g,x} \pi_{y|x,g} \pi_{b|y,g} \pi_{z|g,y} \pi_{c|g,z} \tag{3.5}$$

Under multinomial sampling, the likelihood function for the *GABC* table is

$$\text{Likelihood} = \Pr(GABC) = \text{constant} \times \prod_{g,a,b,c} \pi_{g,a,b,c}^{n_{gabc}}. \tag{3.6}$$

The MLC models we will consider are part of a family of latent class models known as modified path models (Goodman, 1974). Modified path models are essentially a chain of univariate conditional models with latent variables where the ordering in the chain reflects a type of causal sequence. For example, the ordering of the conditional probabilities in equation (3.5) reflects the sequence that variables are observed in the CE. The first variable listed is the grouping variable followed by the latent and manifest variables associated with the first interview, followed by the corresponding latent and manifest variables for the second interview and so on. In the modified path modeling approach, each probability in this chain can be modeled by a univariate logistic model where the parameters are simultaneously estimated using maximum likelihood estimation methods. For example, for the probability $\text{B}_{a|xg}$, several choices of logistic models are possible from the one way dependence model consisting of terms for the interactions *AX* and *AG* to the full model consisting of the second order interaction *AXG*.

All the models we will consider are hierarchical models for which presence of an interaction term implies presence of all lower order interactions and main effect terms containing the sample letters. As an example, the hierarchical log-linear model containing the term *AXG* also contains the terms *A*, *X*, *G*, *AX*, *AG*, and *GX*. However, in specifying logistic log-linear models for conditional probabilities, all terms in the model must contain the dependent variable; thus, the model for $\text{B}_{a|xg}$ containing *AXG* contains the "main effect" terms *A*, *AX*, *AG* as well as the "first order interaction" term *AXG* only. This model is specified in shorthand notation as {*AXG*}.

Modified path models can be represented by path diagrams as in Figure 2. The lines drawn between two variables denote the covariance of the two variables and the arrow denotes the order of the variable in the probability statement (3.5). The line extending from the grouping variable, *G*, to the line between *X* and *A*, say, implies that relationship between *X* and *A* varies by the level of the variable, *G*. This path diagram corresponds to a model where the following is assumed for each conditional probability in (3.5):

Month to month transitions: $B_{x|g} = \{XG\}$, $B_{y|gx} = \{YG, YX\}$, $B_{z|gy} = \{ZG, ZY\}$,
Measurement error: $B_{a|gx} = \{AXG\}$, $B_{b|gy} = \{BYG\}$, $B_{c|gz} = \{CZG\}$

Under the assumptions made previously, the model parameters are estimable using maximum likelihood estimation methods. Van de Pol and de Leeuw (1986) provides the formula for applying the E-M algorithm for estimating the parameters of this model and conditions for their estimability. These methods have been implemented in the REM software which will be applied to the CE data set in the next section.

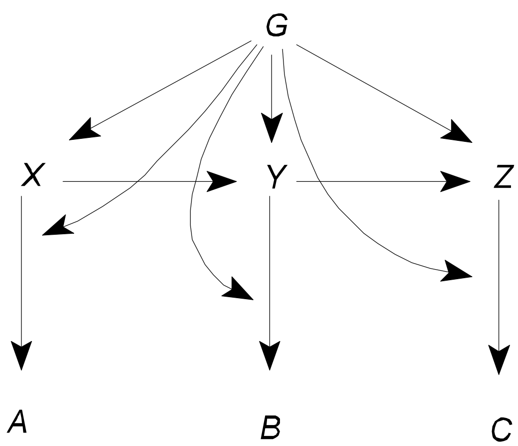

Figure 2. Path Diagram for a Markov Latent Class Model

4. Application to the CE Interview Survey

4.1 Errors in the CE Screening Questions

The CE Interview Survey collects data on monthly expenditures for a wide-range of consumer items. The questionnaire is organized into 24 sections with each section treating a particular area of household expense. Table 1 is a list of the sections by topic. Within each section dealing with expenditures, screening questions precede questions that ask about the details of the expenditures. For example, for in Section 4, Utilities and Fuels for Owned and Rented Properties, the screening question is question 1 of Part A:

> Since the 1st of (*month, 3 months ago*), have you (or any members of your CU) received any bills for telephone services? Do not include bills for telephones used entirely for business purposes.

If the response to this question is “yes,” more detailed questions are asked regarding the number and nature of the bills, the amount of each bill, month received, and what was included on each. If the response is “no,” the detailed questions are skipped and the interviewer moves onto the next consumer item in the section (utilities, in this case).

[INSERT TABLE 1 ABOUT HERE]

The focus of our investigation is the error associated with these screening questions. These questions are critical to the accuracy of the expenditure data since they determine whether or not detailed information on expenditures is collected. There appears to be no published literature on the accuracy of these questions, yet there are many reasons to suspect that responses to the screening questions would be subject to reporting error, particularly under-reporting error. Some of these are the following:

1. ***Encoding Error.*** An item may have been purchased by another household member without the respondent's knowledge. Even if records are used to identify purchases, many purchases may not appear in the records. For example, a checkbook may record that a check was written for $350 at Sam's Wholesale Club with no record of what was purchased there.

2. ***Comprehension Error.*** The respondent may misunderstand the question or not realize that certain items that should be reported are included in the question. For example, the respondent purchased an outdoor awning for a patio and, because it was not listed explicitly in the screening questions, failed to report it.

3. ***Recall Error***. The respondent may forget that a purchase was made during the three-month period or may remember it as happening more than three months ago. Again, records such as checkbooks may not be detailed enough to aid the respondent's recall and other records such as receipts or bills may be incomplete.

4. ***Communication Error.*** The respondent may deny purchasing an item even though he knows it should be reported. This may be the result satisficing (i.e., the respondent is fatigued or pressed for time and wants to shortcut the interview) or social desirability concerns (i.e., the respondent is embarrassed to admit a purchase). For example, the respondent may have spent a disproportionate amount of money on pet supplies and is embarrassed to admit that these items dominate so much of the meager family budget.

The effect of each of these types of errors on under-reporting varies depending upon the characteristics of the item. For example, expenditures that are incurred on a regular (say, monthly) basis such as cable TV subscription fees, trash collection fees, and electricity, are not likely to be affected by comprehension error or recall error. However, expenditures such as small, infrequent purchases that occur with no regular pattern may be subject to all types of errors and these items are expected to have the largest reporting error. Further, small expenditures may be more likely to be subject to encoding errors than large major expenditures. Figure 3 provides a summary of the types of errors one may expect for the various types of expenditures. In general, items that are subject to more types of errors can be expected to have greater levels of reporting error; however, the level of error from any source can still overwhelm the errors from all other sources, so these expectations can be violated. Nevertheless, this table may be useful later in the analysis as we attempt to uncover the causes of error in the CE screening questions.

In additional to the characteristics of the items itself, other external factors related to the characteristics of the interviewer, the respondent or CU, and the interview may influence screener error. For example, interviewers may change the wording of the questions, fail to ask a question, fail to probe when necessary, or may provide feedback that influences reporting accuracy. CU size may also influence responses since larger CU's are likely to have more expenditures and more consumers. Thus, there is a greater potential for encoding error, recall error, and communication (satisficing) error than in smaller CU's. Also, the length of the interview and the use of receipts, checkbook registries, bills, and other records used in the interview can influence screener error. In general, we would expect less error in expenditure amounts when records are used; however, it is unclear how the use of records will effect the accuracy of screening questions.

Interview length is also a determinant of response accuracy. Longer interviews suggest more expenditures and, potentially, greater respondent fatigue and satisficing. However, longer interviews are also associated with greater care and completeness of reporting. Shorter interviews may suggest more negative responses to the screening questions and may therefore be an indicator of under-reporting.

Consumer Item Characteristic	Encoding	Comprehension	Recall	Communication
Small expenditures				
Frequent, random	T	T	T	T
Frequent, regular		T		T
Infrequent	T	T	T	T
Large expenditures				
Frequent, random		T	T	T
Frequent, regular		T		T
Infrequent	T	T		T

Figure 3. Correspondence between Consumer Item Characteristics and Screening Question Reporting Error. A check mark suggests items which may be more prone to the type of error in the column heading.

In addition to length of the interview, the number of times the CU has been interviewed previously may also influence screening error. For example, in the initial interview, the respondent may not realize that responding "yes" to a screening question will always result in a somewhat extensive barrage of questions to record the details of the expenditures. By the second or third interviews, the interview process as well as successful ways to shorten the interview by responding “no” to the screening process may be well-learned. Conversely, accuracy may actually improve with the number of interviews since it is possible that respondents learn what information on expenditures they must have on hand in order to provide accurate information as well as to expedite the interview process.

The position of an item in the questionnaire can influence reporting accuracy. As the interview continues, respondents may tire and become bored with the process. Items that appear in the earlier sections of the questionnaire may be less subject to satisficing behavior than items in the later sections.

In the analysis to follow, we will investigate the effects of some of these factors on screening accuracy using the Markov latent class analysis approach. As mentioned in the introductory section, the major goal of the paper is to demonstrate how MLC models can be used to investigate some of the external and internal correlates of screening accuracy. Therefore, the analysis will not pursue any causal factor or item in much depth, but rather illustrate that useful information regarding the nature of measurement error in the CE can be gleaned from MLCA. As such the results should be considered as preliminary. In the Discussion section of the paper, we offer some ideas for extending this analysis to a wider range of factors and items that could be considered in a comprehensive investigation of screening error in the CE.

4.2 The Analytic Approach

Our analysis of the CE data focuses on a subset of consumer items and a few CU and interview variables that may be correlated with reporting error. In selecting the items for the investigation, we chose items from beginning, middle, and ending sections so that question position (or respondent fatigue) effects can be estimated. The items selected for the study and their corresponding section numbers are shown in Table 2. The items in the table span Sections 4 through 19 of the questionnaire. Some of the items, such as CABLE and TRASH, are regularly incurred small expenditures. Others, such as SHOES, DENTAL, EYES, and DRUGS are infrequently incurred expenditures. There are also large, infrequent expenditures such as FURN and possibly VEQOTH.

[TABLE 2 ABOUT HERE]

Note that some variables, such as GAS and TRASH, are single items and others, such as KITCHN and CLOTH are combinations of items. We would expect that single items should be more prone to under-reporting error than groups of items since the respondent would have to erroneously report no single item in the entire group of items were purchased in order to record a “no” response. For a single item, a false report of “no” to only one item is needed. These potential patterns of error will be investigated in the results that follow.

Five potential correlates of reporting error were considered in our analysis. First, as seen from Figure 1, the sample for all three months of a quarter combined differs by interview frequency; that is, some respondents were interviewed once more than other respondents. In addition, the subsample varies depending upon the month of interview. Thus, our model for screener error should incorporate these design factors and test their significance. Other explanatory variables that will be considered in the analysis include: CU size, interview length, and use of records. Table 2 list the variables that will be used in the MLC models.

Note the absence of the interview frequency variable in Table 2. In the preliminary investigation of the variables, interview frequency was not significant once the other explanatory variables were accounted for in the models, so this variable was dropped from the analysis. Also note that the definitions of the interview length variable, L, and the records use variable, R, are defined at the interview level for all three interviews in the quarter combined. Thus, $L = 1$ denotes a set of CU's whose interviews during a quarter are typically short, $L = 4$ denotes CU's for which all three interviews in the quarter are long, and $L = 2$ or 3 denote mixtures of short, long, and average length interviews. Similarly, $R = 1$ denotes a set of CU's who always or almost always used records for all three interviews in the quarter, $R = 2$ denotes CU's that sometimes use records, and $R = 3$ denotes CU's that never or almost never use records. Finally, as described previously, X, Y, and Z are the latent variables associated with the three quarters and which denote the true purchase status of a CU while A, B, and C are the corresponding observed variables.

A number of MLC models were explored in the initial stages of the analysis. The modeling process began with a very simple model to which additional variables were added until a model that fit the data adequately and parsimoniously was identified. It became clear during the initial model building process that this model building approach was computationally intensive. Two or three hours of computer time could be needed to fit a single model and numerous models would have to be for each item. Moreover, whenever four or more explanatory variables were entered into the model, the algorithm often failed to converge, probably as a result of the large number of 0-cells in the seven-way (i.e., four explanatory variables and three dependent variables) tables . Thus, a different model fitting strategy was needed.

For computational efficiency, the same model structure (i.e., main effect and interaction terms) should be fit to all consumer items since a full-scale analysis of the CE would consider all the items in the questionnaire, not just a subset of items as in Table 2. It would save much time and effort if a model structure could be identified that applied to all the items and then the parameters of this model could be estimated separately for each item. The risk in this approach is that the model could be over-parameterized for some items and inadequate for others. Still, there would be little to be gained by optimizing the model for each item if the objective analysis is solely to estimate probabilities of misclassification. The presence of 0 terms in the model will have minimal impact on the estimated probabilities other than increasing their standard errors somewhat. Therefore, our approach will be to identify a single model structure that will be used for all 19 items and is suggested from theoretical considerations presented in prior discussion.

As mentioned previously, the ℓEM software failed to converge for some items when large complex models were tried. Therefore, only models containing the indicators A, B, C and three explanatory were fit in our analysis. This means that the fit of models containing all four explanatory variables in Table 2 will not be examined.

To specify the models that were fitted, consider MLC models containing the manifest variables L, R, F, A, B, and C. Extending the notation of (3.5), the local independence model formulation for a model containing these variables can be written as

$$\pi_{f,l,r,a,b,c} = \sum_{x,y,z} \pi_{f,l,r} \pi_{x|f,l,r} \pi_{a|f,l,r,x} \pi_{y|x,f,l,r} \pi_{b|y,f,l,r} \pi_{z|f,l,r,y} \pi_{c|f,l,r,z} \tag{4.1}$$

where each of the conditional probability terms is modeled by a logistic log-linear model. One such modified path model that is of interest in our study will be referred to as Model 1 which specifies the following submodels for the terms in (4.1):

Table 3. Terms and Restrictions for Model 1

Term	Submodel	Term	Submodel
$B_{f,l,r}$	$\{FLR\}$	$B_{b\|f,l,r,y}$	same as $B_{a\|f,l,r,x}$
$B_{x\|f,l,r}$	$\{XF, XL, XF\}$	$B_{z\|f,l,r,y}$	$\{ZY, ZF, ZL, ZF\}$
$B_{a\|f,l,r,x}$	$\{AX, AF, AL, AF\}$	$B_{c\|f,l,r,z}$	same as $B_{b\|f,l,r,y}$
$B_{y\|f,l,r,x}$	$\{YX, YF, YL, YF\}$		

The model restricts the error probabilities to be stationary or equal across time points; i.e, $B_{a|f,l,r,x}$ = $B_{b|f,l,r,y}$ = $B_{c|f,l,r,z}$. However, there is no similar restriction of the transition probabilities $B_{x|f,l,r}$, $B_{y|f,l,r,x}$, and $B_{z|f,l,r,y}$. Moreover, the levels of all latent and indicator variables in the model depend upon the explanatory variables in the simplest way that can be specified, viz., first order interactions terms only.

A similar model, denoted by Model 2, was fit substituting the CU size variable, *F*, with the interview subsample variable, *S*. The structure of both models is identical and each model contains a total of 98 parameters. Although we are not particularly interested in the variation of the error estimates across the three levels of *S*, Model 2 was fit primarily as a check on Model 1. Estimates that differ importantly between the two models could be an indication that *S* should not be ignored in the analysis.

The fit statistics associated with these two models is shown in Table 4 for each consumer item. In this table, L^2 is the likelihood ratio chi-square statistic based upon 874 degrees of freedom. For an adequate fit, the p-value of this statistic (see Table 4 under "p-value") should be a least 0.05. The smallest value of p occurs for DRUGS (0.16). However, for all other items the p-values are near 1 indicating that the model almost perfectly replicates the data. Fortunately, both Model 1 and Model 2 fit very well and the estimates of classification error were similar for both models. Therefore, in the next section, the estimates from only Model 1 will be reported.

[INSERT TABLE 4 ABOUT HERE]

4.3 Results and Key Findings

To obtain an overview of the response accuracy of the CE across consumer items, we estimated the probability that the CE process correctly classifies a CU as a purchaser, non-purchaser or mixed consumer for the quarter. Since the error probabilities were restricted to be the same for all three quarters, the probability of a correct classification is given by estimates of $\Pr(\text{correct}|X = x) = \Pr(A = x \mid X = x)$ for $x = 1, 2, 3$; i.e., the error associated with the quarter 1 observation, A, provides the error estimates for the other two quarters as well. The estimates of reporting error from Model 1 as well as $\Pr(X = x)$, the true prevalence of each purchaser status, are provided in Table 5 for all 19 consumer items considered. A number of points can be made from this table as follows:

- Across all items, the probability of correctly classifying a true non-purchaser is high - average of about 97 percent. Non-purchaser is also the most prevalent consumer status among the three statuses.
- The accuracy of classifying true purchasers varies considerably across items and is quite low, about 38 percent on average. The true prevalence rates for this type of purchaser is also quite low for the items considered.
- The accuracy with which mixed consumers are classified is about 76 percent on average with considerable variability across items. With few exceptions, the prevalence of mixed consumers in the sample is higher, often considerably so, than the prevalence of purchasers.
- The accuracy of reporting tends to decrease as the interview progresses. For the first six items in Table 4, the average accuracy is about 50 percent for true purchasers and 80 percent for true mixed consumers. For the bottom six items, the accuracy is only about 5 percent for purchasers and 72 percent for mixed consumers.
- The data provide no evidence that the combined items (SPORTS, TV, FURN, KITCHN, CLOTH, VEQMIN, VEQOTH, DRUGS) are reported with more accuracy than other items.
- Regular purchases such as CABLE, GAS, TRASH, and HOUSKP are among the most accurately reported items for all three types of purchasers.

Next, we examine the effect of the length of the interview on response accuracy and confine the analysis to the mixed consumer group only. Table 6 summarizes the results of this analysis. The last column of the table is the ratio of accuracy rates for the longest interview group (L=4) to the shortest interview group (L=1). The average of this ratio across the 19 items is 1.11 which indicates a tendency toward greater accuracy for the group whose interviews during a quarter were the longest.

The next two tables, Tables 7 and 8, examines the effects of records use (R) and CU size (F) on response accuracy for the mixed purchaser group. Again, the last column of these tables is the ratio of accuracy rates for the two extreme levels which measures the maximum effect that is expected. There appears to be no consistent effects of either R or F on the accuracy rates: in both cases the average ratios are close to 1. However, some items show large effects in the expected direction. For example, records use appears to improve reporting accuracy for CABLE, FURN, VEQOTH, and DENTAL, while having small positive to negative effects on the other items.

Likewise, accuracy is somewhat improved when the CU size is small for TV, FURN, VEQOTH, and PETS while other items show a much smaller or even negative effect of small family size. Such results are quite difficult to explain or interpret.

Finally, we examine the error probabilities for true purchasers and true mixed consumers by item in Table 9. Turning first to the left half of the table, the probability that a true purchaser is misclassified as a non-purchaser is Pr(A = 2|X=1) in the table while Pr(A = 3|X = 1) denotes the probability of misclassifying a true purchaser as a mixed consumer. Averaged across the items, both types of misclassifications appear equally likely; however, as before, the story varies considerably by item. For example, true VEQMIN purchasers are much more likely to be misclassified as non-purchasers than mixed consumers. However, for EYES, the opposite appears to be true.

On the right half of the table, Pr(A = 1|X=3) denotes the probability that a true mixed consumer is classified as a true purchaser while Pr(A = 2|X = 3) is the probability of classifying a true mixed consumer as non-purchaser. Here, there appears to be a stronger tendency to classify mixed consumers as non-purchasers rather than purchasers - leading to under-reporting of expenditures. However, for the items CABLE, GAS, and DRUGS, there is a substantial risk of mixed consumers being misclassified as consumers - leading over-reporting for these items. Thus, although under-reporting appears to the be dominate type of screener error, our analysis indicates that over-reporting of some items is also a potential problem.

The following is a summary of the key findings from the MLCA organized according to the research issues investigated.

1. *What is the general level of reporting error in the CE screening questions?*

This is very little error associated with the misclassification of non-purchasers. However, for purchasers the average error rate is considerable: approximately 63 percent misclassification. Approximately half of the misclassified purchasers are reported as non-purchasers and the other half in the mixed consumer category. There is considerable variation in the classification error rates by item.

For mixed consumers, the average error rate is approximately 25 percent: much lower than purchasers but still substantial. In fact, because the mixed consumer group is more than twice as large as the purchaser group (15 percent of the sample versus 6.5 percent), roughly one third of the purchases missed in the CE are due to mixed consumers reporting as non-purchasers.

Finally, combining the under-reported purchases across both the purchaser and mixed groups and all items in the study, the total number of purchasers missed in the CE is roughly one third of all purchases; in other words, on average, the screener questions successfully record a monthly purchase in roughly 2 out of 3 months that the purchase is made.[1]

2. What items are subject to large under-reporting errors? Is over-reporting a problem for some items?

---

[1]This is computed by assuming that the average number of months a mixed consumer actually purchases an item is 1.5.

Which items will have large under-reporting errors appears to be mostly unpredictable, although much more analysis could be done here to relate item characteristics to reporting error. Some ideas for further analysis will be discussed subsequently. However, this very preliminary study suggests that regularly occurring purchases such as CABLE, GAS, TRASH, and HOUSKP are relatively accurately reported. There appears to be a slight tendency for infrequent purchases such as COMPUT and FURN to be under-reported.

Over-reporting appears to occur for regular purchases that are purchased in only one or two months of the quarter and are reported as purchased in all months of the quarter.

3. Are reporting errors greater for items that occur earlier in the interview than for later occurring items?

There is a tendency for items appearing earlier in the interview to have smaller error rates than items appearing in the later sections. For example, true purchasers are accurately identified in 50 percent of the cases in the early items compared with only 5 percent in the later items. For mixed consumers, the effect is more subtle - 80 percent compared with 72 percent.

4. How does the reporting accuracy compare for CU's whose interviews are usually very short versus households with very long interviews?

CU's who tend to have very long interviews throughout a quarter also tend to have greater accuracy according to our analysis. Among mixed consumers, the average increase in accuracy was approximately 10 percent and among purchasers it was twice that - approximately 20 percent.

5. How does the reporting accuracy compare for small versus large CU's?

Averaging across all 19 items in the study, we observed no discernable effect of CU size on reporting accuracy.

6. How does the reporting accuracy compare for CU's that typically use records for most items versus CU's who seldom use records for any items?

Surprisingly, we detect no general effect of records use on response accuracy. However, the effect seems to be quite item-specific and is attenuated when averaged across all the items in the study. Although records use was a significant factor for most items (as determined by the MLCA Wald statistic), the predicted effect of increasing response accuracy through the use of records is not consistent across items. Thus, these results are very difficult to interpret.

7. Are purchasers misclassified more often as mixed consumers or non-purchasers?

True purchasers appear to be equally likely to be misclassified as mixed-consumers or non-purchasers across the items.

8. Are mixed consumers misclassified more often as purchasers or non-purchasers?

Mixed consumers are much more likely to be misclassified as non-purchasers than purchasers (approximately 19 percent misclassification as non-purchasers versus approximately 3 percent as purchasers). This result is consistent with the general tendency away from over-reporting purchases in the CE. We also noted a small chance, say 3 percent or less, of non-purchasers to be misreported as mixed consumers: a tendency that is slightly less for CU's that use records than for those that do not.

5. Discussion

As mentioned in the Introduction, the primary goal of this project was to demonstrate an application of latent class analysis to the CE with the intent of encouraging more research into this area of data analysis in the future. In our analysis, we applied MLCA to three consecutive quarters of the CE in order to address a number of issues related to error in the screening questions for 19 selected consumer items. The results of this research are promising. The estimates of screener response error appear plausible in most instances and the results of the multivariate analysis of 19 items generally agree with our preconceived expectations from measurement error theory.

The results of this study indicate that MLCA is a useful device for exploratory investigations of measurement error in the CE Interview Survey. Since the analysis requires no data other than what is available from the panel survey and since MLCA software is readily available, measurement error studies using MLCA methods are relatively inexpensive to apply and can be applied to all items that are repeated each quarter. In our analysis, MLCA results provided new insights regarding the nature and magnitudes of the error and suggested areas where more indepth investigations of the error sources are needed.

However, there are a number of limitations of the present research that should be noted. First, our study has not fully addressed the validity of the MLCA estimates. Although the estimates seem plausible in many instances, other results seem puzzling and counterintuitive. For example, it appears that greater use of records by respondents can have a detrimental effect on data accuracy for some items, while positively affecting response, as expected, for others items. In addition, the error rates for some consumer items seem implausibly high. Such anomalies in the results are troublesome and suggest areas were further investigations of the statistical properties of MLCA are needed. The strategy employed in this paper of using MLCA to search for trends across data items and to focus on clear and consistent variations in the error estimates was an attempt to ameliorate these anomalies. We believe this strategy to be robust to many model inadequacies, however, our study stops short of confirming this.

Second, much more study is needed of the potential correlates of screener error. Because our investigation was preliminary and primarily pedagogical, the number of explanatory variables considered was small. Other characteristics of the CU such as income, age, presence of children in the CU as well as characteristics of the interview such as number of call-backs required for completion, interviewer ratings of data quality, and presence of other CU members during the interview could also be explored. In addition, the coding procedures used for two of the variables we considered - records use and interview length - could be further refined. For example, interview length could be entered into the model as a continuous rather than categorical variable. Further, to be an effective explanatory variable, records use is needed at the item level rather than the interview level.

Third, the number of items considered in the present study was quite limited. A much expanded list of items is needed to thoroughly investigate influence of item characteristics on response error. For example, every consumer item in the CE could be coded according to those characteristics that could potentially affect response accuracy. Examples are cost of the item (small, medium, large), frequency with which the item is purchased among CU's that purchase the item, regularity of the purchase (monthly, quarterly, randomly), saliency of the purchase, and so on. Then the MLCA estimates of screener accuracy could be correlated with the item characteristics much more formally than was done in the present study, for example, using analysis of variance techniques in order to identify characteristics of items that predict response error. Such information would be very useful for designing data collection procedures that are tailored to an items specific characteristics.

Since our study was preliminary by design, much more work is needed to understand how best to use MLCA methods for the CE Interview Survey. Nevertheless, we believe this to be an important tool for investigating measurement error issues in the survey and our results to date provide the promise of even greater importance of the method in future studies.

REFERENCES

Biemer, P. and Bushery, J. (to be published). "On the Validity of Markov Latent Class Analysis for Estimating Classification Error in Labor Force Data" in *Survey Methodology*

Goodman, L. (1974). "Analysis of Systems of Qualitative Variables When Some of the Variables are Unobservable. Part I: A Modified Latent Structure Approach," *American Journal of Sociology*, 79, 1179-1259.

Meyers, B. D. (1988). "Classification-Error Models and Labor-Market Dynamics," *Journal of Business & Economic Statistics*, Vol. 6, No. 3, pp. 385-390.

Poulsen, C.S. (1982). *Latent Structure Analysis with Choice Modeling Applications*, doctoral dissertation, Wharton School, University of Pennsylvania.

Singh, A.C. and Rao, J.N.K. (1995). "On the Adjustment of Gross Flow Estimates for Classification Error with Application to Data from the Canadian Labour Force Survey," Journal of the American Statistical Association, Vol. 90, No. 430, pp. 478-488.

Van de Pol, F. and de Leeuw, J. (1986). "A Latent Markov Model to Correct for Measurement Error," *Sociological Methods & Research*, Vol. 15, Nos. 1-2, pp 118-141.

Van de Pol, F. and Langeheine, R. (1997). "Separating Change and Measurement Error in Panel Surveys with an Application to Labor Market Data," in L. Lyberg, et al (eds.) *Survey Measurement and Process Quality*, John Wiley & Sons, NY.

Vermunt, J. (1997). *ℓEM: A General Program for the Analysis of Categorical Data*, Tilburg, University.

Wiggins, L.M. (1973). *Panel Analysis, Latent Probability Models for Attitude and Behavior Processing*, Elsevier S.P.C., Amsterdam.

Table 1. Sections of the CE Interview Survey Questionnaire

Section	**Topic**
Section 1	General Survey Information
Section 2	Rented Living Quarters
Section 3	Owned Living Quarters and Other Owned Real Estate
Section 4	Utilities and Fuels for Owned and Rented Properties
Section 5	Construction, Repairs, Alternations, and Maintenance of Property
Section 6	Appliances, Household Equipment, and other Selected Items
Section 7	Household Equipment Repairs, Service Contracts, and Furniture Repair and Reupholstering
Section 8	Home Furnishings and Related Household Items
Section 9	Clothing and Sewing Materials
Section 10	Rented and Leased Vehicles
Section 11	Owned Vehicles
Section 12	Vehicle Operating Expense
Section 13	Insurance Other than Health
Section 14	Hospitalization and Health insurance
Section 15	Medical and Health Expenditures
Section 16	Educational Expenses
Section 17	Subscriptions, Memberships, Books, and Entertainment Expenses
Section 18	Trips and Vacations
Section 19	Miscellaneous Expenses
Section 20	Expense Patterns for Food, Beverages, and other Selected Items
Section 21	Credit Liability
Section 22	Work Experience and Income
Section 24	Total CU Income

Table 1. The Consumer Variables Used Analyzed in the Study

Variable	Section	Description
CABLE	4	Cable TV, satellite services, or community antenna
GAS	4	Bottled or tank gas
TRASH	4	Trash/garbage collection
SPORTS	6	Combined sports, recreation, and excercise equipment
TV	6	Combined television, radio, video, and sound equipment
FURN	8	Combined furniture
KITCHN	8	Combined kitchenware
ACCESS	9	Accessories
CLOTH	9	Combined clothing
SHOES	9	Footwear (including athletic shoes not specifically purchased for sports)
VEQMIN	12	Combined for minor vehicle repairs
VEQOIL	12	Oil change, lubrication, and oil filter
VEQOTH	12	Combined for other purchases (audio equipment, accessories, etc.)
DENTAL	15	Dental care
DRUGS	15	Combined medicine and medical supplies
EYES	15	Eye exams, eye glasses, other eye services
COMPUT	19	Computer information services and computer games
HOUSKP	19	Housekeeping services
PETS	19	Pet services and veterinary expenses

Table 2. Definition of Variables for the Analysis

Explanatory Variables	
S	Interview panel S = 1 Y January, April, July panel S = 2 Y February, May, August panel S = 3 Y March, June, September panel
F	Family size F = 1 Y single person CU F = 2 Y 2 or 3 person CU F = 3 Y 4 or more person CU
L	Typical interview length (defined at the panel level) L = 1 Y very short; i.e., interview lengths in {*s*,*s*,*s*} or {*s*,*s*,*m*} L = 2 Y short; i.e., any combination other than those in L = 1, 3, or 4 L = 3 Y medium; i.e., interview lengths in {R,*m*,*m*} L = 4 Y long; i.e, interview lengths in {R,R,R} or {R,R,*m*} where *s* Y 45 minutes or less; *m* Y 45 to 90 minutes; R Y more than 90 minutes
R	Typical use of records (defined at the panel level) R = 1 Y always or almost always used records R = 2 Y sometimes used records R = 3 Y never or almost never used records
Latent Variables	
X	Latent variable defined for the first quarter indicating the **true** purchase frequency of a consumer item for the quarter X = 1 Y purchase for each month of the quarter X = 2 Y no purchase for any month X = 3 Y purchase for some months
Y, Z	Same as X except defined for the second and third quarters, respectively.
Indicator Variables	
A	Indicator variable defined for the first quarter indicating the **observed** purchase frequency of a consumer item for the quarter A = 1 Y purchase for each month of the quarter A = 2 Y no purchase for any month A = 3 Y purchase for some months
B, C	Same as A except defined for the second and third quarters, respectively.

Table 4. Fit of the Model with Interview Length, Records Use, Family Size and Interview Month

Variable Name	Model with Family Size (Model 1)		Model with Survey Panel (Model 2)	
	L^2	p-value	L^2	p-value
CABLE	607	1.00	554	1.00
GAS	375	1.00	356	1.00
TRASH	578	1.00	566	1.00
SPORTS	328	1.00	362	1.00
TV	454	1.00	493	1.00
FURN	263	1.00	278	1.00
KITCHN	435	1.00	404	1.00
ACCESS	249	1.00	237	1.00
CLOTH	732	0.99	796	0.94
SHOES	459	1.00	465	1.00
VEQMIN	333	1.00	309	1.00
VEQOIL	432	1.00	452	1.00
VEQOTH	384	1.00	398	1.00
DENTAL	289	1.00	510	1.00
DRUGS	857	0.53	91	0.16
EYES	248	1.00	220	1.00
COMPUT	289	1.00	274	1.00
HOUSKP	262	1.00	245	1.00
PETS	400	1.00	403	1.00

Table 5. True Prevalence of Consumer Class and the Probability of Correct Classification by Consumer Item (Cell entries are percentages)

	Purchaser		**Nonpurchaser**		**Mixed**	
Variable	**Pr(X = 1)**	**Pr(Correct)**	**Pr(X = 2)**	**Pr(Correct)**	**Pr(X = 3)**	**Pr(Correct)**
CABLE	22.50	98.42	47.62	97.19	29.88	78.63
GAS	**5.57**	**93.84**	**86.12**	**99.86**	**8.31**	**64.65**
TRASH	3.77	91.99	81.55	99.15	14.68	88.01
SPORTS	5.04	5.25	86.33	100.00	8.63	97.85
TV	2.40	13.42	78.40	92.76	19.20	89.27
FURN	4.52	5.73	90.52	92.77	4.96	58.94
KITCHN	10.80	9.56	79.76	96.14	9.44	87.73
ACCESS	3.04	2.50	87.58	100.00	9.38	88.40
CLOTH	11.29	55.74	56.10	85.29	32.61	82.31
SHOES	5.34	10.02	85.80	100.00	8.86	70.37
VEQMIN	5.79	1.48	73.74	99.63	20.46	95.92
VEQOIL	11.80	8.13	65.56	93.88	22.64	65.83
VEQOTH	2.14	19.99	77.48	97.68	20.38	38.60
DENTAL	2.89	53.40	74.41	97.41	22.70	59.66
DRUGS	8.56	86.41	67.24	87.19	24.20	77.43
EYES	12.20	2.02	76.66	98.69	11.14	73.33
COMPUT	0.23	94.63	92.41	99.80	7.36	56.22
HOUSKP	0.64	50.52	98.08	100.00	1.28	94.31
PETS	5.34	10.02	85.80	100.00	8.86	70.37
Average	6.52	37.53	78.48	96.71	15.00	75.68

Table 6. Probability of a Correct Response by Typical Interview Length for Mixed Cus
(Cell entries are percentages)

Variable	L = 1	L = 2	L = 3	L = 4	(L=4)/(L=1)
CABLE	72.55	74.98	83.96	91.94	1.27
GAS	82.58	69.51	46.50	60.00	0.73
TRASH	87.26	89.36	82.56	88.40	1.01
SPORTS	n/a	98.89	99.49	94.98	n/a
TV	75.90	93.50	85.78	90.62	1.19
FURN	n/a	60.97	57.34	59.37	n/a
KITCHN	88.57	87.75	79.72	87.72	0.99
ACCESS	64.99	92.55	90.99	89.13	1.37
CLOTH	91.63	81.50	78.27	86.61	0.95
SHOES	79.23	71.89	73.71	63.67	0.80
VEQMIN	82.24	97.04	99.23	99.99	1.22
VEQOIL	63.26	62.55	92.49	68.19	1.08
VEQOTH	28.04	47.66	27.07	48.12	1.72
DENTAL	45.52	57.96	60.97	75.15	1.65
DRUGS	80.90	75.50	74.15	83.83	1.04
EYES	72.21	77.55	82.96	63.86	0.88
COMPUT	59.59	57.05	46.90	67.28	1.13
HOUSKP	n/a	89.73	99.88	99.87	n/a
PETS	79.23	71.89	73.71	63.67	0.80
Average	72.11	76.73	75.56	78.02	1.11

Table 7. Probability of a Correct Response by Records Use for Mixed Cus
(Cell entries are percentages)

Variable	**R = 1**	**R = 2**	**R = 3**	**(R=1)/(R=3)**
CABLE	86.54	78.19	73.12	1.18
GAS	61.54	73.82	61.45	1.00
TRASH	87.83	89.95	85.43	1.03
SPORTS	96.39	98.66	99.91	0.96
TV	92.38	91.69	82.64	1.12
FURN	63.54	64.00	51.22	1.24
KITCHN	92.07	84.26	86.47	1.06
ACCESS	93.58	87.95	83.61	1.12
CLOTH	84.02	81.87	81.42	1.03
SHOES	65.78	71.89	70.03	0.94
VEQMIN	98.52	92.22	98.77	1.00
VEQOIL	66.45	66.60	65.06	1.02
VEQOTH	41.70	39.71	31.32	1.33
DENTAL	66.52	57.02	55.47	1.20
DRUGS	79.63	76.94	75.57	1.05
EYES	60.84	99.67	99.51	0.61
COMPUT	51.91	52.93	85.29	0.61
HOUSKP	88.36	99.37	99.67	0.89
PETS	65.78	71.89	70.03	0.94
Average	75.97	77.82	76.63	1.02

Table 8. Probability of a Correct Response by CU Size for Mixed Cus
(Cell entries are percentages)

Variable	**F = 1**	**F = 2**	**F = 3**	**(F=1)/(F=3)**
CABLE	79.82	79.34	75.98	1.05
GAS	62.84	71.39	64.75	0.97
TRASH	87.78	85.60	89.84	0.98
SPORTS	99.64	99.83	95.69	1.04
TV	96.84	94.32	81.67	1.19
FURN	63.56	64.00	51.22	1.24
KITCHN	84.72	90.32	89.96	0.94
ACCESS	87.95	81.30	94.16	0.93
CLOTH	83.22	76.39	87.02	0.96
SHOES	73.95	75.95	64.19	1.15
VEQMIN	98.50	87.81	97.87	1.01
VEQOIL	64.63	66.67	68.23	0.95
VEQOTH	39.31	55.63	33.01	1.19
DENTAL	59.26	57.23	62.21	0.95
DRUGS	73.43	83.96	81.55	0.90
EYES	64.84	81.63	87.52	0.74
COMPUT	49.19	68.44	55.87	0.88
HOUSKP	91.96	96.67	98.55	0.93
PETS	73.95	75.95	64.19	1.15
Average	75.69	78.53	76.24	1.00

Table 9. Probability of an Error
(Cell entries are percentages)

	Misclassification of True Purchasers		Misclassification of True Mixed Consumers	
Variable	**Pr(A=2|X=1)**	**Pr(A=3|X=1)**	**Pr(A=1|X=3)**	**Pr(A=2|X=3)**
CABLE	0.25	1.33	16.36	5.01
GAS	0.95	5.22	10.81	24.53
TRASH	3.00	5.01	4.93	7.06
SPORTS	46.52	48.23	2.15	0.00
TV	53.15	33.44	1.84	8.89
FURN	60.73	50.16	0.00	10.92
KITCHN	33.97	56.47	3.24	9.03
ACCESS	47.34	50.16	0.68	10.92
CLOTH	1.93	42.32	5.49	12.19
SHOES	32.29	57.69	0.39	29.24
VEQMIN	96.62	1.90	0.18	3.91
VEQOIL	22.94	68.94	0.00	34.17
VEQOTH	50.06	29.95	0.11	61.29
DENTAL	17.75	28.86	0.00	40.34
DRUGS	2.91	10.68	16.87	5.70
EYES	82.76	15.22	0.00	26.67
COMPUT	5.24	0.13	0.63	43.15
HOUSKP	26.19	23.29	0.00	5.69
PETS	32.29	57.69	0.39	29.24
Average	32.47	32.30	3.28	19.37