RC23329 (W0409-040) September 7, 2004
Mathematics

IBM Research Report

Latent Class Regression on Latent Factors

Jia Guo, Melanie Wall
Division of Biostatistics
School of Public Health
University of Minnesota
A460 Mayo Building, MMC 303
Minneapolis, MN 55455-0378

Yasuo Amemiya
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Latent Class Regression on Latent Factors

Jia Guo†, Melanie Wall†[1] and Yasuo Amemiya††

† Division of Biostatistics

School of Public Health

University of Minnesota

A460 Mayo Building, MMC 303

Minneapolis, MN 55455-0378, USA

†† IBM T.J. Watson Research Center

June 16, 2004

[1]Corresponding author. Email: melanie@biostat.umn.edu, phone: (612)626-2138, fax: (612)626-0660.

ABSTRACT

In the research of public health, psychology, and social sciences, many research questions investigate the relationship between a categorical outcome variable and continuous predictor variables. The focus of this paper is to develop a model to build this relationship when both the categorical outcome and the predictor variables are latent (i.e. not observable directly). This model extends the latent class regression model so that it can include regression on latent predictors. Maximum likelihood estimation is used and two numerical methods for performing it are described: the Monte Carlo Expectation and Maximization algorithm (MCEM) and Gaussian quadrature followed by quasi-Newton algorithm. A simulation study is carried out to examine the behavior of the model under different scenarios. A data example involving adolescent health is used for demonstration where the latent classes of eating disorders risk are predicted by the latent factor body satisfaction.

KEY WORDS:

latent class models, factor analysis, Monte Carlo EM.

1. Introduction

In the research of public health, psychology, and social sciences, it is very common to have variables or constructs that cannot be measured directly by a single observable variable but instead are hypothesized to be the driving force underlying a series of observed variables. The underlying or unobservable variables are called latent variables and a long literature exists and is evolving of methods for measuring them and examining relationships among them (for a brief history see Bartholomew and Knott, 1999). Just as observed variables can be characterized as continuous or categorical, so can latent variables. Latent class analysis (Lazarsfeld and Henry, 1968; Clogg, 1995; Hagenaars and McCutcheon, 2002) considers models for measuring latent categorical variables hypothesized to take on a finite number of values which partition the population into a finite number of discrete groups. Applications of the latent class model are common in the health science literature including, e.g., Uebersax and Grove, 1990, measuring distinct diagnostic categories given presence/absence of several symptoms; Flaherty, 2002, measuring smoking initiation; Croudace et al., 2003, studying typologies for nocturnal enuresis. Factor analysis or general latent factor (trait) analysis (Lawley and Maxwell, 1971; Moustaki and Knott, 2000) considers models for measuring continuous latent variables. Numerous health science applications exist where hypothesized continuous latent factors are used, e.g., Bowling, 1997, measuring quality of life; Neumark-Sztainer, et.al, 2003a and 2003b, utilizing social cognitive theory of health behaviors; Lee, et.al, 2003, measuring attitudes toward drinking alcohol.

Many research questions propose to investigate the relationship between a categorical response variable and continuous predictors. Logistic regression is the obvious technique to use when both the outcome and the predictors are observed directly. But, when either the categorical outcome or the continuous predictors cannot be observed directly, different methods are needed (see Table 1). If the response variable is measured via a latent class model and the predictors are observable, the approach called latent class regression can be applied to analyze the relationship (Dayton and Macready, 1988; Bandeen-Roche, et al., 1997). Another case is where the predictors are latent factors and the categorical response variable is observable. This case may be considered as a kind of errors-in-variables problem (Fuller, 1987; Carroll, et al., 1995). In particular since the outcome is categorical, this would be a nonlinear errors-in-variables problem and has been extensively studied with several methods proposed (Burr, 1988; Carroll, et al., 1995). Following the above reasoning, the need remains to develop a model and method for including both categorical and continuous latent variables, which will be the focus of this paper.

To motivate the need for such an analysis technique, we consider an example from a project conducted to study adolescent nutrition and obesity called Project EAT (Neumark-Sztainer, et al., 2002). One research question of interest is whether an adolescent girl's body satisfaction could predict her eating disorders risk class. Body satisfaction had been hypothesized by the researchers to be a continuous latent factor measured by a battery of self report Likert items related to satisfaction with different parts of one's body (e.g. hips, shoulders, waist, etc.). The outcome variable of interest, eating disorders risk class, was hypothesized to be a categorical latent variable. The researchers hypothesized that there were different types of eating disorders risk related to girls engaging in purging vs. those engaging in restriction behaviors. A checklist of 9 unhealthy weight control behaviors was asked on the questionnaire. No absolute classification rule based on the checklist of 9 behaviors exits, but a latent class model can be used to measure the different latent classes of eating disorders risk. In section 6, this example will be examined further in particular the modeling of the relationship between the latent body satisfaction and eating disorders risk.

The paper is organized as follows. In section 2, we review the latent factor model, the latent class model, the latent class regression model and introduce notation. In section 3, based on the models in section 2, we propose a new model for the case when both the categorical response and continuous predictors are latent variables. In section 4, the maximum likelihood method is considered and two different computational algorithms are proposed. In particular, the Monte Carlo Expectation and Maximization (MCEM) algorithm is demonstrated with flexible assumptions of the distribution of the latent factors and the Gaussian quadrature approximation followed by quasi-Newton maximization method is proposed for the case when the latent factors are normally distributed (available in SAS PROC NLMIXED). A simulation study examining the behavior of the model is shown in section 5. In section 6, we apply this model to the Project EAT data to analyze the relationship between body satisfaction and the characteristics of the latent classes for eating disorders risk formed in the data. Discussion and future work is given in section 7.

2. Motivating Models

2.1 Latent factor analysis model

Suppose we have a data set with n individuals. Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{iP})^T$ be a P dimensional vector of continuous observations for each individual i $(i = 1, \ldots, n)$. The factor analysis model (Lawley and Maxwell, 1971) assumes there are Q $(Q < P)$ latent factors $\mathbf{f}_i = (f_{i1}, \ldots, f_{iQ})^T$, which relate to $\mathbf{X}_i$ in the following way,

$$\mathbf{X}_i = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{f}_i + \boldsymbol{\epsilon}_i. \tag{1}$$

Here the P dimensional vector $\boldsymbol{\mu}$ and $P \times Q$ matrix $\boldsymbol{\Gamma}$ contain known and unknown scalars and $\boldsymbol{\epsilon}_i$ is a P dimensional vector of random error with $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$, $Var(\boldsymbol{\epsilon}_i) = \boldsymbol{\Psi}$, and $\boldsymbol{\epsilon}_i$ is assumed independent of $\mathbf{f}_i$. Furthermore, $\boldsymbol{\Psi}$ is assumed to be diagonal, which implies along with the assumption that $\boldsymbol{\epsilon}_i$ and $\mathbf{f}_i$ are independent, that any correlations found between the elements in the observed vector $\mathbf{X}_i$ are due to their relationship with common $\mathbf{f}_i$ and not due to some spurious correlation between $\boldsymbol{\epsilon}_i$. As is, model (1) is not generally identifiable and so the following error-in-variables parametrization will be considered,

$$\mathbf{X}_i = \begin{pmatrix} \boldsymbol{\lambda}_0 \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Lambda} \\ \mathbf{I} \end{pmatrix} \mathbf{f}_i + \boldsymbol{\epsilon}_i, \tag{2}$$

where $\mathbf{I}$ is a $Q \times Q$ identity matrix, $\mathbf{0}$ is a Q dimensional vector of zero, $\boldsymbol{\lambda}_0$ and $\boldsymbol{\Lambda}$ are known or unknown scalars. This form allows straightforward interpretation of $\mathbf{f}_i$ and does not lose the generality of model (1). A measurement model of the form (2) possibly after reordering the elements of $\mathbf{X}_i$ and centering all qualities is widely used in practice (see e.g. Fuller, 1987; Jöreskog and Sorbom, 1996).

Estimation and inference for model (2) have been considered in many different ways relying on varying assumptions of the distributions of $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ (Bollen, 1989). Under the assumption that $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are normally distributed, the maximum likelihood method is commonly used to estimate $\boldsymbol{\lambda}_0$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ because of its optimal properties. Furthermore, it has been shown (Anderson and Amemiya, 1988; Browne and Shapiro, 1988) that the estimators for model (2) using the maximum likelihood method assuming normality are asymptotically consistent even if the normality assumptions for $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are violated. All estimation methods are based on the number of factors Q being fixed and known. A common technique for choosing the number of factors in the exploratory model (i.e. where the elements of $\boldsymbol{\Lambda}$ are freely estimated) is to examine a screen plot of the eigenvalues of $\mathbf{S}$, i.e. the covariance matrix for $\mathbf{X}$. The number of factors can be chosen to be the number of eigenvalues before the elbow in the plot.

More recently the latent factors model has been considered in the following more general form,

$$f(\mathbf{X}_i) = \int f(\mathbf{X}_i|\mathbf{f}_i) f(\mathbf{f}_i) d\mathbf{f}_i. \tag{3}$$

Assuming $f(\mathbf{X}_i|\mathbf{f}_i) \sim N\left(\begin{pmatrix} \boldsymbol{\lambda}_0 \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Lambda} \\ \mathbf{I} \end{pmatrix} \mathbf{f}_i, \boldsymbol{\Psi}\right)$ and $f(\mathbf{f}_i) \sim N(\boldsymbol{\mu}_f, \boldsymbol{\Phi})$, this is just the normal factor analysis model (2). In the form (3), more flexible distributions can be assumed for $f(\mathbf{X}_i|\mathbf{f}_i)$, e.g. any exponential family (Moustaki and Knott, 2000; Sammel, et al., 1997, Dunson, 2000 and 2003).

2.2 Latent Class model

Let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ})^T$ denote the vector of the observed variables for each individual i, the basic finite mixture model (McLachlan and Peel, 2000) is,

$$f(\mathbf{Y}_i) = \sum_{k=1}^{K} f(c_i = k) f_k(\mathbf{Y}_i|c_i = k). \tag{4}$$

Here K is the number of latent classes, c_i is the latent class variable, $f_k(\mathbf{Y}_i|c_i = k)$ is the class-specific distribution for $\mathbf{Y}_i$, and $\pi_k \equiv f(c_i = k)$ denotes the marginal probability of belonging to latent class k such that $\sum_{k=1}^{K} \pi_k = 1$.

The special case of model (4) when all elements of $\mathbf{Y}_i$ are binary variables is often called the "latent class model" in the literature (Clogg, 1981 and 1995). Moreover, in the latent class model it is generally assumed that the elements of $\mathbf{Y}_i$ are mutually independent within classes. The latent class model is then,

$$f(\mathbf{Y}_i) = \sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} \pi_{j|k}^{Y_{ij}} (1 - \pi_{j|k})^{1-Y_{ij}}, \tag{5}$$

where $\pi_{j|k} \equiv Pr(Y_{ij} = 1 | c_i = k)$ is the probability that $Y_{ij} = 1$ when the i^{th} individual is in the latent class k. Although it is possible to constrain certain parameters $\pi_{j|k}$ to zero, it is most common to allow them to be freely estimated by the data. Given a fixed number of classes K, maximum likelihood estimates can be obtained straightforward by the EM algorithm. It is common to fit models with different numbers of classes and compare them by BIC values and choose the model with the smallest BIC (Collins, et al., 1993).

2.3 Latent Class Regression model

Latent class regression models (Dayton and Macready, 1988; Bandeen-Roche, et al., 1997) build a relationship between a latent class variable and other observable continuous or categorical predictor variables of interest. Consider again the latent class model (5), it may be thought of as a measurement model relating $\mathbf{Y}_i$ to c_i. That is, a model for how c_i is measured by $\mathbf{Y}_i$. Now consider additional observed covariates $\mathbf{X}_i$ where it is of interest to investigate if $\mathbf{X}_i$ can directly effect the latent class membership, i.e. if $\mathbf{X}_i$ can be a good predictor for the latent classes. It will be assumed that $\mathbf{Y}_i$ is conditionally independent of $\mathbf{X}_i$ given c_i. That is, intuitively the model assumes the only reason that $\mathbf{X}_i$ is related to $\mathbf{Y}_i$ is because it is related to c_i. The following is the latent class regression model,

$$f(\mathbf{Y}_i | \mathbf{X}_i) = \sum_{k=1}^{K} f(c_i = k | \mathbf{X}_i) \prod_{j=1}^{J} \pi_{j|k}^{Y_{ij}} (1 - \pi_{j|k})^{1-Y_{ij}}. \tag{6}$$

Bandeen-Roche, et al. 1997 assumed a generalized logit model for $f(c_i = k|\mathbf{X}_i)$, i.e.

$$Pr(c_i = k|\mathbf{X}_i) = exp(\mathbf{X}_i^{oT}\boldsymbol{\beta}_k) / \sum_{k=1}^{K} exp(\mathbf{X}_i^{oT}\boldsymbol{\beta}_k) \;\; k = 1, \ldots, K,$$

where $\mathbf{X}_i^o = (1, X_{i1}, \ldots, X_{iP})^T$ and $\boldsymbol{\beta}_k = (\beta_{0k}, \beta_{1k}, \ldots, \beta_{Pk})^T$ are $P+1$ dimensional vectors. In the above equation, the K^{th} class is taken as the reference class, i.e. $\boldsymbol{\beta}_K = (0, 0, \ldots, 0)^T$. Then, similar to the latent class model, given the number of classes K fixed, estimates can be obtained by the EM algorithm.

3. Latent Class Regression on Latent Factors model

In the previous section, we have introduced three models that construct relationships between latent variables and observed variables. As an extension to these models, we will build on the idea of latent class regression and introduce a new model that considers the structural relationship between a latent class variable and latent factors.

Let c_i be a latent class variable with K categories and $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ})^T$ be a J dimensional observed vector with binary elements, which is used to measure c_i via the latent class model (5). Let $\mathbf{X}_i = (X_{i1}, \ldots, X_{iP})^T$ be a P dimensional observed vector with continuous elements used to measure a Q dimensional latent factor $\mathbf{f}_i$ via the latent factor model (2) or more generally (3).

Similar in spirit to latent class regression model (6), our new model considers regressing the latent class c_i on the latent factors $\mathbf{f}_i$. One of the fundamental assumptions of this new model is that $\mathbf{Y}_i$ is conditionally independent of $\mathbf{X}_i$ given the latent variables c_i and $\mathbf{f}_i$. This means that the model assumes $\mathbf{Y}_i$ and $\mathbf{X}_i$ are only related because the variables they are measuring are related. This is a natural assumption when modeling relationships between variables measured with error, i.e. we want to model the relationship between the underlying variables, not the ones with error. Thus we consider the joint distribution of $\mathbf{Y}_i$ and $\mathbf{X}_i$ to be

$$f(\mathbf{Y}_i, \mathbf{X}_i) = \sum_{k=1}^{K} \int f(\mathbf{Y}_i|c_i = k) f(\mathbf{X}_i|\mathbf{f}_i) f(c_i = k|\mathbf{f}_i) f(\mathbf{f}_i) d\,\mathbf{f}_i, \tag{7}$$

where

$$\begin{array}{rcll} f(\mathbf{Y}_i|c_i=k) & = & \prod_{j=1}^{J} \pi_{j|k}^{Y_{ij}}(1-\pi_{j|k})^{1-Y_{ij}} & \text{(latent class model)} \\ f(\mathbf{X}_i|\mathbf{f}_i) & \sim & N_p\left(\left(\begin{array}{c}\boldsymbol{\lambda}_0\\ \mathbf{0}\end{array}\right)+\left(\begin{array}{c}\boldsymbol{\Lambda}\\ \mathbf{I}\end{array}\right)\mathbf{f}_i, \boldsymbol{\Psi}\right) & \text{(latent factor model)} \\ f(c_i=k|\mathbf{f}_i) & = & \pi_k(\mathbf{f}_i)=\frac{exp(\mathbf{f}_i^{oT}\boldsymbol{\beta}_k)}{\sum_{k=1}^{K} exp(\mathbf{f}_i^{oT}\boldsymbol{\beta}_k)} & \text{(generalized logit model)} \\ f(\mathbf{f}_i) & \sim & F(\boldsymbol{\Phi}) & \text{(specified distribution for latent factors).} \end{array} \tag{8}$$

The parameters $\pi_{j|k}$, $\boldsymbol{\lambda}_0$, $\boldsymbol{\lambda}_1$ and $\boldsymbol{\Psi}$ are all as defined before in section 2. While most commonly the distribution $F(\boldsymbol{\Phi})$ for the underlying factors will be normal, it is not necessary so we present it here to be any specified distribution with unknown parameters $\boldsymbol{\Phi}$. Note that like the latent class regression model in section 2.3, we use the generalized logit link, i.e. $\log\left(\frac{\pi_k(\mathbf{f}_i)}{\pi_K(\mathbf{f}_i)}\right)=\mathbf{f}_i^{oT}\boldsymbol{\beta}_k$, where $\mathbf{f}_i^o=(1,\mathbf{f}_i)^T$, $\boldsymbol{\beta}_k=(\beta_{0k},\beta_{1k},\ldots,\beta_{Qk})^T$ and $\boldsymbol{\beta}_K=(0,0,\ldots,0)^T$ indicating class K as the reference class. The parameter $\boldsymbol{\beta}_k$ is a $Q+1$ dimensional vector relating the latent factors to the probability of being in a particular latent class. The big difference between this model and that in section 2.3 is that here $\mathbf{f}_i$ is not observed directly. In this model, the parameters of interest are the vector $\boldsymbol{\beta}_k$'s, which describe the relationship between the latent factors and the latent classes.

4. Maximum Likelihood Estimation

Given the parametric model (7-8) and the *i.i.d.* data $(\mathbf{Y}_i,\mathbf{X}_i)$, for $i=1,\ldots,n$, estimation of the model parameters can proceed via the maximum likelihood method. Let $\mathbf{Z}_i=(\mathbf{Y}_i,\mathbf{X}_i)$, $\mathbf{d}_i=(c_i,\mathbf{f}_i)$, and $\boldsymbol{\theta}=(\{\pi_{j|k}\},\boldsymbol{\lambda}_0,\boldsymbol{\Lambda},\boldsymbol{\Psi},\{\boldsymbol{\beta}_k\},\boldsymbol{\Phi})$ is the vector of parameters relating $\mathbf{Z}_i$ and $\mathbf{d}_i$. Thus the likelihood function for the model (7-8) can be written as

$$L_o=\prod_{i=1}^{n} f(\mathbf{Z}_i;\boldsymbol{\theta})=\prod_{i=1}^{n}\int f(\mathbf{Z}_i,\mathbf{d}_i;\boldsymbol{\theta})d\,\mathbf{d}_i, \tag{9}$$

where the notation for the integral over $\mathbf{d}_i$ is taken very generally to include the continuous integral for $\mathbf{f}_i$ and the summation over c_i. This likelihood function is hard to maximize due to the integration of the latent variables for which there is no closed form solution. Hence two numerical methods for performing the full maximum likelihood are described in this section: Monte Carlo Expectation and Maximization algorithm (MCEM) and Gaussian quadrature followed by quasi-Newton algorithm.

4.1 MCEM Algorithm

It is natural to consider the latent variables, $\mathbf{d}_i$, as missing data and implement the EM algorithm for maximizing (9). Since it is hard to maximize the observed data likelihood L_o directly, we construct the complete data likelihood and apply the EM algorithm to maximize it. The complete data likelihood is

$$L_c = \prod_{i=1}^{n} f(\mathbf{Z}_i, \mathbf{d}_i; \boldsymbol{\theta}).$$

The E-step obtains the expectation of the log complete data likelihood given the observed data and the current parameter estimates, i.e. $\boldsymbol{\theta}_l$.

$$\begin{aligned} E(\log L_c|\mathbf{Z}_1, \ldots, \mathbf{Z}_n; \boldsymbol{\theta}_l) &= \sum_{i=1}^{n} \int \log f(\mathbf{Z}_i, \mathbf{d}_i; \boldsymbol{\theta}) f(\mathbf{d}_i|\mathbf{Z}_i; \boldsymbol{\theta}_l) d\,\mathbf{d}_i \\ &\equiv g_{\boldsymbol{\theta}_l}(\boldsymbol{\theta}; \mathbf{Z}). \end{aligned}$$

For the latent class regression on latent factor model (7-8), unfortunately we do not have a closed form for $f(\mathbf{d}_i|\mathbf{Z}_i; \boldsymbol{\theta}_l)$ and consequently we do not have a closed form solution for the integral in $g_{\boldsymbol{\theta}_l}(\boldsymbol{\theta}; \mathbf{Z})$. Hence, we propose to use the Monte Carlo method to obtained an approximation to $g_{\boldsymbol{\theta}_l}(\boldsymbol{\theta}; \mathbf{Z})$. First note that

$$\begin{aligned} g_{\boldsymbol{\theta}_l}(\boldsymbol{\theta}; \mathbf{Z}) &= \sum_{i=1}^{n} \int \log f(\mathbf{Z}_i, \mathbf{d}_i; \boldsymbol{\theta}) f(\mathbf{d}_i|\mathbf{Z}_i; \boldsymbol{\theta}_l) d\,\mathbf{d}_i \\ &= \sum_{i=1}^{n} \int \log f(\mathbf{Z}_i, \mathbf{d}_i; \boldsymbol{\theta}) \frac{f(\mathbf{Z}_i|\mathbf{d}_i; \boldsymbol{\theta}_l)}{\int f(\mathbf{Z}_i|\mathbf{d}_i; \boldsymbol{\theta}_l) f(\mathbf{d}_i; \boldsymbol{\theta}_l) d\,\mathbf{d}_i} f(\mathbf{d}_i; \boldsymbol{\theta}_l) d\,\mathbf{d}_i \\ &= \sum_{i=1}^{n} E\left(\log f(\mathbf{Z}_i, \mathbf{d}_i; \boldsymbol{\theta}_l) \frac{f(\mathbf{Z}_i|\mathbf{d}_i; \boldsymbol{\theta}_l)}{\int f(\mathbf{Z}_i|\mathbf{d}_i; \boldsymbol{\theta}_l) f(\mathbf{d}_i; \boldsymbol{\theta}_l) d\,\mathbf{d}_i} \right), \end{aligned}$$

where the expectation is taken with respect to the random variable $\mathbf{d}_i$. Given the current $\boldsymbol{\theta}_l$, a Monto Carlo sample $(\mathbf{d}_i^1, \ldots, \mathbf{d}_i^M)$ is generated from $f(\mathbf{d}_i; \boldsymbol{\theta}_l)$, then the expectation can be approximated by an average

$$g_{\boldsymbol{\theta}_l}(\boldsymbol{\theta}; \mathbf{Z}) \approx \sum_{i=1}^{n} \frac{1}{M} \sum_{m=1}^{M} [\log f(\mathbf{Z}_i, \mathbf{d}_i^m; \boldsymbol{\theta})\ W_i^m] \equiv g_{\boldsymbol{\theta}_l}^{MC}(\boldsymbol{\theta}; \mathbf{Z}),$$

where $W_i^m = \frac{f(\mathbf{Z}_i|\mathbf{d}_i;\boldsymbol{\theta}_l)}{\frac{1}{M}\sum_{m=1}^{M} f(\mathbf{Z}_i|\mathbf{d}_i^m;\boldsymbol{\theta}_l)}$.

Note that the same Monte Carlo sample is used for evaluate the integral in the denominator of the weights in the expectation. Now we note that $\log f(\mathbf{Z}_i, \mathbf{d}_i^m; \boldsymbol{\theta})$ can be factorized into four parts corresponding to the four parts of the model (7-8), i.e.

$$\begin{aligned} \log f(\mathbf{Z}_i, \mathbf{d}_i^m; \boldsymbol{\theta}) &= \log f(\mathbf{Y}_i | c_i^m; \{\pi_{j|k}\}) + \log f(\mathbf{X}_i | \mathbf{f}_i^m; \boldsymbol{\lambda}_0, \boldsymbol{\Lambda}, \boldsymbol{\Psi}) \\ &+ \log f(c_i^m | \mathbf{f}_i^m; \{\boldsymbol{\beta}_k\}) + \log f(\mathbf{f}_i^m; \boldsymbol{\Phi}). \end{aligned} \tag{10}$$

The M-step is to maximize $g_{\boldsymbol{\theta}_l}(\boldsymbol{\theta}; \mathbf{Z})$ with respect to $\boldsymbol{\theta}$ and then update $\boldsymbol{\theta}_l$. Based on (10), we see that $g_{\boldsymbol{\theta}_l}(\boldsymbol{\theta}; \mathbf{Z})$ has four parts with distinct parameters associated with each. Hence, we can maximize each component separately as a straightforward weighted regression (weighted by W_i^m) in order to obtain $\boldsymbol{\theta}_{l+1}$.

The MCEM algorithm will iterate between the E-step and M-step until the parameter estimates converge according to some criteria. In order to decrease the Monte Carlo error at the E-step, a large M should be used. But it has been pointed out that it is inefficient to choose a large M when $\boldsymbol{\theta}_l$ is far from the ML estimate (Wei and Tanner, 1990; Booth and Hobert, 1999). Following recommendation, it is preferable to start with a small M and increase it for each iteration to $M_l = M_0 + Tl$, where M_l is the sample size for the Monte Carlo step at the l^{th} iteration and M_0 and T are positive constants. We monitor the convergence of the EM algorithm by plotting $\boldsymbol{\theta}_l$ versus the iteration l.

Standard Error Estimates of the parameter estimates from MCEM can be obtained by inverting the information matrix of the log likelihood function based on the observed data. We apply Louis' formula (Louis, 1982)

$$I_{\mathbf{Z}}(\boldsymbol{\theta}) = E_{\mathbf{d}}\left(-\frac{\partial^2 L_c(\mathbf{Z}, \mathbf{d}; \boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right) - Var_{\mathbf{d}}\left(\frac{\partial L_c(\mathbf{Z}, \mathbf{d}; \boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right),$$

evaluated at the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$. The expectation and variance are taken with respect to the conditional distribution of the latent variable $\mathbf{d} = (\mathbf{d}_1, \ldots, \mathbf{d}_n)^T$ given the observed data $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)^T$ and parameter $\boldsymbol{\theta}$. These conditional expectations are difficult to evaluate because as before in the EM algorithm the conditional distribution of $\mathbf{d}$ given $\mathbf{Z}$ is unavailable. Hence we cannot sample from the conditional distribution or get the closed forms for the expectations. By a similar method described for the parameter estimation, we can switch the conditional expectation to a weighted unconditional expectation with respect to $\mathbf{d}$ and then use the Monte Carlo method to approximate the expectation. Here a large M is used as it is unnecessary to iterate.

4.2 Gaussian quadrature with quasi-Newton algorithm

We note that the MCEM algorithm introduced above is flexible with regard to the assumptions of the distribution of the latent factors. That is, it was not necessary to assume $\mathbf{f}_i$ as normally distributed. Consider again the likelihood function L_o associated with model (7-8). Because the latent classes are discrete, it can be written as

$$L_o = \prod_{i=1}^{n} \sum_{k=1}^{K} \int f(\mathbf{Y}_i|c_i = k) f(\mathbf{X}_i|\mathbf{f}_i) f(c_i = k|\mathbf{f}_i) f(\mathbf{f}_i) d\,\mathbf{f}_i.$$

We note that the observed data likelihood is a function of the integral of the latent factors $\mathbf{f}_i$. In the special case, when the $\mathbf{f}_i$ is normally distributed, this can be approximated by adaptive Gaussian quadrature method (Golub and Welsch 1969, or Table 25.10 of Abramowitz and Stegun 1972). Then given a closed form approximation to the integral involving the normal factors $\mathbf{f}_i$, the observed likelihood can then be approximate in a closed form. With the closed form approximation for the likelihood, the maximization of it can be carried out through a quasi-Newton algorithm.

In fact, this method of Gaussian quadrature approximation followed by quasi-Newton maximization can be implemented using the “general” likelihood function in PROC NLMIXED in SAS. Appendix A gives code demonstrating how this can be done.

Although detailed investigation of the computational speed and accuracy of this method as compared to MCEM is beyond the scope of the current paper, the estimation for the example considered herein takes 5 times longer using MCEM. It should also be noted that for increasing numbers of factors, the integration in both methods may be computationally prohibitive.

5. Simulation Study

5.1 Simulation

In order to investigate the behavior of the new model under different scenarios, we will do a small simulation study. We are particularly interested in examing the effect that sample size and reliability of the measurement of both $\mathbf{f}_i$ and c_i have on the inference for the $\{\boldsymbol{\beta}_k\}$ parameters in model (7-8). By reliability we are refereing to the amount of measurement error in the latent class model and latent factor model in (8). Consider the model with $P = 4$ dimensional observed $\mathbf{X}_i$ variable measuring $Q = 1$ latent factor; $J = 5$ dimensional observed $\mathbf{Y}_i$ variable measuring $K = 2$ latent classes. Furthermore, assume that the underlying factor f_i is normally distributed. Specifically, we consider,

$$\begin{aligned}
f_i &\overset{i.i.d}{\sim} N(0,1) \\
\epsilon_{\mathbf{i}} &\overset{i.i.d}{\sim} N_4(0, 0.5I_{4\times4}) \\
\begin{pmatrix} X_{i1} \\ X_{i2} \\ X_{i3} \\ X_{i4} \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \lambda_{11} \\ \lambda_{12} \\ \lambda_{13} \\ \lambda_{14} \end{pmatrix} f_i + \epsilon_{\mathbf{i}} \\
c_i | f_i &\sim Bernoulli(logit^{-1}(0.5 + 1 \cdot f_i)) \\
Y_{ij} | c_i &\sim Bernoulli(\pi_{j|k}) \;\; k = 1, 2; \;\; j = 1, \ldots, 5.
\end{aligned} \tag{11}$$

Consider two different sets of factor loadings for the latent factor. First we generate the data with $(\lambda_{11}, \lambda_{12}, \lambda_{13}, \lambda_{14})^T = (0.4, 0.4, 0.5, 0.6)^T$, which implies that f_i has low reliability (< 0.6) and then we choose $(\lambda_{11}, \lambda_{12}, \lambda_{13}, \lambda_{14})^T = (1.5, 1.6, 1.0, 1.7)^T$ so that f_i has high reliability (> 0.9). Furthermore, we consider two different sets of values for $\pi_{j|1}$ and $\pi_{j|2}$ where $j = 1, \ldots, 5$, which describe the way that the observed data Y_{ij} are related to the latent classes. One case considers where $\pi_{j|1} = 0.1$ and $\pi_{j|2} = 0.8$ for all j. This implies that when a person is in class 1 he has a low probability, i.e. 0.1, of responding "yes" to each of the five questions, and when being in class 2 he has a high probability, i.e. 0.8, of responding "yes" to each of the five questions. We refer to this as the parallel probabilities case since the probabilities for each response Y_{ij} in two latent classes are the same and with big differences. The other case we consider is where three of the five observed variables have $\pi_{j|1} = 0.1$ and $\pi_{j|2} = 0.8$ ($j = 1, 2, 3$) but the other two variables (i.e. $j = 4, 5$) have $\pi_{j|1} = 0.2$ and $\pi_{j|2} = 0.3$. Note that the probabilities of being 1 for the three corresponding elements of $\mathbf{Y}_i$ are quite different (0.1 vs. 0.8) and the other two are similar (0.2 vs. 0.3). We refer to this as the non-parallel case and would expect it to be a less precise measurement model for c_i. Finally, we consider 2 sample sizes, $n = 200$ and 2000. So in total we have $2 \times 2 \times 2 = 8$ different scenarios. For each scenario, we generate 1000 data sets; to each of the simulated data sets in each of the eight different scenarios, we perform maximum likelihood as described in section 4. Specifically, because the factor is normally distributed the adaptive Gauss quadrature method for approximating the observed data likelihood and quasi-Newton optimization as the optimization technique is implemented via PROC NLMIXED in SAS.

5.2 Results

The simulation results are shown in Table 2. From this table, we find that as expected when the sample size increases from 200 to 2000, the bias and the standard errors are decreasing. We see that in all scenarios for both β_0 (the intercept) and β_1 (the slope) the high reliability of f case shows smaller standard errors than the respective low reliability case. Likewise since the parallel case can be consider more precise than the nonparallel case, we find smaller standard errors for both β_0 and β_1 for the parallel case as compared to the nonparallel case. What is more interesting is the relative impact each of these on the estimates of β_0 and β_1. The impact that the reliability of f has on the efficiency of β_1 is greater than the impact it has on β_0. On the other hand, the impact that the parallel vs. nonparallel has on the efficiency of β_0 is much more substantial than that on β_1. This suggests that the slope of the relationship between c and f is more sensitive to the model for f, whereas the intercept is more sensitive to the model for c. The coverage probability in

Table 2 is calculated in the following way. For each data set we obtain the 95% confidence intervals of the parameters based on the estimated standard errors output from PROC NLMIXED. The coverage probability is the number of the confidence intervals covering the true value over 1000. In all scenarios, the coverage probabilities are close to 95%.

6. Example

Project EAT (Neumark-Sztainer, et al., 2002) is a comprehensive study of adolescent nutrition and obesity. Self-report survey data were collected from students in 7^{th} and 10^{th} grade at 31 Twin Cities schools in the 1998-99 school year. A portion of the survey data collected for the 2113 girls in Project EAT is used in this current example. The particular relationship of interest in this study is that between body satisfaction (a latent factor) and eating disorders risk (a latent class).

In the Project EAT data set, for each individual i ($i = 1, \ldots, 2113$) let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{i9})^T$ be the 9 dichotomous questionnaire items indicating the self-reported use of unhealthy weight control behaviors within the past year (i.e. "Have you done any of the following things in order to lose weight or keep from gaining weight during the past year: fasting, eating very little food, taking diet pills, making myself vomit, using laxatives, using diuretics, using food substitute, skipping meals, smoking more cigarettes?"). Furthermore, let $\mathbf{X}_i = (X_{i1}, \ldots, X_{i5})^T$ indicate the five items measuring body satisfaction (i.e. "How satisfied are you with your: body shape, waist, hips, thigh, stomach?"). Each element was measured on a 5 point Likert scale where the anchors were 1="very dissatisfied" and 5="very satisfied". Despite the discrete nature of these Likert responses, we will treat $\mathbf{X}_i$ as a continuous variable in this data analysis and center each element to mean zero.

6.1 Exploratory data analysis

Assume that underlying the observed responses $\mathbf{Y}_i$ is a latent class variable c_i with categories representing different typologies of eating disorders risk. In practice, we will not know the "correct" number of latent classes in the model. The number of latent classes K needs to be investigated before fitting the relationship between latent variables. Here we present the exploratory latent class analysis of the 9 observed indicators asking which unhealthy weight control behaviors had been used within the past year. Table 3 shows the estimated latent class model parameters and associated BIC values. The 3-class model shows the best BIC fit value. Examining the $\{\pi_{j|k}\}$ for the 3-class model leads to a class of girls who are basically not doing any of the behaviors (56.4%), a class who are doing just the restricting behaviors (i.e. eating very little and skipping meals) (35.2%), and a high risk class who have high probability of doing everything (8.4%). The 4-class model reveals a classification worth discussing. In the 4-class model, the high risk group has been split into girls who are more likely to be using external substances (i.e. diet pills, laxatives, diuretics, food substances) to lose weight separate from girls who are restricting food intake and vomiting as well as smoking cigarettes to lose weight. This 4-class model did not match the researchers' theory as well as the 3-class model and since the 3-class model empirically fitted the best, it will be used for the latent class regressed on a latent factor model.

Now we explore the observed body satisfaction variables $\mathbf{X}_i$ as measurements of a latent factor f_i. The researchers hypothesize that these questions are measuring one dimension of body satisfaction. The correlations between the variables in $\mathbf{X}_i$ range between 0.57 and 0.75. The eigenvalues of the covariance matrix are (3.703, 0.476, 0.362, 0.261, 0.198), which indicates that 1-dimension is well described by these variables providing empirical support for the 1-factor model. Thus, we will consider the body satisfaction, a 1-dimensional continuous latent factor f_i underlying the observed $\mathbf{X}_i$.

6.2 Model Fitting

Consider the parametric model (7-8), where $P = 5$, $Q = 1$, $J = 9$ and $K = 2, 3$ for the example data set. Note, although the 3 latent class outcome model was chosen based on exploratory data analysis, the 2 latent class outcome is shown as comparison. Table 4 shows the parameter estimates for different models, the standard errors of estimates and the p-values for each parameter where the "low" eating disorders risk class is treated as the reference class 0. The AIC and BIC values indicate the model with 3-class outcome fits the data better. The estimates of the log ORs for class 1 (β_{11}) and class 2 (β_{12}) are negative and statistically significant, which are interpreted as the effect of a 1-unit increase in body satisfaction on the log odds of being in class $k(k = 1, 2)$ rather than class 0. It makes sense that these are negative since as a girl's satisfaction with body increases, she would be less likely to be in one of the high eating disorders risk classes. The intercepts represent the log odds of being in class k rather than class 0 for a girl with body satisfaction at the center of the scale, i.e. 0, since the elements of $\mathbf{X}_i$ have been centered. These are related to the overall prevalence in each class where we see that the prevalence in class 2, the high risk class, is small.

7. Discussion

This paper proposes a new model for fitting the relationship between a latent class outcome and latent factor predictors. It is demonstrated that maximum likelihood estimation is possible by the MCEM algorithm, or in the special case where the factors are normally distributed and Q is small by Gaussian quadrature and quasi-Newton. The model presented in this paper is a natural extension of the latent class regression model and more generally is an extension of structural equation modeling. Structural equation modeling focuses on the relationships among latent variables, but almost exclusively continuous latent variables. Here we present a method for examining the relationship among latent variables that are of mixed types.

Similar to a common practice in structural equation modeling, we recommend that the latent class regressed on latent factor model be built in steps. In particular the measurement models (i.e. the latent class and the latent factor models) can be examined first to assess the fit of different numbers of classes or factors. Then once these measurement models are settled upon, the "structural model" i.e. the relationship among the latent variables can be modeled simultaneously with the measurement models. This is how we presented the method for the Project EAT example and feel this allows for appropriately focused model checking.

In the simulation study and the example, we only considered models where the underlying latent factor was assumed to be normally distributed. Although other distribution could be considered, it is not clear that this modeling choice can be checked as the latent factors are not directly observable. Rather, methods that only require weak assumptions on the distribution of the underlying factors should be developed in the future.

Finally, identifiability of model parameters is always an important issue in latent variable models. In our proposed model, the parameters are well identified theoretically when the latent class model and the latent factor model satisfy appropriate conditions. But identifiability is not a property of just the model but also of the data combined with the model. A model may have weak data identifiability if perhaps due to small sample size, some parameters are difficult or even impossible to estimate. While generally it is hard to assess this data identifiability problem, some work has been done to quantify it for the latent class model within a fully Bayesian framework by comparing the prior and posterior of parameters (Garrett and Zeger, 2000). Similar methods might be possible to consider for the model presented here.

References

Abramowitz, M. and Stegun, I.A. (1972). *Handbook of Mathematical Functions*, New York: Dover Publications, Inc.

Anderson T.W., Amemiya Y. (1998). The asymptotic normal distribution of estimators in factor analysis under general conditions. *The Annals of Statistics*, 16, 759-771.

Bandeen-Roche, K., Miglioretti, D., Zeger, S., Rathouz, P (1997). *JASA*, 92, No. 440,1375-1386.

Bartholomew, D.J., and Knott, M. (1999). Latent Variable Models and Factor Analysis, 2nd. ed., Kendall's Library of Statistics

Bollen, K.A. (1989). *Structural Equations with Latent Variables.* New York, Wiley.

Booth, J.G. and Hobert, J.H. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm, *Journal of the Royal Statistical Society* Series B 62: 265-285.

Bowling, A. (1997). *Measuring health: A review of quality of life measurement scales.* Open University Press.

Browne, M.W., Shapiro A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Psychology*, 41, 193-208.

Burr, D. (1988). On Errors-in-Variables in Binary Regression-Berkson Case. *JASA*, 83,739-743

Carroll, R.J., Ruppert, D., and Stefanski, L.A. (1995), *Measurement Error in Nonlinear Models*, Chapman & Hall, London.

Clogg, C.C. (1981). New developments in latent structure analysis. In D.J. Jackson and E.F. Borgotta (eds.) *Factor Analysis and Measurement in Sociological Research.* Beverly Hills: Sage.

Clogg, C.C. (1995). Latent class models. In G. Arminger, C.C. Clogg, and M.E. Sobel (eds.) *Handbook of Statistical Modeling for the Social and Behavioral Sciences.* New York: Plenum Press, pp. 311-59.

Collins, L.M., Fidler, P.L., Wugalter, S.E., Long, J. (1993). Goodness-of-fit testing for latent class models. *Multivariate Behavioral Research*, 28, 375-389.

Croudace, T.J., Jarvelin, M.R., Wadsworth, M. and Jones, P.B. (2003). Developmental typology of trajectories to nighttime bladder control: Epidemiologic applications of longitudinal latent class analysis. *American Journal of Epidemiology*, 157, No. 9, 834-842.

Dayton, M., and Macready, G. (1988). Concomitant-variable latent class models. *JASA*, 83, 173-178.

Dunson, D.B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society* Series B, 62, 355-366.

Dunson, D.B. (2003). Dynamic latent trait models for multidimensional longitudinal data. *JASA*, 98, No. 463 555-563.

Flaherty, B.P. (2002). Assessing reliability of categorical substance use measures with latent class analysis, *Drug and Alcohol Dependency*, 68, Suppl 7-20.

Fuller, W. (1987). *Measurement Error Models.* New York: John Wiley and Son.

Jöreskog, K.G. and Sorborm, D. (1996). *LISREL8: User's Reference Guide*, Chicago: Scientific Software International.

Hagenaars, J.A. and McCutcheon, A.L. (eds.) (2002) *Applied Latent Class Analysis Models.* Cambridge University Press.

Garrett, Elizabeth S. and Zeger, Scott L. (2000). Latent Class model Diagnosis. *Biometrics* Vol 56, 971-1295.

Golub, G.H., and Welsch, J.H. (1969). Calculation of Gaussian Quadrature Rules. *Mathematical Computing*, 23, 221 -230.

Lawley, D.N., and Maxwell, A.E. (1971). *Factor Analysis as a Statistical Method*, 2nd ed. London: Butterworths.

Lazarsfeld, P.F., and Henry, N.W. (1968). *Latent Structure Analysis.* Boston, MA: Houghton Mifflin.

Lee NK, Oei, TPS, Greeley, JD, Baglioni, AJ. (2003). Psychometric properties of the drinking expectancy questionnaire: A review of the factor structure and a proposed new scoring method. *Journal of Studies on Alcohol*, 64(3): 432-436.

Louis, T. (1982). Finding the observed Information Matrix when Using the EM Algorithm. *Journal of the Royal Statistical Society.* Series B, Vol.44, No.2.

McLachlan, G.J., and Peel, G. (2000). *Finite Mixture Models.* New York, Wiley.

Moustaki, I. and Knott, M. (2000). Generalized latent trait models. *Psychometrika*, 65, 391-411.

Neumark-Sztainer, D., Story, M., Hannan, P., and Moe, J. (2002). Overweight status and eating patterns among adolescents: Where do youth stand in comparison to the Healthy People 2010 Objectives? *Am J Public Health*, 92: 844-51.

Neumark-Sztainer, D., Wall, M.M., Story, M., and Perry, C. (2003a). Correlates of unhealthy weight control behaviors among adolescent girls and boys: Implications for the primary prevention of disordered eating. *Health Psychology*, Vol 22, No. 1, 88-98.

Neumark-Sztainer, D., Wall, M.M., Perry, C., Story, M. (2003b). Correlates of fruit and vegetable intake among adolescents: Findings from Project EAT, accepted and to appear *Preventive Medicine.*

Sammel, M., Ryan, L., and Legler, J. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society* Series B, 59, 667-678.

Uebersax, J.S. and Grove, W.M. (1990). Latent class analysis of diagnostic agreement. *Statistics in Medicine*, 9:559-572.

Wei, G.C.G. and Tanner, M.A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *JASA*, 85, 699-704.

Appendix A

```
**** Data (dat) is generated as in simulation study from model (11);
**** Example of data frame;
id      x1           x2            x3            x4      y1 y2 y3 y4 y5  dummy
1 -0.417476996 -0.103693622  -0.596811257  -0.827308222  0  1  1  1  1    1
2 -1.113301965 -0.098232232   0.2814094176 -0.577613842  0  0  1  1  0    1
3 -0.395122307  0.2969895587  0.1303007521  0.5346021666 0  0  1  0  1    1
4  0.375280428  1.1090689506 -0.576731743   0.1241954368 1  1  1  0  0    1
5 -0.363147587 -0.86985922   -1.039424695  -1.147074384  1  0  1  1  1    1
                           ... ...
                           ... ...
**** Note the X variables are centered (not standardized);
proc standard data = dat mean = 0 out = dat1;
     var x1-x5;
run;

**** PROC NLMIXED for the fitting model (7-8);
proc nlmixed data = dat1 tech = quanew lis = 2 method = gauss
     maxiter = 1000 gconv = .00000000001 fconv = .00000000001;
**** starting values (starting values are given for the logit(pi_i|j);
parms
     beta0 = 0.5  beta1 = 1
     lam11 = 1    lam12 = 1     lam13 = 1
     psi1  = 0.5  psi2  = 0.5   psi3  = 0.5   psi4  = 0.5
     lpi11 = 1    lpi12 = 1     lpi13 = 0     lpi14 = 0     bpi15 = -.5
     bpi21 = 0.5  bpi22 = 0.5   bpi23 = -0.5  bpi24 = -0.5  bpi25 = -1;
bounds -6 <= bpi11-bpi15 bpi21-bpi25 <= 6;

**** latent class part;
pi11 = 1/(1+exp(-bpi11)); pi21 = 1/(1+exp(-bpi21));
pi12 = 1/(1+exp(-bpi12)); pi22 = 1/(1+exp(-bpi22));
pi13 = 1/(1+exp(-bpi13)); pi23 = 1/(1+exp(-bpi23));
pi14 = 1/(1+exp(-bpi14)); pi24 = 1/(1+exp(-bpi24));
pi15 = 1/(1+exp(-bpi15)); pi25 = 1/(1+exp(-bpi25));
```

```
prod11 = (pi11**y1)*(1-pi11)**(1-y1); prod21 = (pi21**y1)*(1-pi21)**(1-y1);
prod12 = (pi12**y2)*(1-pi12)**(1-y2); prod22 = (pi22**y2)*(1-pi22)**(1-y2);
prod13 = (pi13**y3)*(1-pi13)**(1-y3); prod23 = (pi23**y3)*(1-pi23)**(1-y3);
prod14 = (pi14**y4)*(1-pi14)**(1-y4); prod24 = (pi24**y4)*(1-pi24)**(1-y4);
prod15 = (pi15**y5)*(1-pi15)**(1-y5); prod25 = (pi25**y5)*(1-pi25)**(1-y5);

**** relation between latent class and latent factor (structural model);
eta1=exp(beta0+beta1*fi)/(1+exp(beta0+beta1*fi));
eta2=1/(1+exp(beta0+beta1*fi));

**** the part of the likelihood coming from latent class part;
l_latclass=eta1*prod11*prod12*prod13*prod14*prod15
          +eta2*prod21*prod22*prod23*prod24*prod25;
ll_latclass = log(l_latclass);

**** factor analysis part;
mu1 = lam11*fi; mu2 = lam12*fi;
mu3 = lam13*fi; mu4 =     1*fi;

**** the part of the likelihood coming from latent factor part;
ll_factpart = -.5*log(psi1) - (1/(2*psi1)) * (x1 - mu1)**2
              -.5*log(psi2) - (1/(2*psi2)) * (x2 - mu2)**2
              -.5*log(psi3) - (1/(2*psi3)) * (x3 - mu3)**2
              -.5*log(psi4) - (1/(2*psi4)) * (x4 - mu4)**2;

**** dummy is just a place holder so that SAS has something on;
**** the left side of equation;
model dummy ~ general(ll_latclass + ll_factpart);
random fi ~ normal(0,phi) subject = id;
run;
```

Table 1: Modeling Continuous Predictors and Categorical Outcome

	Predictors Observed	Predictors Latent
Outcome Observed	Logistic Regression	Error-in-Variable Technique
Outcome Latent	Latent Class Regression	**Current Study**

Table 2: Simulation Results

				Reliability of f					
				Low (< 0.6)			High (> 0.9)		
	$\{\pi_{j\|1}\}$ $\{\pi_{j\|2}\}$	Sample Size	True Value	Mean	SE	Coverage Probability	Mean	SE	Coverage Probability
β_0	Parallel	200	0.5	0.5103	0.1897	0.9560	0.5067	0.1826	0.9570
		2000	0.5	0.5005	0.0589	0.9390	0.5002	0.0571	0.9370
	Non-parallel	200	0.5	0.5186	0.2559	0.9310	0.5129	0.2482	0.9290
		2000	0.5	0.5002	0.0746	0.9540	0.4998	0.0720	0.9480
β_1	Parallel	200	1	1.0360	0.2820	0.9450	1.0186	0.2191	0.9510
		2000	1	1.0028	0.0823	0.9520	1.0016	0.0666	0.9430
	Non-parallel	200	1	1.0482	0.3148	0.9420	1.0296	0.2491	0.9530
		2000	1	1.0033	0.0867	0.9570	1.0019	0.0705	0.9530

Table 3: Estimated $\pi_{j|k}$ (probability of saying yes to the variable j given that the individual is in latent class k) under latent class models with different K

	marginal	2-class		3-class			4-class			
To control weight	1	0	1	0	1	2	0	1	2	3
fasted	17.9	2.8	38.8	2.6	32.6	58.5	2.6	29.2	71.4	24.9
ate little	44.1	9.0	92.5	7.9	89.9	94.2	6.9	87.5	100.0	74.1
diet pills	6.3	1.1	13.6	1.2	6.5	40.4	0.8	6.1	31.0	49.4
vomit	6.3	0.1	15.0	0.1	7.1	45.0	0.1	5.6	43.6	33.2
laxatives	1.6	0.2	3.5	0.2	0.1	17.1	0.2	0.0	11.7	20.7
diuretics	1.4	0.1	3.3	0.1	0.1	16.1	0.1	0.0	7.9	29.4
food substitutes	9.3	2.1	19.2	2.1	13.1	41.6	1.7	12.0	34.5	54.4
skipped meals	44.4	11.1	89.7	9.5	89.7	85.7	8.5	87.4	100.0	43.5
smoked more cigs	9.3	2.5	18.6	2.5	13.1	39.1	2.4	11.2	47.6	9.8
percent in each class	100%	58.0%	42.0%	56.4%	35.2%	8.4%	54.7%	34.7%	8.0%	2.6%
BIC		11131.3		11044.4			11072.7			

Table 4: Estimation results for Project EAT data

Model	Parameter	Estimate	SE	P-value	AIC	BIC
2-class	β_{01}	-0.3998	0.0654	$< .0001$	20180	20350
	β_{11}	-1.0181	0.0636	$< .0001$		
3-class	β_{01}	-0.9523	0.0924	$< .0001$	20062	20294
	β_{02}	-1.2006	0.1081	$< .0001$		
	β_{11}	-0.9208	0.0804	$< .0001$		
	β_{12}	-1.1472	0.0927	$< .0001$		