3

LOG-MULTIPLICATIVE ASSOCIATION MODELS AS LATENT VARIABLE MODELS FOR NOMINAL AND/OR ORDINAL DATA

*Carolyn J. Anderson**
Jeroen K. Vermunt†

Associations between multiple discrete measures are often due to collapsing over other variables. When the variables collapsed over are unobserved and continuous, log-multiplicative association models, including log-linear models with linear-by-linear interactions for ordinal categorical data and extensions of Goodman's (1979, 1985) RC(M) association model for multiple nominal and/or ordinal categorical variables, can be used to study the relationship between the observed discrete variables and the unobserved continuous ones, and to study the unobserved variables. The derivation and use of log-multiplicative association models as latent variable models for discrete variables are presented in this paper. The models are based on graphical models for discrete and continuous variables where the variables follow a conditional Gaussian distribution. The models have many desirable properties, including having schematic or graphical rep-

This research was supported by grants from the National Science Foundation (#SBR96-17510) and the Bureau of Educational Research at the University of Illinois. We thank Ulf Böckenholt, Rung-Ching Tsai, and Jee-Seon Kim for useful comments and suggestions. Correspondence can be addressed to Carolyn J. Anderson, Educational Psychology, University of Illinois, 1310 South Sixth Street, Champaign, IL 61820 or Jeroen K. Vermunt, Department of Methodology, PO Box 90153, 5000 LE Tilburg, Tilburg University, The Netherlands. E-mail addresses are cja@uiuc.edu and J.K.Vermunt@kub.nl.

*University of Illinois

†Tilburg University, The Netherlands

resentations of the system of observed and unobserved variables, the log-multiplicative models can be read from the graphs, and estimates of the means, variances, and covariances of the latent variables given values on the observed variables are a function of the log-multiplicative model parameters. To illustrate some of the advantageous aspects of these models, two examples are presented. In one example, responses to items from the General Social Survey (Davis and Smith 1996) are modeled, and in the other example, panel data from two groups (Coleman 1964) are analyzed.

1. INTRODUCTION

Associations in multivariate categorical data are often due to collapsing over other variables. For example, consider the following four items from the 1994 General Social Survey (Davis and Smith 1996):

- A_1 "Do you approve or disapprove of a married woman earning money in business or industry if she has a husband capable of supporting her?" (approve, disapprove).
- A_2 "It is much better for everyone involved if the man is the achiever outside the home and the woman takes care of the home and family." (strongly agree, agree, disagree, strongly disagree).
- A_3 "A man's job is to earn money; a woman's job is to look after the home and family." (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree).
- A_4 "It is not good if the man stays at home and cares for the children and the woman goes out to work." (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree).

We would expect associations to be present between the responses to these items because all of the items appear to be indicators of a single continuous variable—namely, attitude regarding the proper roles of wives and husbands in terms of employment inside/outside the home.

As a second example, consider the Coleman (1964) panel data that consist of responses made at two time points by boys and girls to two items: their attitude toward (positive, negative) and their self-perception of membership in (yes, no) the leading or popular crowd. These two questions may be indicators of the same (continuous) variable or they may be indicators of different but correlated variables. It is also possible that there may be change over time and/or differences between boys and girls. The underlying latent variable structure has implications for what associations should be present in the observed data and the nature of these associations.

Log-linear models are very effective tools for determining what associations are present in categorical data; however, they are less useful for describing the nature of multiple observed associations. When the associations arise because we have collapsed over unobserved or not directly measurable continuous variables, the description and interpretation of the associations would be greatly facilitated if our models represented the observed associations in terms of the unobserved or latent variables. The models should also allow a researcher to study the underlying structural relationships between the unobserved variables. Ideally, researchers should be able to transform their specific theories and hypotheses about the relationships between the observed and unobserved variables into statistical models, which in turn can be readily fit to observed data. We propose a latent variable model that meets these requirements.

The latent variable models proposed here are based on graphical models for discrete and continuous variables (Lauritzen and Wermuth 1989; Wermuth and Lauritzen 1990; see also Edwards 1995; Lauritzen 1996; Whittaker 1990), and they belong to a family of "location models" for discrete and continuous variables (Olkin and Tate 1960; Afifi and Elashoff 1969; Krzanowski 1980, 1983, 1988). The models presented here differ from previously discussed cases in that the continuous variables are unobserved and we restrict our attention to cases where the discrete (observed) variables are conditionally independent given the continuous (latent) ones. The models implied for the observed data are log-multiplicative association models.

In log-multiplicative association models, which are extensions of log-linear models, dependencies between discrete variables are represented by multiplicative terms. Special cases of these models include many well-known models for categorical data such as linear-by-linear interaction models, ordinal-by-nominal association models, the uniform association model for ordinal categorical variables, the $RC(M)$ association model for two variables, and many generalizations of the $RC(M)$ association model for three or more variables (e.g., Agresti 1984; Becker 1989; Clogg 1982; Clogg and Shihadeh 1994; Goodman 1979, 1985).

A simple case of the models was discussed by Lauritzen and Wermuth (1989; Wermuth and Lauritzen 1990), who provided a latent continuous variable interpretation of Goodman's (1979) RC association model for two items. Whittaker (1989) discusses the case of multiple, uncorrelated latent variables for two and three observed variables. In this paper, we consider more general graphical models for multiple *correlated* latent variables for any number of observed variables. Additionally, we allow the covariance matrix of the latent variables to differ over values of the observed, discrete variables.

The models have many desirable properties, including having schematic or graphical representations. The graphs are useful pictorial representations of theories about phenomena, and the corresponding log-multiplicative models can be read from the graph. In many cases, estimates of the means, variances, and covariances of the latent variables are by-products of the estimation of the parameters of the log-multiplicative model.

The remainder of this paper is structured as follows. In Section 2, we present the basic ideas and approach for cases where each observed (discrete) variable is related to only one latent (continuous) variable. In Section 3, we extend the basic model to cases where the observed variables may be related to multiple latent variables. The models discussed in Sections 2 and 3 are illustrated in Section 4, using data from the 1994 General Social Survey (Davis and Smith 1996) and Coleman's (1964) panel data for the boys. In Section 5, we further generalize the latent variable model by allowing the covariance matrix of the latent variables to differ over levels of the observed discrete variables. In Section 6, we illustrate these heterogeneous covariance models by analyzing Coleman's (1964) data for the girls, and doing a combined analysis of the boys and girls data. In Section 7, we conclude with a discussion of additional possible generalizations and areas for further study. Two appendixes are included. The first describes how log-multiplicative models can be read from graphs representing the latent variable model, and the second describes the maximum-likelihood estimation of the models by the unidimensional Newton method.

2. SINGLE LATENT VARIABLE PER INDICATOR

In this section, we present models where each observed discrete variable is an indicator of only one latent variable. In Section 2.1, we derive the log-multiplicative model for the graphical model where there is one continuous latent variable, and in Section 2.2, we extend the model to cases where there are two or more correlated latent variables. In Section 2.3, we discuss identification constraints for the parameters of single indicator models and their implications.

2.1. *The One Latent Variable Model*

The example of the four items from the General Social Survey is a case where we hypothesize that each item is an indicator of a single, common latent variable; that is, we expect that a single indicator, one latent variable model should fit the data. Such a model is presented in Figure 1, where the discrete variables are represented by the squares and the continuous variable by a circle. The absence of a line connecting two variables indicates that the variables are conditionally independent given all the other variables, while the presence of a line connecting two variables, indicates that the variables are dependent.

In the latent variable models proposed in this paper, the joint distribution of the discrete and continuous variables is assumed to be conditional Gaussian (Lauritzen and Wermuth 1989; see also Edwards 1995; Lauritzen 1996; Whittaker 1990). In a conditional Gaussian distribution, the marginal distribution of the discrete variables is multinomial and the conditional distribution of the continuous variables given the discrete ones is multivariate normal where the mean and covariance matrix may differ over levels of the discrete variables. For now, we assume that the covariance matrix does not differ over levels of the discrete variables; however, this restriction is relaxed in Section 5. In other words, differences between cells of a cross-classification of observations is dealt with by allowing the means of the continuous latent variables to differ between cells. Individual differences within cells are captured by the within-cell variances of continuous variables.

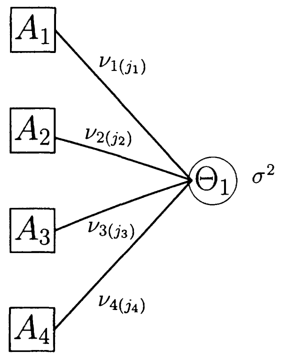

FIGURE 1. Single indicator, one latent variable model for four observed variables.

Let $\mathcal{A} = \{A_1, \ldots, A_I\}$ be a set of I discrete variables, and Θ represent the continuous variable. We denote a realization of the continuous variable by θ and an observation on the discrete variables by $\mathbf{a} = (a_{1(j_1)}, \ldots, a_{I(j_I)})$ (i.e., $\mathbf{a}$ is a cell in the cross-classification of the I discrete variables). The levels of discrete variable A_i are indexed by j_i where $j_i = 1, \ldots, J_i$. The probability that an observation falls into cell $\mathbf{a}$ is denoted by $P(\mathbf{a})$. To obtain the joint distribution of the discrete and continuous variables, we take the product of the marginal distribution of the discrete variables, which is multinomial, and the conditional distribution of the continuous variables, which is a conditional normal distribution; that is,

$$f(\mathbf{a}, \theta) = P(\mathbf{a}) f(\theta | \mathbf{a})$$
$$= P(\mathbf{a}) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{(\theta - \mu(\mathbf{a}))^2}{\sigma^2} \right] \tag{1}$$

where $f(\theta | \mathbf{a})$ is a normal distribution with mean $\mu(\mathbf{a})$, which depends on $\mathbf{a}$, and variance σ^2. Equation (1) is the moment form of a (homogeneous) conditional Gaussian distribution.

Since the continuous variable θ is unobserved, we do not have readily available estimates of $\boldsymbol{\mu}(\mathbf{a})$ and σ^2. While the observed cell proportions provide estimates of $P(\mathbf{a})$, we want a model for $P(\mathbf{a})$ that is implied by our specific hypotheses regarding the relationships between the discrete variables and the continuous variable. The model for $P(\mathbf{a})$ will contain interactions between the discrete variables that result from having collapsed over the continuous variable.

Rather than working with the moment form of the conditional Gaussian distribution, it is more useful to work with the canonical form of the distribution. The canonical form of the distribution can be obtained by re-writing equation (1) as

$$f(\mathbf{a}, \theta) = \exp\left[\log(P(\mathbf{a})) - \log(\sqrt{2\pi\sigma^2}) - \frac{1}{2}\left(\frac{\mu(\mathbf{a})^2}{\sigma^2}\right) + \frac{\mu(\mathbf{a})}{\sigma^2}\theta - \frac{1}{2}\left(\frac{\theta^2}{\sigma^2}\right)\right]. \tag{2}$$

We define

$$h(\mathbf{a}) = \frac{\mu(\mathbf{a})}{\sigma^2}, \tag{3}$$

and

$$g(\mathbf{a}) = \log(P(\mathbf{a})) - \log(\sqrt{2\pi\sigma^2}) - \tfrac{1}{2}\sigma^2 h(\mathbf{a})^2 \ , \tag{4}$$

which are both functions of **a**. Substituting definitions (3) and (4) in equation (2) we obtain

$$f(\mathbf{a},\theta) = \exp\left[g(\mathbf{a}) + h(\mathbf{a})\theta - \frac{1}{2}\left(\frac{\theta^2}{\sigma^2}\right)\right], \tag{5}$$

which is the canonical form of the conditional Gaussian distribution. In the canonical form, the joint distribution factors into three components: a discrete part $g(\mathbf{a})$, which represents the discrete variables and the dependencies among them after controlling for the continuous variables; a linear part $h(\mathbf{a})\theta$, which represents the dependencies between the discrete variables and the continuous variable; and a quadratic part $-(1/2)\theta^2/\sigma^2$, which represents the continuous variable after controlling for the discrete ones.

The model for the observed data is obtained by rewriting equation (4) in terms of $P(\mathbf{a})$,

$$P(\mathbf{a}) = \sqrt{2\pi\sigma^2}\exp\left[g(\mathbf{a}) + \frac{1}{2}\sigma^2 h(\mathbf{a})^2\right]. \tag{6}$$

Equation (6) does not include θ, which is unobserved but depends only on observed data. The hypothesis that the discrete variables are conditionally independent given the continuous variable is incorporated into the model through the parameterization we specify for $g(\mathbf{a})$, and the hypothesis that each of the discrete variables is directly related to the latent variable θ is incorporated through the parameterization we specify for $h(\mathbf{a})$.

The function $g(\mathbf{a})$ is set equal to the sum of effect terms as in log-linear models.[1] Since the discrete variables are conditionally independent given the continuous variable, $g(\mathbf{a})$ equals the sum of marginal effect terms for each of the discrete variables; that is,

$$g(\mathbf{a}) = \sum_{i=1}^{I} \lambda_{i(j_i)}, \tag{7}$$

where $\lambda_{i(j_i)}$ is the marginal or main effect term for level j_i of variable A_i. This parameterization of $g(\mathbf{a})$ is used throughout this paper, because in all

[1] If there were no continuous variables, we would have a log-linear model.

of the models considered here, the discrete variables are independent of each other given the continuous variable(s).

From equation (5), we see that $h(\mathbf{a})$ is a coefficient for the strength of the association between the discrete variables and continuous variable. Since each discrete variable is directly related to the continuous variable, we define $h(\mathbf{a})$ as

$$h(\mathbf{a}) = \sum_{i=1}^{I} \nu_{i(j_i)} \tag{8}$$

where $\nu_{i(j_i)}$ is the category score or scale value for level j_i of variable A_i. The category scale values may be estimated from the data or specified *a priori*.

Replacing $g(\mathbf{a})$ and $h(\mathbf{a})$ in equation (6) by the parameterizations given in (7) and (8) yields a log-multiplicative model for the observed data; that is,

$$\begin{aligned}\log(P(\mathbf{a})) &= \lambda + \sum_i \lambda_{i(j_i)} + \frac{1}{2}\sigma^2\left(\sum_{i=1}^{I}\nu_{i(j_i)}\right)^2 \\ &= \lambda + \sum_i \lambda^*_{i(j_i)} + \sigma^2\sum_i\sum_{k>i}\nu_{i(j_i)}\nu_{k(j_k)} \end{aligned}\tag{9}$$

where λ is a normalizing constant and $\lambda^*_{i(j_i)} = \lambda_{i(j_i)} + (1/2)\sigma^2\nu^2_{i(j_i)}$. Since the term $(1/2)\sigma^2\nu^2_{i(j_i)}$ is only indexed by j_i, it gets "absorbed" into the marginal effect term.

If there are only two discrete variables (i.e., $I = 2$), then equation (9) reduces to the $RC(1)$ association model (Goodman 1979, 1985; see also Clogg and Shihadeh 1994). For our General Social Survey example where $I = 4$ (i.e., Figure 1), we have

$$\begin{aligned}P(\mathbf{a}) = \lambda &+ \lambda^*_{1(j_1)} + \lambda^*_{2(j_2)} + \lambda^*_{3(j_3)} + \lambda^*_{4(j_4)} \\ &+ \sigma^2\nu_{1(j_1)}\nu_{2(j_2)} + \sigma^2\nu_{1(j_1)}\nu_{3(j_3)} + \sigma^2\nu_{1(j_1)}\nu_{4(j_4)} \\ &+ \sigma^2\nu_{2(j_2)}\nu_{3(j_3)} + \sigma^2\nu_{2(j_2)}\nu_{4(j_4)} + \sigma^2\nu_{3(j_3)}\nu_{4(j_4)}.\end{aligned}\tag{10}$$

In equation (10) and the more general equation (9), we have multiplicative terms with the same association parameter in each term (i.e., σ^2) and a single set of category scores for each of the variables, which appear in the different multiplicative terms.

Equations (9) and (10) are log-multiplicative association models with bivariate interactions between all pairs of the discrete variables. The best fit that can be attained using equation (9) or (10) is given by the all two-way interaction log-linear model (see Becker 1989). If an all two-way interaction log-linear model fits a data set, then we are justified in considering models such as equation (9) or (10).

2.2. *Multiple Latent Variables*

The one latent variable model is a relatively simple model. In many data sets, the observed variables may be indicators of different latent variables; therefore we generalize the model to the case of multiple latent variables. The derivation given in Section 2.1 is extended to obtain a more general model for the observed data. The one latent variable model is a special case of this more general model. In this section, we will examine two additional special cases, including the most complex single indicator model.

Let $\boldsymbol{\Theta} = \{\Theta_1, \ldots, \Theta_M\}$ be a set of M continuous variables where $M \leq I$ and the $(M \times 1)$ vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)'$ be a realization of the M latent variables. The moment form of the joint distribution of the I discrete and M latent variables is obtained by multiplying the marginal distribution of the discrete variables, which is multinomial, and the conditional distribution of the continuous variables, which is multivariate normal; that is,

$$\begin{aligned} f(\mathbf{a}, \boldsymbol{\theta}) &= P(\mathbf{a}) f(\boldsymbol{\theta}|\mathbf{a}) \\ &= P(\mathbf{a})(2\pi)^{-M/2}|\boldsymbol{\Sigma}|^{-1/2} \\ &\quad \times \exp[-\tfrac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}(\mathbf{a}))'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta} - \boldsymbol{\mu}(\mathbf{a}))] \end{aligned} \tag{11}$$

where $f(\boldsymbol{\theta}|\mathbf{a})$ is a multivariate normal distribution with the $(M \times 1)$ mean vector $\boldsymbol{\mu}(\mathbf{a})$, which is a function of $\mathbf{a}$, and the $(M \times M)$ covariance matrix $\boldsymbol{\Sigma}$. The canonical form is obtained from equation (11) by multiplying the terms in the exponent and redefining parameters:

$$\begin{aligned} f(\mathbf{a}, \boldsymbol{\theta}) &= P(\mathbf{a})(2\pi)^{-M/2}|\boldsymbol{\Sigma}|^{-1/2} \\ &\quad \times \exp[-\tfrac{1}{2}\boldsymbol{\mu}(\mathbf{a})'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}(\mathbf{a}) + \boldsymbol{\mu}(\mathbf{a})'\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta} - \tfrac{1}{2}\boldsymbol{\theta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}] \\ &= \exp[g(\mathbf{a}) + \mathbf{h}(\mathbf{a})'\boldsymbol{\theta} - \tfrac{1}{2}\boldsymbol{\theta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}] \end{aligned} \tag{12}$$

where $\mathbf{h}(\mathbf{a})$ is the $(M \times 1)$ vector valued function

$$\mathbf{h}(\mathbf{a}) = \mathbf{\Sigma}^{-1}\boldsymbol{\mu}(\mathbf{a}), \tag{13}$$

and

$$g(\mathbf{a}) = \log(P(\mathbf{a})) - \frac{M}{2}\log(2\pi) - \frac{1}{2}\log(|\mathbf{\Sigma}|) - \frac{1}{2}\mathbf{h}(\mathbf{a})'\mathbf{\Sigma}\mathbf{h}(\mathbf{a}). \tag{14}$$

The model for the observed data is found by rewriting equation (14) in terms of $P(\mathbf{a})$,

$$P(\mathbf{a}) = (2\pi)^{M/2}|\mathbf{\Sigma}|^{1/2}\exp[g(\mathbf{a}) + \tfrac{1}{2}\mathbf{h}(\mathbf{a})'\mathbf{\Sigma}\mathbf{h}(\mathbf{a})]. \tag{15}$$

In the single indicator multiple latent variable model, to obtain a specific model for the observed data, we must parameterize $g(\mathbf{a})$, $\mathbf{h}(\mathbf{a})$, and $\mathbf{\Sigma}$.

An example of a multiple latent variable model that may fit the Coleman panel data is shown in Figure 2. In this model, two items are directly related to one latent variable, the other two items are directly related to a second latent variable, and the two latent variables are correlated. Since the discrete variables (the items) are conditionally independent given the two latent variables, $g(\mathbf{a})$ has the same definition as in the one latent variable model—i.e., equation (4).

The items (discrete variables) have been partitioned into two mutually exclusive sets $\mathcal{A}_1 = \{A_1, \ldots, A_r\}$ and $\mathcal{A}_2 = \{A_{r+1}, \ldots, A_I\}$. For the

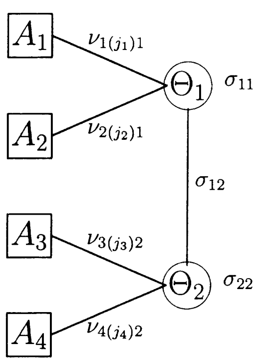

FIGURE 2. Single indicator, two correlated latent variable model for four observed variables.

Coleman data, $r = 2$ and $I = 4$. Since the variables in $\mathcal{A}_1$ are directly related to Θ_1, the first element of the (2×1) vector $\mathbf{h}(\mathbf{a})$ contains coefficients that relate the variables in $\mathcal{A}_1$ and Θ_1. Likewise, the second element of the $\mathbf{h}(\mathbf{a})$ contains coefficients that relate the variables in $\mathcal{A}_2$ and Θ_2. Thus we parameterize $\mathbf{h}(\mathbf{a})$ as

$$\mathbf{h}(\mathbf{a}) = \begin{pmatrix} \sum_{i=1}^{r} \nu_{i(j_i)1} \\ \sum_{i=r+1}^{I} \nu_{i(j_i)2} \end{pmatrix}, \tag{16}$$

where $\nu_{i(j_i)m}$ is the category score or scale value for level j_i of discrete variable A_i for latent variable Θ_m.

Any hypotheses that we have about the relationship between the latent variables are incorporated into the model by the parameterization we specify for $\mathbf{\Sigma}$. To complete our model, we define $\mathbf{\Sigma}$ as

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}. \tag{17}$$

Replacing $g(\mathbf{a})$, $\mathbf{h}(\mathbf{a})$, and $\mathbf{\Sigma}$ in equation (15) by their definitions in equations (4), (16), and (17) gives us the model for the observed data,

$$\begin{aligned} \log(P(\mathbf{a})) = \lambda &+ \sum_{i=1}^{I} \lambda^*_{i(j_i)} + \sigma_{11}\left(\sum_{i=1}^{r-1} \sum_{k=i+1}^{r} \nu_{i(j_i)1}\nu_{k(j_k)1} \right) \\ &+ \sigma_{22}\left(\sum_{i=r+1}^{I-1} \sum_{k=i+1}^{I} \nu_{i(j_i)2}\nu_{k(j_k)2} \right) \\ &+ \sigma_{12}\left(\sum_{i=1}^{r} \sum_{k=r+1}^{I} \nu_{i(j_i)1}\nu_{k(j_k)2} \right), \end{aligned} \tag{18}$$

where

$$\lambda^*_{i(j_i)} = \begin{cases} \lambda_{i(j_i)} + (1/2)\sigma_{11}\nu^2_{i(j_i)1} & \text{if } A_i \in \mathcal{A}_1 \\ \lambda_{i(j_i)} + (1/2)\sigma_{22}\nu^2_{i(j_i)2} & \text{if } A_i \in \mathcal{A}_2. \end{cases}$$

For the Coleman data where $I = 4$ and $r = 2$ (i.e., Figure 2), the specific log-multiplicative model based on equation (18) is

$$\begin{aligned} P(\mathbf{a}) = {} & \lambda + \lambda^*_{1(j_1)} + \lambda^*_{2(j_2)} + \lambda^*_{3(j_3)} + \lambda^*_{4(j_4)} \\ & + \sigma_{11}\nu_{1(j_1)1}\nu_{2(j_2)1} + \sigma_{22}\nu_{3(j_3)2}\nu_{4(j_4)2} \\ & + \sigma_{12}\nu_{1(j_1)1}\nu_{3(j_3)2} + \sigma_{12}\nu_{1(j_1)1}\nu_{4(j_4)2} \\ & + \sigma_{12}\nu_{2(j_2)1}\nu_{3(j_3)2} + \sigma_{12}\nu_{2(j_2)1}\nu_{4(j_4)2}. \end{aligned} \tag{19}$$

From equation (19), we can see more clearly that the model contains multiplicative terms for all bivariate associations and there is a single set of scale values for each variable. Unlike the one latent variable model, equation (9), where there is a single association parameter for each of the multiplicative terms, in equations (18) and (19) there are three different association parameters for the multiplicative terms: σ_{11}, σ_{22}, and σ_{12}. When the discrete variables within a set are related because they are all indicators of the same latent variable, the association parameter is the variance. When discrete variables from the two different sets are related because the corresponding latent variables are related, the association parameter is the covariance between the latent variables.

In the most complex, single factor per indicator model, each discrete variable is an indicator of (i.e., directly related to) a different latent variable and all the latent variables are correlated. For $I = 4$, the graph of this model is given in Figure 3. With this structure, $M = I$ and

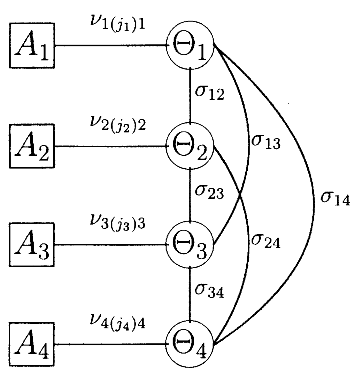

FIGURE 3. The most complex single indicator model for four variables.

$$\mathbf{h}(\mathbf{a}) = \begin{pmatrix} \nu_{1(j_i)1} \\ \vdots \\ \nu_{I(j_I)I} \end{pmatrix}.$$

The definition of $g(\mathbf{a})$ remains the same—i.e., equation (4)—and $\boldsymbol{\Sigma}$ is now an $(I \times I)$ matrix of variances and covariances between the latent variables. Using these definitions, the model for the observed data is

$$\log(P(\mathbf{a})) = \lambda + \sum_i \lambda^*_{i(j_i)} + \sum_i \sum_{k>i} \sigma_{ik} \nu_{i(j_i)i} \nu_{k(j_k)k} \tag{20}$$

where $\lambda^*_{i(j_i)} = \lambda_{i(j_i)} + (1/2)\sigma_{ii} \nu^2_{i(j_i)i}$.

Equation (20) is a multivariate generalization of the $RC(1)$ association model, which for three variables is equivalent to models discussed by Clogg (1982; see also Agresti 1984). If category scores are known, then equation (20) is a log-linear model with linear-by-linear interaction terms for each pair of the observed variables (i.e., $\sigma_{ik} x_{i(j_i)i} x_{k(j_k)k}$ where the x's are known scores). If scores for some variables are known but not for others, then equation (20) includes some ordinal-by-nominal interaction terms (e.g., $\sigma_{ik} \nu_{i(j_i)i} x_{k(j_k)k}$).

If no partial association between a pair of discrete variables exists, then the corresponding covariance can be set to zero, which sets the interaction term for the pair of variables equal to zero. But if partial associations between all pairs of variables are present, then it is possible to obtain simpler models by imposing certain restrictions on the model parameters. The models that have fewer latent variables—for example, equations (9) and (18)—can be thought of as special cases of the most complex single indicator model, equation (20), where equality restrictions have been imposed on the association parameters (i.e., covariances) across the multiplicative terms.

All of the single indicator per latent variable models include bivariate interactions between pairs of discrete variables; therefore, a log-linear model that provides a baseline fit (best fit) for the latent variable, log-multiplicative models always exists. Log-linear models are useful in that they indicate whether particular latent variable models may be appropriate. The use of log-linear models in conjunction with the graphical/latent variable, log-multiplicative models is illustrated in the analyses presented in Sections 4 and 6.

When each discrete variable is an indicator of a different latent variable (i.e., equation 20), we can only estimate the covariances between the latent variables. We cannot estimate the variances because they are absorbed into the marginal effect terms (i.e., $\lambda^*_{i(j_i)} = \lambda_{i(j_i)} + (1/2)\sigma_{ii}\nu^2_{i(j_i)i}$). There is no way to tease apart the term $\lambda_{i(j_i)}$ from the terms involving the variances. However, in simpler models such as (9) and (18), we can estimate the variances and the covariances. Given estimates of variances and covariances and using the fact that $\boldsymbol{\mu}(\mathbf{a}) = \boldsymbol{\Sigma}\mathbf{h}(\mathbf{a})$ (see equation 13), we can estimate the conditional means of the latent variables, $\boldsymbol{\mu}(\mathbf{a})$.

2.3. *Identification Constraints*

Identification constraints are required to estimate the parameters of the log-multiplicative models. The choice of constraints sets the scale of the conditional means of the latent variables. Adding conditions beyond those needed for identification correspond to more restrictive latent variable models.

For all log-multiplicative models, location constraints are required for the marginal effect terms, $\lambda^*_{i(j_i)}$, and for the scale values, $\nu_{i(j_i)m}$. These may be setting one value equal to zero (e.g., $\nu_{i(1)m} = 0$), or setting the sum equal to zero (e.g., $\sum_{j_i} \nu_{i(j_i)m} = 0$). We use zero sum constraints in the examples presented in Sections 4 and 6.

One additional constraint is required for each latent variable. While the variance of each latent variable could be set to a constant (e.g., $\sigma_{mm} = 1$ for all m), for reasons that become clear below, it is advantageous to set the scale of the category scores for one discrete variable that is directly related to the latent variable. For example, if A_1 and Θ_m are directly related, then $\sum_{j_1} \nu^2_{1(j_1)m} = 1$. The rule adopted here (assuming $I > 2$) is that a scaling condition is imposed on the scale values of one observed variable per latent variable. For the one common latent variable model, equation (9), the scale values of one variable need to be scaled, and for the two correlated latent variable model in equation (18), the scale values of one variable in $\mathcal{A}_1$ and one variable in $\mathcal{A}_2$ need to be scaled. For model (20), we take this rule to the limit and impose scaling constraints on the scale values for each of the discrete variables.

The category scale values provide two types of information about how the mean of a latent variable differs over levels of an observed variable. This can be seen by expressing the scale values as $\nu_{i(j_i)m} = \omega_{im}\nu^*_{i(j_i)m}$ where $\omega_{im} = (\sum_{j_i} \nu^2_{i(j_i)m})^{1/2}$ and $\sum_{j_i} \nu^{*2}_{i(j_i)m} = 1$. The ω_{im}'s can be interpreted as measures of the overall (relative) strength of the relationship between variable A_i and latent variable Θ_m, and the $\nu^*_{i(j_i)m}$'s represent category specific information about this relationship. For identification purposes, if we impose the scaling condition on the scale values of, for example, A_1 where A_1 is an indicator of Θ_1 (i.e., $\sum_{j_1} \nu^2_{1(j_1)1} = 1$), then for $i \neq 1$, the ω_{i1}'s are free to vary and the variance of Θ_1 is an estimated parameter. Imposing a scaling condition on the scale values of more than one variable per latent variable is a restriction. This restriction can be interpreted as placing equality restrictions on the overall strength of the relationship between the observed variables and the latent variables (i.e., the ω_{im}'s).

We can now show that the case of $I = 3$ is special. The single indicator latent variable model for three observed variables implies the following log-multiplicative model

$$\log(P(\mathbf{a})) = \lambda + \lambda^*_{1(j_1)} + \lambda^*_{2(j_2)} + \lambda^*_{3(j_3)} + \sum_i \sum_{k>i} \sigma^*_{11} \nu_{i(j_i)1} \nu_{k(j_k)1}. \tag{21}$$

Suppose that for identification, the condition $\sum_{j_1} \nu^2_{1(j_1)1} = 1$ is imposed. Since we can represent the scale values for the other two variables as $\nu_{2(j_2)1} = \omega_2 \nu^*_{2(j_2)1}$ and $\nu_{3(j_3)1} = \omega_3 \nu^*_{3(j_3)1}$, model (21) is empirically indistinguishable from model (20), which is seen by setting $\sigma_{12} = \omega_2 \sigma^*_{11}$, $\sigma_{13} = \omega_3 \sigma^*_{11}$, and $\sigma_{23} = \omega_2 \omega_3 \sigma^*_{11}$. This equivalence provides an alternative interpretation for the partial association model for three variables discussed by Clogg (1982; see also Agresti 1984).

3. MULTIPLE LATENT VARIABLES PER INDICATOR

Observed variables may be directly related to more than one latent variable. Adding this complexity to the models does not require the derivation of a more complex model. We use the same general model derived in Section 2.2 (i.e., equation 15), but specify a more complex parameterization for $\mathbf{h}(\mathbf{a})$. Unlike the single indicator models where there is a single set of scale values for each discrete variable, in the multiple indicator models, a discrete variable may have multiple sets of scale values.

The major difficulty in using log-multiplicative models as multiple indicator models is determining the necessary and sufficient constraints needed to uniquely identify the parameters of the log-multiplicative models. For all models, the identification constraints described in Section 2.3 (i.e., location constraints on the marginal effect terms and the scale values and a scaling constraint on the category scores of one observed variable per latent variable) are needed. The additional identification constraints (if any) depend on the complexity of the model.

Since the number of possible multiple indicator models is far too large to consider here, we derive the log-multiplicative models for three of the four models that are used in the examples presented in Sections 4 and 6,[2] and show how to determine the identification constraints for these models. In the first two examples, the latent variables are uncorrelated, and in the third example, the latent variables are correlated.

3.1. *Uncorrelated Latent Variables*

Consider the General Social Survey data where all the items appear to be indicators of one latent variable (i.e., attitude). If the single indicator, one latent variable does not fit, then one possibility is that there is extra pair-specific association that is not accounted for by the common latent variable. To model pair-specific association, we can introduce additional latent variables for pairs of discrete variables. For example, suppose that in addition to being indicators of the latent attitude variable Θ_1, items A_1 and A_2 are directly related to Θ_2 (a pair specific variable), which is uncorrelated with Θ_1. In this case, $\boldsymbol{\Sigma}$ equals a (2×2) diagonal matrix, and $\mathbf{h}(\mathbf{a})$ is parameterized as

$$\mathbf{h}(\mathbf{a}) = \begin{pmatrix} \sum_{i=1}^{4} \nu_{i(j_i)1} \\ \nu_{1(j_1)2} + \nu_{2(j_2)2} \end{pmatrix}.$$

Using this parameterization of $\mathbf{h}(\mathbf{a})$, the parameterization $g(\mathbf{a})$ in equation (7) and a diagonal $\boldsymbol{\Sigma}$ in our general model, equation (15), gives us the log-multiplicative model

$$\log(P(\mathbf{a})) = \lambda + \sum_i \lambda^*_{i(j_i)} + \sigma_{11} \sum_i \sum_{k>i} \nu_{i(j_i)1} \nu_{k(j_k)1} + \sigma_{22} \nu_{1(j_1)2} \nu_{2(j_2)2}. \tag{22}$$

In addition to the identification constraints needed for the common part of the model, the scale values for each discrete variable related to the pair-specific latent variable must have a scaling condition im-

[2]The fourth model, which has a heterogeneous covariance matrix, is discussed in Section 6.2.

posed on them (i.e., $\sum_{j_1} \nu^2_{1(j_1)2} = \sum_{j_2} \nu^2_{2(j_2)2} = 1$). To see this, replace $\nu_{1(j_1)2}$ with $\nu^*_{1(j_1)2} = c\nu_{1(j_1)2}$ where c is a constant. The value of the term $\sigma_{22}\nu_{1(j_1)2}\nu_{2(j_2)2}$ in equation (22) remains the same; that is,

$$\sigma_{22}\nu_{1(j_1)2}\nu_{2(j_2)2} = \sigma^*_{22}\nu^*_{1(j_1)2}\nu^*_{2(j_2)2}$$

where $\sigma^*_{22} = \sigma_{22}/c^2$ and $\nu^*_{2(j_2)2} = c\nu_{2(j_2)2}$.

Extra association may also be due to multiple uncorrelated latent variables to which each discrete variable is directly related. This would give us

$$\mathbf{h}(\mathbf{a}) = \begin{pmatrix} \sum_i \nu_{i(j_i)1} \\ \vdots \\ \sum_i \nu_{i(j_i)M} \end{pmatrix}. \tag{23}$$

Since the latent variables are uncorrelated, $\mathbf{\Sigma}$ equals a diagonal matrix and the log-multiplicative model is

$$\log(P(\mathbf{a})) = \lambda + \sum_i \lambda^*_{i(j_i)} + \sum_i \sum_{k>i} \sum_m \sigma_{mm}\nu_{i(j_i)m}\nu_{k(j_k)m}. \tag{24}$$

To identify the parameters in equation (24), we need to use only the identification constraints given in Section 2.3 (assuming $I > 2$).

3.2. *Correlated Latent Variables*

Often in the social sciences, latent variables are correlated; therefore, we consider the situation where each of I observed variables is directly related to each of M latent variables ($M > 1$), and the latent variables are correlated. In this case, the parameterization of $\mathbf{h}(\mathbf{a})$ is given in equation (23). Assuming that all of the latent variables are correlated and using equation (15) gives us the log-multiplicative model

$$\log(P(\mathbf{a})) = \lambda + \sum_i \lambda^*_{i(j_i)} + \sum_i \sum_{k>i} \sum_m \sigma_{mm}\nu_{i(j_i)m}\nu_{k(j_k)m} + \sum_i \sum_{k>i} \sum_m \sum_{m'>m} \sigma_{mm'}\nu_{i(j_i)m}\nu_{k(j_k)m'}, \tag{25}$$

where λ is a normalizing constant and $\lambda^*_{i(j_i)} = \lambda_{i(j_i)} + (1/2)\sum_m \times \sum_{m'} \sigma_{mm'} \nu_{i(j_i)m} \nu_{i(j_i)m'}$. Since the sum $(1/2)\sum_m \sum_{m'} \sigma_{mm'} \nu_{i(j_i)m} \nu_{i(j_i)m'}$ is only indexed by j_i, it gets "absorbed" into the marginal effect term. Equation (25) is the most complex multiple indicator model possible, and as shown below, it has more parameters than can be estimated from data.

To show what constraints are needed for equation (25), as well as other multiple indicator models, let $\mathbf{N}_i$ equal the $(J_i \times M)$ matrix whose columns contain the scale values for the categories of variable A_i; that is, $\mathbf{N}_i = (\boldsymbol{\nu}_{i1}, \ldots, \boldsymbol{\nu}_{iM})$ where $\boldsymbol{\nu}_{im}$ equals the $(J_i \times 1)$ vector of scale values $\nu_{i(j_i)m}$. If A_i is conditionally independent of latent variable Θ_m, then all the scale values relating A_i and latent variable Θ_m equal zero and the corresponding column of $\mathbf{N}_i$ contains zeros (i.e., $\boldsymbol{\nu}_{im} = \mathbf{0}$). The interaction term for levels j_i and j_k of variables A_i and A_k equals the (j_i, j_k) element of the matrix product $\mathbf{N}_i \boldsymbol{\Sigma} \mathbf{N}_k'$ where $\boldsymbol{\Sigma}$ is the covariance matrix of the latent variables. For each cell in the cross-classification of the discrete variables, the interaction terms in the model equal the appropriate elements from the matrices in the set

$$\{\mathbf{N}_i \boldsymbol{\Sigma} \mathbf{N}_k' \,|\, i < k\}. \tag{26}$$

Determining the additional constraints needed to identify a model consists of determining whether transformations of the $\mathbf{N}_i$'s and $\boldsymbol{\Sigma}$ exist that have no effect on the value of the elements of the matrix products in (26).

For model (25), none of the columns of the $\mathbf{N}_i$'s equals $\mathbf{0}$. Given any $(M \times M)$ nonsingular matrix T, we can always set $\mathbf{N}_i^* = \mathbf{N}_i \mathbf{T}$ for all i and $\boldsymbol{\Sigma}^* = \mathbf{T}^{-1} \boldsymbol{\Sigma} \mathbf{T}^{-1}$ without changing the values of any of the elements of the matrix products in (26). Given this indeterminacy (and for convenience), we can arbitrarily set all covariances equal to zero and estimate the M variances. This leads us back to the uncorrelated latent variable model in equation (24).

If at least one observed variable is not an indicator of a latent variable, then restrictions exist on the set of possible parameters. For example, consider the case of four variables and two latent variables where A_1 and A_4 are indicators of Θ_1 and Θ_2, respectively, and A_2 and A_3 are indicators of both Θ_1 and Θ_2. The graph for this model is given in Figure 4 and $\mathbf{h}(\mathbf{a})$ is set equal to

$$\mathbf{h}(\mathbf{a}) = \begin{pmatrix} \nu_{1(j_1)1} + \nu_{2(j_2)1} + \nu_{3(j_3)1} \\ \nu_{2(j_2)2} + \nu_{3(j_3)2} + \nu_{4(j_4)2} \end{pmatrix}.$$

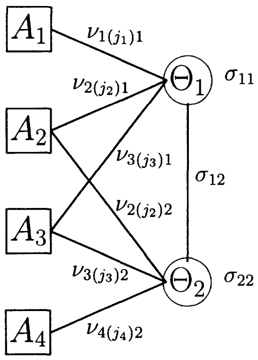

FIGURE 4. Multiple indicator, two correlated latent variable model for four observed variables.

The log-multiplicative model for this case is given in equation (35) in Appendix A. For the log-multiplicative model, the matrices of scale values equal $\mathbf{N}_1 = (\boldsymbol{\nu}_{11}, \mathbf{0})$, $\mathbf{N}_2 = (\boldsymbol{\nu}_{21}, \boldsymbol{\nu}_{22})$, $\mathbf{N}_3 = (\boldsymbol{\nu}_{31}, \boldsymbol{\nu}_{32})$ and $\mathbf{N}_4 = (\mathbf{0}, \boldsymbol{\nu}_{42})$. The covariance σ_{12} cannot be arbitrarily set equal to zero, because $\mathbf{N}_1 \boldsymbol{\Sigma} \mathbf{N}_4' = (\sigma_{12} \boldsymbol{\nu}_{11} \boldsymbol{\nu}_{42}')$. Setting $\sigma_{12} = 0$ implies that there is no (partial) association between A_1 and A_4. After imposing location and scaling conditions, we only need one additional constraint: one variance needs to be set equal to a constant.

When the latent variables are correlated, it can be especially difficult to determine the identification constraints; therefore, we suggest empirically verifying them. Given a set of conditions on the parameters that are believed to be needed for identification, fit the model with fewer conditions. We suggest starting by only imposing those conditions given in Section 2.3, which are known to be required for identification. Successively add conditions on the parameters. If a condition is only needed for identification, then the fit statistics for the model will be exactly the same as the model fit without all of the identification conditions. If a condition on the parameters is a restriction, then the model will not fit the data as well.[3] Once it has been determined that none of the conditions is a restriction, the model with these conditions imposed should be refit several times

[3]In Appendix B, where maximum-likelihood estimation of log-multiplicative models is presented, we discuss a second method for checking whether a condition imposed on the scale values is a restriction.

with random starting values. If the conditions are sufficient for identification, then the parameter estimates will be exactly the same. However, if the conditions imposed are not sufficient, different parameter estimates will likely be obtained, in which case, additional conditions are needed on the parameters to identify a unique solution.

4. EXAMPLES OF SINGLE AND MULTIPLE INDICATOR MODELS

Two example analyses are given here that illustrate the graphical/latent variable models presented in Sections 2 and 3. The models are used here in both an exploratory and a confirmatory fashion.

The log-multiplicative models fit to data in this section and in Section 6 were fit using the unidimensional Newton method described in Appendix A, which was implemented in an experimental version of ℓ_{EM} (Vermunt 1997).

4.1. *General Social Survey Data*

For this example, we analyze the $(2 \times 4 \times 5 \times 5)$ cross-classification of 899 responses from the 1994 General Social Survey (Davis and Smith 1996) to the four items listed in Section 1. Statistics for the models fit to the data are reported in Table 1. Since the data contain many zeros, to assess model goodness-of-fit, we report dissimilarity indices (D) in addition to likelihood ratio statistics (G^2). For model comparisons (most of which are not nested), we use the BIC statistic to take into account goodness-of-fit, sample size, and model complexity.

As baseline models, the independence and all two-way interaction log-linear models were fit to the data. The all two-way model fits the data ($G^2 = 117.93$, $df = 136$, $p = .87$); however, it is complex and estimating the parameters is problematic due to zeros in the observed bivariate margins. While the items appear to measure the same attitude, Model (c), the single indicator one latent variable model (i.e., equation 10), is unsatisfactory. The two uncorrelated latent variable model (Model d) where each item is an indicator of both latent variables (i.e., equation 24), fits the data; however, this model is complex and difficult to interpret.

Given that all the items appear to be indicators of the same attitude, we considered models with one common latent variable and additional

TABLE 1

Statistics for models fit to four items from the 1994 General Social Survey*

	df	G^2	*p*	*D*	BIC
Model					
(a) Independence	187	1063.25	<.01	.378	−209
(b) All 2-way interaction log-linear	136	117.93	.87	.089	−807
(c) One latent variable	175	279.50	<.01	.187	−911
(d) Two uncorrelated latent variables	163	170.61	.33	.116	−938
One common latent variable with extra latent variables for					
(e) $A_1A_2, A_1A_3, A_1A_4, A_2A_3, A_2A_4$ & A_3A_4	145	143.60	.52	.100	−843
(f) $A_1A_2, A_1A_3, A_2A_3, A_2A_4$ & A_3A_4	149	144.30	.59	.102	−869
(g) A_1A_2, A_1A_3, A_2A_3 & A_3A_4	155	161.36	.35	.113	−893
(h) A_1A_3, A_2A_3 & A_3A_4	158	164.98	.34	.124	−910
(i) A_2A_3 & A_3A_4	162	168.76	.34	.127	−933
(j) A_3A_4	168	194.41	.08	.138	−948
(k) A_2A_3	169	240.94	<.01	.168	−908
(l) A_2A_4	169	260.16	<.01	.177	−899
(m) A_1A_2	172	271.65	<.01	.183	−898
(n) A_1A_3	171	275.56	<.01	.189	−887
(o) A_1A_4	171	275.61	<.01	.189	−887

*Respondents were asked whether wives and/or husbands should work outside of the home.

uncorrelated latent variables to represent associations between pairs of items not captured by the common variable. Model (e), which has six extra latent variables, fits the data; therefore, we sought simpler models by successively deleting pair-specific latent variables, Models (f)–(j). We also fit the one common latent variable model plus one uncorrelated variable for a pair of items, Models (j)–(o). Since Model (j) has the smallest BIC statistic, fits the data reasonably well,[4] and its interpretation is similar to Models (c) and (i), we report the results from Model (j).

Model (j) has one common latent variable and a second uncorrelated variable that accounts for extra A_3A_4 association. Table 2 contains the estimated association parameters and their standard errors, as well as $\hat{\omega}_{i1}$ computed for each item. The common latent variable is an attitude variable pertaining to the proper roles of wives and husbands in terms of employment inside/outside the home. From the $\hat{\omega}_{i1}$'s, items A_3 and A_2 are most strongly related to the common latent variable, followed by items A_4 and A_1. The conditional mean of the common latent variable is proportional to the sum of the scale values corresponding to a given response pattern (see equation 13). The order of category scores for the common latent variable corresponds to the order of the response options, except for item A_4 where the scale values for "strongly agree" and "agree" are nearly equal but out of order (i.e., $\hat{\nu}_{A_4(1)1} = -.135$ and $\hat{\nu}_{A_4(2)1} = -.145$). The greater the agreement with a statement, the greater the value on the conditional mean of the latent variable.

Relative to the common latent variable, the variable for the extra A_3A_4 association accounts for inconsistent extreme responses "strongly agree" to item A_3 but "strongly disagree" to item A_4, and overly consistent responses for the more moderate responses. These inconsistencies and consistencies may be due in part to the location of the items on the survey and to the wording of item A_4. Item A_4 immediately follows A_3, while A_1 and A_2 are from two different sections of the survey. Item A_4 differs from the other items in that the traditional roles of husbands and wives are reversed and children are explicitly mentioned.

4.2. *Coleman Panel Data: The Boys*

The Coleman (1964) panel data, which are reported in Table 3, consist of responses made at two time points by 3398 boys and 3260 girls to two

[4]There are two large standardized residuals; however, these were cells where the observed count equals 1 and the fitted values are between .01 and .02.

TABLE 2

Estimated Parameters (and Standard Errors) from Model (j) in Table 1, Fit to Four Items from the 1994 General Social Survey*

			Response Options				
			Strongly Agree	Agree (approve)†	(neither agree nor disagree)	Disagree (disapprove)†	Strongly Disagree
$\hat{\sigma}_{11} =$ 11.294	(4.252)						
$\hat{\omega}_{11} =$ .077‡		$\hat{\nu}_{1(j_1)1} =$		.055 (.016)		−.055 (.016)	
$\hat{\omega}_{21} =$ .587		$\hat{\nu}_{2(j_2)1} =$	−.306 (.094)	−.211 (.065)		.066 (.030)	.450 (.138)
$\hat{\omega}_{31} =$ 1.00		$\hat{\nu}_{3(j_3)1} =$	−.679 (.063)	−.305 (.074)	.083 (.050)	.324 (.024)	.577 (.051)
$\hat{\omega}_{41} =$ .287		$\hat{\nu}_{4(j_4)1} =$	−.135 (.048)	−.145 (.034)	.020 (.016)	.064 (.032)	.196 (.049)
$\hat{\sigma}_{22} =$ 2.642	(1.094)						
		$\hat{\nu}_{3(j_3)2} =$	.783 (.137)	−.333 (.186)	−.052 (.163)	−.511 (.112)	.113 (.233)
		$\hat{\nu}_{4(j_4)2} =$	.108 (.164)	−.471 (.104)	.097 (.138)	−.467 (.119)	.734 (.122)

*See text for the items.

†The response options for item A_1 were "approve" and "disapprove."

‡$\hat{\omega}_{i1} = (\sum_{j_i} \hat{\nu}^2_{i(j_i)1})^{1/2}$.

TABLE 3
Panel Data Where A_t and B_t Refer to the Attitude and Membership Items at Time Point t*

				Boys			Girls		
B_1[a]	A_1[b]	B_2	A_2	Count	Fitted	Std resid	Count	Fitted	Std resid
2	2	2	2	458	454.83	.15	484	470.58	.62
2	2	2	1	140	151.39	−.93	93	102.49	−.94
2	2	1	2	110	121.46	−1.04	107	103.71	.32
2	2	1	1	49	51.68	−.37	32	29.74	.41
2	1	2	2	171	167.63	.26	112	113.49	−.14
2	1	2	1	182	177.15	.36	110	112.44	−.23
2	1	1	2	56	57.22	−.16	30	32.93	−.51
2	1	1	1	87	77.30	1.10	46	42.96	.46
1	2	2	2	184	171.87	.93	129	146.76	−1.47
1	2	2	1	75	73.13	.22	40	42.09	−.32
1	2	1	2	531	534.85	−.17	768	766.94	.04
1	2	1	1	281	290.89	−.58	321	289.60	1.85
1	1	2	2	85	80.97	.45	74	74.00	.00
1	1	2	1	97	109.38	−1.18	75	60.80	1.82
1	1	1	2	338	322.09	.89	303	320.66	−.99
1	1	1	1	554	556.17	−.09	536	550.80	−.63

Source: Coleman (1964).

*Fitted values and standardized residuals are from Model (f) in Table 6 (i.e., graph in Figure 5 with heterogeneous $\boldsymbol{\Sigma}$ and $\tau\delta_{1122,G(j)}$).

[a]For items B_1 and B_2, $j = 1$ for "no" and $j = 2$ for "yes."

[b]For items A_1 and A_2, $j = 1$ for "negative" and $j = 2$ for "positive."

items: their attitude toward (positive, negative) and their self-perception of membership in (yes, no) the leading or popular crowd. The data for the boys have been analyzed extensively (e.g., Agresti 1997; Andersen 1988; Goodman 1978; Langeheine 1988; Whittaker 1990), while the data for the girls has not. We analyze the boys in this section, and in Section 6, we model the girls data.

The fit statistics for models estimated for the boys data are reported in the left side of Table 4. We find that the independence log-linear model fails to fit ($G^2 = 1421.68$, $df = 11$, $p < .001$), but the all two-way interaction log-linear model provides a good fit for the boys ($G^2 = 1.21$, $df = 5$, $p = .94$). Given that the all two-way model fits well, we consider log-multiplicative models. The simplest model with one common latent variable (i.e., equation 10) fails to fit ($G^2 = 243.59$, $df = 7$,

TABLE 4

Fit Statistics for Models Estimated Separately to the Boys and Girls Data

	Boys Data			Girls Data					
							with $\tau\delta_{1122}$		
Model	*df*	G^2	*p*	*df*	G^2	*p*	*df*	G^2	*p*
Baseline Models									
(a) Independence	11	1421.68	<.01	11	1845.03	<.01	10	1725.65	<.01
(b) All 2–way loglinear	5	1.21	.94	5	8.39	.14	4	4.44	.34
Latent Variable Models									
(c) 1 latent variable	7	243.59	<.01	7	314.32	<.01	6	307.59	<.01
(d) 2 correlated variables, multiple indicators	5	1.21	.94	5	8.70	.12	4	4.44	.35
(e) 2 correlated variables, single indicator	6	1.21	.98	6	17.13	.01	5	5.22	.39
(f) Model (e) with scaling restrictions	8	5.43	.71	8	23.29	<.01	7	9.73	.20
(g) Model (f) with $\sigma_{12} = 0$	9	97.52	<.01	9	128.66	<.01	8	115.72	<.01

$p < .001$). We next estimate a multiple indicator, two correlated latent variable model with the following characteristics: attitude at time one, A_1, is related to one latent variable; membership at time two, B_2, is related to a second latent variable; and the remaining two variables, A_2 and B_1, are allowed to be related to both latent variables (i.e., equation A.1 in Appendix A, and Figure 4 where A_3 and A_4 correspond to B_1 and B_2, respectively). For identification, the category scores for A_1 and B_2 are scaled and $\sigma_{22} = 1$. This model, Model (d) in Table 4, has the same fit and degrees of freedom as the all two-way interaction log-linear model; however, the log-multiplicative model provides us with information regarding the structure underlying the data. The estimated scale values for the boys data from Model (d) are given in Table 5.

The scale values in Table 5 suggest that the two attitude items are indicators of the same latent variable, "attitude," and the two membership items are indicators of a second correlated latent variable, "membership perception," (i.e., Figure 2, where A_3 and A_4 correspond to B_1 and B_2). Model (e), the corresponding single indicator, two correlated latent variable model (i.e., equation 19), fits the data nearly as well as the multiple indicator, two correlated latent variable model ($G^2 = 1.21, df = 6, p = .98$). Also suggested by the estimates in Table 5 is that the strength of the relationship between the observed and latent variables may be equal for all items. Imposing this restriction, Model (f) which is Model (e) with the restriction that $\sum_{j_i} \nu^2_{i(j_i)m} = 1$ for all four items, yields $G^2 = 5.43$, $df = 8$, and $p = .71$. Lastly, to check whether $\sigma_{12} = 0$, we estimate the uncorrelated latent variable version of Model (f); however this model, Model (g), fails to fit ($G^2 = 97.52$, $df = 9, p < .001$).

TABLE 5
Estimated Parameters from the Multiple Indicator, Two Correlated Latent Variable Model*

	$\hat{\nu}_{i(j_i)1}$	$\hat{\nu}_{i(j_i)2}$
A_1	±.707	.000
A_2	±.789	±.009
B_1	±.102	±.865
B_2	.000	±.707

Note: $\hat{\sigma}_{11} = .520$, $\hat{\sigma}_{12} = .076$, and $\sigma_{22} = 1.00$.
*Model (d) in Table 4, fit to the boys' data.

Our final model for the boys data, Model (f), is a linear-by-linear interaction model with restrictions across the association parameters. The estimated variances (and standard errors[5]) equal $\hat{\sigma}_{11} = .580(.037)$ for attitude and $\hat{\sigma}_{22} = 1.231(.043)$ for membership, and the covariance equals $\hat{\sigma}_{12} = .123(.013)$. Given the identification constraints and restrictions on the scale values, the category scores for the two levels of each variable equal $-.707$ for $j = 1$ (i.e., "negative" or "no") and $.707$ for $j = 2$ (i.e., "positive" or "yes").

5. MODELS WITH HETEROGENEOUS COVARIANCE MATRICES

In the models considered so far, $\mathbf{\Sigma}$ has been restricted to be constant or homogeneous across levels of the discrete variables. We further generalize the models by allowing $\mathbf{\Sigma}$ to differ over cells of the cross-classification of the discrete variables. To make this generalization, we replace $\mathbf{\Sigma}$ in the joint distribution by $\mathbf{\Sigma}(\mathbf{a})$. Using the canonical form given in equation (12), we obtain

$$f(\mathbf{a}, \boldsymbol{\theta}) = \exp[g(\mathbf{a}) + \mathbf{h}(\mathbf{a})'\boldsymbol{\theta} - \tfrac{1}{2}\boldsymbol{\theta}'\mathbf{\Sigma}(\mathbf{a})^{-1}\boldsymbol{\theta}], \tag{27}$$

where

$$\mathbf{h}(\mathbf{a}) = \mathbf{\Sigma}(\mathbf{a})^{-1}\boldsymbol{\mu}(\mathbf{a}), \tag{28}$$

and

$$g(\mathbf{a}) = \log(P(\mathbf{a})) - \frac{M}{2}\log(2\pi) - \frac{1}{2}\log(|\mathbf{\Sigma}(\mathbf{a})|) - \frac{1}{2}\mathbf{h}(\mathbf{a})'\mathbf{\Sigma}(\mathbf{a})\mathbf{h}(\mathbf{a}) \tag{29}$$

(see Lauritzen and Wermuth 1989; Edwards 1995; Lauritzen 1996; Whittaker 1990). The model for observed data is obtained by rewriting equation (29) in terms of $P(\mathbf{a})$,

$$P(\mathbf{a}) = (2\pi)^{M/2}|\mathbf{\Sigma}(\mathbf{a})|^{1/2}\exp[g(\mathbf{a}) + \tfrac{1}{2}\mathbf{h}(\mathbf{a})'\mathbf{\Sigma}(\mathbf{a})\mathbf{h}(\mathbf{a})]. \tag{30}$$

[5]The estimated standard errors from multidimensional Newton-Raphson and from the jackknife of the unidimensional Newton procedure are equal to within $\pm.0001$.

Much of what is true for the homogeneous models is also true for heterogeneous models. Hypotheses about the relationship between observed variables given the latent variables is incorporated through the parameterization of $g(\mathbf{a})$, hypotheses about the relationship between the observed and latent variables are incorporated into the models through $\mathbf{h}(\mathbf{a})$, and hypotheses about the relationship between the latent variables are incorporated through $\boldsymbol{\Sigma}(\mathbf{a})$. The identification constraints given in Section 2.3 are still required. Whether additional constraints are required depends on how the covariance matrix differs over $\mathbf{a}$. Furthermore, the log-multiplicative models for heterogeneous models can be read from graphs (see Appendix A).

What is different between homogeneous and heterogeneous models is that we must specify how the covariance matrix differs over $\mathbf{a}$. Heterogeneous models may include extra terms relative to homogeneous models due to $|\boldsymbol{\Sigma}(\mathbf{a})|^{1/2}$ in equation (30). For homogeneous models, $|\boldsymbol{\Sigma}(\mathbf{a})|^{1/2} = |\boldsymbol{\Sigma}|^{1/2}$ and it is absorbed into the constant λ. In heterogeneous models, depending on how the covariance matrix differs over cells of the table, $|\boldsymbol{\Sigma}(\mathbf{a})|^{1/2}$ may be absorbed into other terms in the log-multiplicative model or may require the addition of extra parameters. For example, if the covariance matrix differs over the categories of just one observed variable, then $|\boldsymbol{\Sigma}(\mathbf{a})|^{1/2}$ is absorbed into the marginal effect term for that variable. As another example, if the covariance matrix is different for a single cell in the table, then there is one value of $|\boldsymbol{\Sigma}(\mathbf{a})|^{1/2}$ for the single cell and another value of $|\boldsymbol{\Sigma}(\mathbf{a})|^{1/2}$ for the rest of the table. Only one element of $\boldsymbol{\Sigma}(\mathbf{a})$ needs to differ and the single cell will be fit perfectly. In such cases, a parameter needs to be included in the log-multiplicative model such that the cell is fit perfectly (e.g., $\tau\delta_{\mathbf{a}}$ where the indicator $\delta_{\mathbf{a}} = 1$ if $\mathbf{a}$ is the cell with the different covariance matrix, and 0 otherwise).

For graphical/latent variable models with homogeneous and heterogeneous covariance matrices, there is always a log-linear model that provides a baseline (best fit) for a log-multiplicative model.[6] With homogeneous covariance matrices, only bivariate associations are implied for the observed (discrete) variables, and the best fit that could be achieved by a log-multiplicative model is given by some log-linear model with two-

[6]Given enough latent variables, the log-multiplicative model derived from a graphical model will be equivalent to some log-linear model, which implies that a graphical representation of any log-linear model can always be found provided that one is willing to assume the existence of underlying continuous variables.

way interactions. With heterogeneous covariance matrices, three- or higher-way interactions may be present depending on how $\mathbf{\Sigma}(\mathbf{a})$ differs over $\mathbf{a}$.

Since there are many possible ways in which the covariance matrix could differ over levels of the discrete variables, we proceed with an example that requires heterogeneous covariance matrices.

6. THE COLEMAN PANEL DATA REVISITED

In Section 6.1, we analyze the Coleman (1964) data for the girls, and in Section 6.2, we analyze the boys and girls data together with gender as a fifth variable.

6.1. *Girls Data*

For the girls data, we repeat the same analyses performed on the boys data in Section 4.2. It is reasonable to expect that the same structural model should fit both the girls and boys data; however, the simplest latent variable model that fits the girls data is Model (d), the two correlated, multiple indicator model given in Figure 4. Models (e) and (f), the latter of which was the best one for the boys data, fail to fit the girls data;[7] however, the lack-of-fit appears to be due to one cell. The response pattern A_1 = "negative," B_1 = "no," A_2 = "positive," and B_2 = "yes"—i.e., the (1,1,2,2) cell—has a relatively large residual.

For the girls, the covariance matrix for the (1,1,2,2) cell may not equal the one for all the other response patterns. If so, then as discussed in Section 5, we could add a single parameter, τ, to fit the cell perfectly. Refitting all the models adding the term $\tau\delta_{1122}$—where $\delta_{1122} = 1$ for cell (1,1,2,2) and 0 otherwise—greatly improves the fit of Models (d), (e), and (f) for the girls data.[8] Of the models that include the extra term, the best model for the girls data is Model (f).

It would be desirable to compare the boys and girls conditional mean values on the attitude and membership perception (latent) variables;

[7] We could argue that Model (f) is the best, because taking sample size and model complexity into account the most parsimonious model is Model (f). The *BIC* statistics for Models (d), (e), and (f) equal −31.75, −31.41, and −41.42, respectively. Furthermore, Model (f) fits well based on the dissimilarity index for models (d), (e) and (f), which equal .016, .021, and .026, respectively.

[8] BIC statistics for Models (d), (e), and (f) with the τ parameter equal −27.92, −35.23, and −46.89, respectively, and the dissimilarity indices equal .013, .014, and .016, respectively. These statistics again point to Model (f) as the best.

however, to make such comparisons regarding the mean values, gender must be included as an observed variable in the model. An additional reason to include gender in the model is to test whether $\boldsymbol{\Sigma}_{\text{boys}} = \boldsymbol{\Sigma}_{\text{girls}}$. The estimates of elements of $\boldsymbol{\Sigma}$ for the girls are slightly larger than those for the boys. The estimates (and standard errors) for the girls are $\hat{\sigma}_{11,\text{girls}} = .760(.040)$, $\hat{\sigma}_{22,\text{girls}} = 1.586(.052)$, and $\hat{\sigma}_{12,\text{girls}} = .138(.014)$, whereas for the boys, they are $\hat{\sigma}_{11,\text{boys}} = .580(.037)$, $\hat{\sigma}_{22,\text{boys}} = 1.231(.043)$, and $\hat{\sigma}_{12,\text{boys}} = .123(.013)$.

6.2. *Combined Analysis*

Given the results from separately estimating models for the boys and girls, we expect that A_1 and A_2 are related to an unobserved attitude variable, B_1 and B_2 are related to an unobserved membership perception variable, and scale restrictions can be imposed on the scale values for A_1, A_2, B_1, and B_2. We would like to test whether the means of the unobserved variables differ for boys and girls and whether $\boldsymbol{\Sigma}$ differs. This underlying model is shown in Figure 5.

To derive the most general log-multiplicative model for the figure, we define $g(\mathbf{a})$ as

$$g(\mathbf{a}) = \lambda + \lambda_{A_1(j)} + \lambda_{A_2(j)} + \lambda_{B_1(j)} + \lambda_{B_2(j)} + \lambda_{G(j)}, \tag{31}$$

where λ is a constant, and $\lambda_{A_1(j)}$, $\lambda_{A_2(j)}$, $\lambda_{B_1(j)}$, $\lambda_{B_2(j)}$, and $\lambda_{G(j)}$ are marginal effect terms for the observed variables. For simplicity, we have

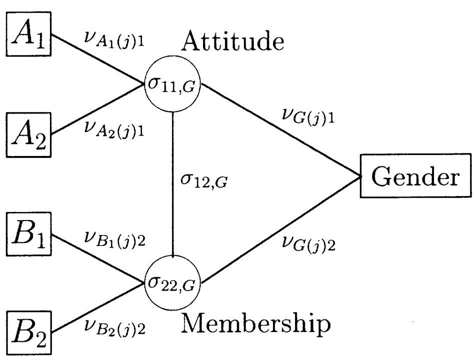

FIGURE 5. Graph corresponding to log-multiplicative Models (c–f) in Table 6 fit to the Coleman panel with gender as the fifth variable.

dropped the subscripts on the j indices. We parameterize $\mathbf{h}(\mathbf{a})$ as

$$\mathbf{h}(\mathbf{a}) = \begin{pmatrix} \nu_{A_1(j)1} + \nu_{A_2(j)1} + \nu_{G(j)1} \\ \nu_{B_1(j)2} + \nu_{B_2(j)2} + \nu_{G(j)2} \end{pmatrix}. \tag{32}$$

The first row in equation (32) equals the sum of the scale values for the unobserved attitude variable and the second row equals the sum of the scale values for the unobserved membership variable. Lastly, we specify a heterogeneous covariance matrix:

$$\mathbf{\Sigma}_{G(j)} = \begin{pmatrix} \sigma_{11G(j)} & \sigma_{12G(j)} \\ \sigma_{12G(j)} & \sigma_{22G(j)} \end{pmatrix}, \tag{33}$$

where this matrix is different for $j = 1$ (boys) and 2 (girls). For the homogeneous models, we set $\mathbf{\Sigma}_{G(j)} = \mathbf{\Sigma}$. Replacing $g(\mathbf{a})$, $\mathbf{h}(\mathbf{a})$, and $\mathbf{\Sigma}(\mathbf{a})$ in equation (30) by their parameterizations in equations (31), (32), and (33), respectively, yields

$$\begin{aligned} \log(P(\mathbf{a})) = {} & \lambda + \lambda_{A_1(j)} + \lambda_{A_2(j)} + \lambda_{B_1(j)} + \lambda_{B_2(j)} + \lambda^*_{G(j)} \\ & + \tfrac{1}{2}\sigma_{11G(j)}[\nu^2_{A_1(j)1} + \nu^2_{A_2(j)1}] + \tfrac{1}{2}\sigma_{22G(j)}[\nu^2_{B_1(j)2} + \nu^2_{B_2(j)2}] \\ & + \sigma_{11G(j)}[\nu_{A_1(j)1}\nu_{A_2(j)1} + \nu_{A_1(j)1}\nu_{G(j)1} + \nu_{A_2(j)1}\nu_{G(j)1}] \\ & + \sigma_{22G(j)}[\nu_{B_1(j)2}\nu_{B_2(j)2} + \nu_{B_1(j)2}\nu_{G(j)2} + \nu_{B_2(j)2}\nu_{G(j)2}] \\ & + \sigma_{12G(j)}[\nu_{A_1(j)1}\nu_{B_1(j)2} + \nu_{A_1(j)1}\nu_{B_2(j)2} + \nu_{A_2(j)1}\nu_{B_1(j)2} \\ & \quad + \nu_{A_2(j)1}\nu_{B_2(j)2} + \nu_{A_1(j)1}\nu_{G(j)2} + \nu_{A_2(j)1}\nu_{G(j)2} \\ & \quad + \nu_{B_1(j)2}\nu_{G(j)1} + \nu_{B_2(j)2}\nu_{G(j)1}], \end{aligned} \tag{34}$$

where $\lambda^*_{G(j)} = \lambda_{G(j)} + \log(|\mathbf{\Sigma}_{G(j)}|^{1/2}) + (1/2)\sum_{m=1}^{2}\sum_{m'=1}^{2}\sigma_{mm'G(j)} \times \nu_{G(j)m}\nu_{G(j)m'}$. While this log-multiplicative model is quite complex, its interpretation is relatively simple and greatly facilitated by Figure 5. The model can be read from its graph using the method outlined in Appendix A.

Based on the previous results, we set $\nu_{A_1(j)1}$, $\nu_{A_2(j)1}$, $\nu_{B_1(j)2}$, and $\nu_{B_2(j)2}$ equal to $\pm.7071$ rather than estimating them. Thus the only scale values estimated are those for gender, $\nu_{G(j)m}$. Other than location constraints on the marginal effects and the scale values for gender, no additional identification constraints are required on the parameters in either the homogeneous or heterogeneous versions of equation (34).

The fit statistics for models with gender as a fifth observed variable are reported in Table 6. While the all two-way interaction log-linear model is the baseline model for the homogeneous version of equation (34), the log-linear model with all three-way interactions that involve gender (A_1A_2G,A_1B_1G,A_1B_2G,A_2B_1G,A_2B_2G,B_1B_2G), is the baseline model for the heterogeneous version of equation (34). Since the all two-way interaction log-linear model, Model (a) in Table 6, fails to fit, the homogeneous latent variable model, Model (c), should also fail. Not only does the homogeneous model fail to fit, but so does the homogeneous model with an extra parameter for the (1,1,2,2) cell for the girls—i.e., $\tau\delta_{1122,G(j)}$ where $\delta_{1122,\text{girls}} = 1$ for the (1,1,2,2) cell for the girls, and 0 otherwise.

Since the log-linear model with the three-way interactions, Model (b), fits the data, we try a heterogeneous model where the covariance matrix differs for boys and girls. The heterogeneous model nearly fits the data ($G^2 = 30.39$, $df = 18$, $p = .03$), and when $\tau\delta_{1122,G(j)}$ is added to the model, the model clearly fits ($G^2 = 19.47$, $df = 17$, $p = .30$). Model (f) is the most parsimonious model that fits the data, so we select it as our final model.

The estimated parameters for Model (f) are given in Table 7. The estimated covariance matrices for the boys and girls are similar to those from the models estimated separately for the boys and girls. Given the scale values and estimated covariance matrices, we compute estimates of the mean values on the latent attitude and membership variables for the cells of the cross-classification of the observed variables using $\boldsymbol{\mu}(\mathbf{a}) = \boldsymbol{\Sigma}(\mathbf{a})\mathbf{h}(\mathbf{a})$ (see equation 28). Since there are only two levels of the variables A_1, A_2, B_1, and B_2 and their scale values are equal, there are only five unique values of the means for the boys and five for the girls. Cells that have the same number of positive responses and yes's have the same mean—for example, the conditional mean for the cell (2,2,2,1) is the same as the mean for (1,2,2,2). The estimated conditional means for attitude and membership perception are plotted in Figure 6 against the numbers 0 through 4, which equal the number of positive responses and yes's. Separate curves are given for boys and girls.

From Figure 6, we see that for response patterns with more negative responses and no's, the boys means are larger than the girls means, while for response patterns with more positive responses and yes's, the girls means are larger than the boys. In both figures, the slopes for the girls are larger than those for the boys. The slopes of the lines for boys and girls differ, because $\hat{\boldsymbol{\Sigma}}_{\text{boys}} \neq \hat{\boldsymbol{\Sigma}}_{\text{girls}}$. If $\hat{\boldsymbol{\Sigma}}_{\text{boys}} = \hat{\boldsymbol{\Sigma}}_{\text{girls}}$, then the lines for boys and girls would be parallel and any difference between them would be due to

TABLE 6
Fit Statistics for Models Fit to Coleman (1964) Panel Data with Gender as a Variable

Model	*df*	G^2	*p*	BIC
Baseline loglinear models				
(a) All 2-way interactions	16	56.31	<.001	−84.31
(b) $(A_1A_2G, A_1B_1G, A_1B_2G, A_2B_1G, A_2B_2G, B_1B_2G)$	10	9.60	.48	−78.44
Latent variable models (equation (34), Figure 5)				
(c) Homogeneous $\boldsymbol{\Sigma}$	21	63.41	<.001	−121.47
(d) Model (c) with $\tau\delta_{1122,G(j)}$	20	60.90	<.001	−115.17
(e) Heterogeneous $\boldsymbol{\Sigma}_{G(j)}$	18	30.39	.03	−128.08
(f) Model (e) with $\tau\delta_{1122,G(j)}$	17	19.47	.30	−130.19

TABLE 7
Estimated Parameters from Model (f) in Table 6 Fit to the Panel Data with Gender as a Variable

Parameter	Value(s) $j = 1$	Value(s) $j = 2$	Parameter	Value(s) $j = 1$	Value(s) $j = 2$
λ	4.796		$\lambda_{G(j)}$	.276	−.276
$\lambda_{A_1(j)}$	−.134	.134	$\lambda_{B_1(j)}$	.291	−.291
$\lambda_{A_2(j)}$	−.185	.185	$\lambda_{B_2(j)}$	.118	−.118
$\nu_{G(j)1}$	−.125	.125	$\nu_{G(j)2}$	.060	−.060
$\sigma_{11,\text{boys}}$	.578		$\sigma_{11,\text{girls}}$	.757	
$\sigma_{22,\text{boys}}$	1.228		$\sigma_{22,\text{girls}}$	1.583	
$\sigma_{12,\text{boys}}$	.123		$\sigma_{12,\text{girls}}$	.138	
			τ	.462	

Note: Due to restrictions on the scale values for variables A_1, A_2, B_1, and B_2, the scale values $\nu_{A_1(j)1}$, $\nu_{A_2(j)1}$, $\nu_{B_1(j)2}$, and $\nu_{B_2(j)2}$ equal −.707 for $j = 1$ and .707 for $j = 2$.

the scale values for gender. The positive covariance between attitude and membership is reflected by the fact that the higher a child's perception of being a member of the leading crowd, the more positive his or her attitude is toward the leading crowd (and vice versa).

7. DISCUSSION

Log-multiplicative models provide a powerful and flexible approach to studying the relationships between nominal and/or ordinal variables in

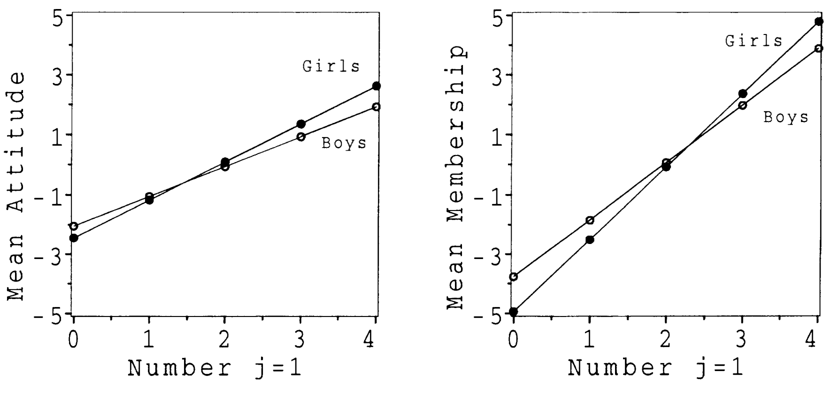

FIGURE 6. Plot of estimated attitude (a) and membership (b) means for boys (circles) and girls (dots) using scale values and estimated covariance matrix from Model (f) in Table 6 fit to the data from Coleman (1964).

terms of unobserved, continuous variables. The approach presented here provides a logical way to incorporate substantive knowledge about a phenomenon into models for studying associations between multiple discrete variables. In the examples presented, we show how to use the models in both an exploratory and confirmatory fashion, how to study group differences in terms of underlying variables, and how to obtain measurements for individuals on the latent variables. Measurement of individuals' values on the latent variables is a byproduct of the estimation of the parameters of log-multiplicative models. Additional possibilities include adding individual level covariates to the models (Anderson and Böckenholt 2000) and imposing inequality restrictions on the category scale values (Ritov and Gilula 1991; Vermunt 1998).

With the conditional Gaussian assumption, the marginal distribution of the continuous variables is a mixture of multivariate normals. This differs from traditional factor analytic and item response theory models where the marginal distribution of latent variables is typically assumed to be multivariate normal. In the traditional models, the conditional distribution within a cell is a mixture of multivariate normals. The models proposed here are alternatives to the more traditional factor analytic models. In some cases, the proposed models may be more appropriate or at least as appropriate as traditional models. Which is better is both a theoretical and an empirical question whose answer depends on the particular phenomenon being studied. A full discussion of the relationships between the latent variable models proposed here and more traditional models is beyond the scope of this paper. Thus, areas for future work include studying the relationship between the log-multiplicative models and traditional factor analytic and item response theory models, and further exploration of the use of log-multiplicative models to estimate individuals' values on latent variables.

APPENDIX A: READING MODELS FROM GRAPHS

Reading log-multiplicative models from graphs is essentially the same for both homogeneous and heterogeneous models. For all models, marginal effect terms are always included for each discrete variable, as well as a constant to ensure that the fitted values sum up to the observed total. In the graphs, the lines connecting the observed and latent variables have been labeled by the corresponding scale values. The interaction terms in the log-multiplicative models equal half the sum of the products of pairs of scale values and the covariance between latent variables from all directed paths between observed variables. There are two types of paths in the graphs: paths from a discrete variable back to itself and paths from one discrete variable to another. Both types of paths may involve either one latent variable or a pair of latent variables.

To illustrate, consider the multiple latent variables per indicator model depicted in Figure 4. The log-multiplicative model for this graph is

$$\begin{aligned}\log(P(\mathbf{a})) = \lambda &+ \lambda^*_{1(j_1)} + \lambda^*_{2(j_2)} + \lambda^*_{3(j_3)} + \lambda^*_{4(j_4)} \\ &+ \sigma_{11}[\nu_{1(j_1)1}\nu_{2(j_2)1} + \nu_{1(j_1)1}\nu_{3(j_3)1} + \nu_{2(j_2)1}\nu_{3(j_3)1}] \\ &+ \sigma_{22}[\nu_{2(j_2)2}\nu_{3(j_3)2} + \nu_{2(j_2)2}\nu_{4(j_4)2} + \nu_{3(j_3)2}\nu_{4(j_4)2}] \\ &+ \sigma_{12}[\nu_{1(j_1)1}\nu_{2(j_2)2} + \nu_{1(j_1)1}\nu_{3(j_3)2} + \nu_{1(j_1)1}\nu_{4(j_4)2} \\ &\quad + \nu_{2(j_2)1}\nu_{3(j_3)2} + \nu_{2(j_2)1}\nu_{4(j_4)2} + \nu_{3(j_3)1}\nu_{2(j_2)2} \\ &\quad + \nu_{3(j_3)1}\nu_{4(j_4)2}],\end{aligned} \tag{35}$$

where $\lambda^*_{i(j_i)} = \lambda_{i(j_i)} + (1/2)\sum_m \sum_{m'} \sigma_{mm'}\nu_{i(j_i)m}\nu_{i(j_i)m'}$.

With respect to paths from a variable back to itself, when the path goes through a single latent variable, this results in terms such as $(1/2)\sigma_{11}\nu^2_{1(j_1)1}$. This term comes from the directed path $A_1 \to \Theta_1 \to A_1$. The covariance of a variable with itself is the variance, so we multiply $(1/2)\nu^2_{1(j_1)1}$ by the variance of Θ_1. Paths from a variable back to itself that involve a pair of latent variables are found only in multiple factors per indicator models. For example, the directed paths $A_2 \to \Theta_1 \to \Theta_2 \to A_2$ and $A_2 \to \Theta_2 \to \Theta_1 \to A_2$ result in the term $(1/2)\sigma_{12}\nu_{2(j_2)1}\nu_{2(j_2)2} + (1/2)\sigma_{12}\nu_{2(j_2)2}\nu_{2(j_2)1} = \sigma_{12}\nu_{2(j_2)1}\nu_{2(j_2)2}$. In homogeneous models, terms that arise from paths from a variable back to itself are absorbed into the marginal effects; however, in heterogeneous models, they are not necessarily absorbed (an example of this is given in Section 6.2).

The second type of path, which connects two different discrete variables, may involve either one latent variable or a pair of correlated latent variables. In the former case, the association parameter is the variance of the latent variable, and in the later, the association parameter is the covariance. For example, the term $\sigma_{11}\nu_{1(j_1)1}\nu_{2(j_2)1}$ results from the directed paths $A_1 \to \Theta_1 \to A_2$ and $A_2 \to \Theta_1 \to A_1$. The term $\sigma_{12}\nu_{1(j_1)1}\nu_{2(j_2)2}$ results from the directed paths $A_1 \to \Theta_1 \to \Theta_2 \to A_2$ and $A_2 \to \Theta_2 \to \Theta_1 \to A_1$.

As discussed in Section 5, some heterogeneous models may include extra terms due to $|\boldsymbol{\Sigma}(\mathbf{a})|^{1/2}$ in equation (30).

APPENDIX B: MAXIMUM-LIKELIHOOD ESTIMATION

The maximum-likelihood estimation of the parameters of the log-multiplicative models presented in this paper is described here. The starting point is the most general, homogeneous latent variable model, which was given in equation (25). All other (homogeneous) latent variable models can be derived from this model by imposing fixed-value restrictions on some parameters—for instance, by fixing particular sets of category scores to zero, particular variances to one, or particular covariances to zero. The heterogeneous models can be estimated by the same procedure described here. The only difference is that some of the maximum-likelihood equations differ slightly.

Assuming either a multinomial or Poisson sampling scheme, the likelihood equations for the parameters $\lambda^*_{i(j_i)}$, σ_{mm}, $\sigma_{mm'}$, and $\nu_{i(j_i)m}$, which equal zero at the maximum value of the likelihood function, are

$$\frac{\partial \log L}{\partial \lambda^*_{i(j_i)}} = \sum_{\mathbf{a}|j_i} [n(\mathbf{a}) - P(\mathbf{a})],$$

$$\frac{\partial \log L}{\partial \sigma_{mm}} = \sum_{\mathbf{a}} \sum_{i} \sum_{k>i} \nu_{i(j_i)m} \nu_{k(j_k)m} [n(\mathbf{a}) - P(\mathbf{a})],$$

$$\frac{\partial \log L}{\partial \sigma_{mm'}} = \sum_{\mathbf{a}} \sum_{i} \sum_{k \neq i} \nu_{i(j_i)m} \nu_{k(j_k)m'} [n(\mathbf{a}) - P(\mathbf{a})],$$

$$\frac{\partial \log L}{\partial \nu_{i(j_i)m}} = \sum_{\mathbf{a}|j_i} \sum_{k \neq i} \sum_{m'} \sigma_{mm'} \nu_{k(j_k)m'} [n(\mathbf{a}) - P(\mathbf{a})],$$

respectively. Here, $n(\mathbf{a})$ denotes an observed cell entry, $\sum_{\mathbf{a}}$ indicates the summation over all cells, and $\sum_{\mathbf{a}|j_i}$ indicates the summation over the cells in which variable A_i has the value $a_{i(j_i)}$.

A simple algorithm to solve these maximum-likelihood equations is the unidimensional Newton algorithm. This procedure is implemented in an experimental version of the program ℓ_{EM} (Vermunt 1997). We have found that this iterative method, which has also been used by others to obtain ML estimates of log-multiplicative models (for instance, see Goodman 1979; Clogg 1982; Becker 1989), works well for the models discussed in this paper. The method involves updating one parameter at a time fixing the other parameters at their current value. A unidimensional Newton update of a particular parameter, say γ, at the tth iteration cycle is of the form

$$\gamma^{(t)} = \gamma^{(t-1)} - \frac{\partial \log L/\partial\gamma}{\partial^2 \log L/\partial^2\gamma},$$

where the derivatives are evaluated at the current values of all model parameters. The relevant second-order derivatives for the parameters appearing in equation (25) are

$$\frac{\partial^2 \log L}{\partial^2 \lambda^*_{i(j_i)}} = -\sum_{\mathbf{a}|j_i} P(\mathbf{a}),$$

$$\frac{\partial^2 \log L}{\partial^2 \sigma_{mm}} = -\sum_{\mathbf{a}}\sum_{i}\sum_{k>i} [\nu_{i(j_i)m}\nu_{k(j_k)m}]^2 P(\mathbf{a}),$$

$$\frac{\partial^2 \log L}{\partial^2 \sigma_{mm'}} = -\sum_{\mathbf{a}}\sum_{i}\sum_{k\neq i} [\nu_{i(j_i)m}\nu_{k(j_k)m'}]^2 P(\mathbf{a}),$$

$$\frac{\partial^2 \log L}{\partial^2 \nu_{i(j_i)m}} = -\sum_{\mathbf{a}|j_i}\sum_{k\neq i}\sum_{m'} [\sigma_{mm'}\nu_{k(j_k)m'}]^2 P(\mathbf{a}).$$

The location and scaling constraints, which are necessary for identification, can be imposed at each iteration cycle after updating a particular set of λ or ν parameters.

As mentioned in Section 2.3, we sometimes might want to impose a scaling condition on a particular set of the ν parameters that is not necessary for identification. Suppose that the scaling of the mth set of category scores for variable A_i is a model restriction. In such a situation, we have to work with Lagrange terms to obtain the restricted ML solution. The Lagrange likelihood equations for the $\nu_{i(j_i)m}$ parameters, which equal zero at the saddle point of the Lagrange likelihood function, are

$$\frac{\partial \log L}{\partial \nu_{i(j_i)m}} + \beta_{im1} + 2\nu_{i(j_i)m}\beta_{im2}.$$

Here, β_{im1} and β_{im2} are the Lagrange parameters corresponding to the location and scaling restrictions (i.e., $\sum_{j_i} \nu_{i(j_i)m} = 0$ and $\sum_{j_i} \nu^2_{i(j_i)m} = 1$).

Only a slight modification of the unidimensional Newton method is needed with these types of restrictions. Setting the Lagrange likelihood equations for the $\nu_{i(j_i)m}$'s equal to zero, we can compute β_{im1} and β_{im2} by a simple linear regression. This can be seen by rewriting the resulting equations as

$$-\frac{\partial \log L}{\partial \nu_{i(j_i)m}} = \beta_{im1} + 2\nu_{i(j_i)m}\beta_{im2}. \tag{36}$$

The provisional values for β_{im1} and β_{im2} can be obtained by regressing the term on the left-hand side of equation (36) on $2\nu_{i(j_i)m}$. After obtaining new Lagrange terms, the ν's are updated and subsequently centered and rescaled. A nice feature of the Lagrange terms is that they converge to zero if the corresponding location or scaling constraint is necessary for identification. In the models presented in this paper, this is always the case for the location constraints but not always for the scaling conditions.

Since the log-likelihood function of log-multiplicative models is not concave, there may be local maxima. Therefore, models should be estimated multiple times using different sets of random starting values to prevent reporting a local solution.

Contrary to multidimensional Newton methods, the above simple estimation method does not provide standard errors or covariances of the parameter estimates as a by-product. Asymptotic standard errors and covariances of parameter estimates can be obtained by means of jackknifing, which is a method that has been used by a number of authors for this purpose in the context of log-multiplicative models (e.g., Anderson and Böckenholt 2000; Clogg and Shihadeh 1994; Eliason 1995).

REFERENCES

Afifi, Abdelmonem A., and R.M. Elashoff. 1969. "Multivariate Two Sample Tests with Dichotomous and Continuous Variables. I. The Location Model." *Annals of Mathematical Statistics* 40:290–98.

Agresti, Alan. 1984. *Analysis of Ordinal Categorical Data*. New York: Wiley.

———. 1997. "A Model for Repeated Measurements of a Multivariate Binary Response." *Journal of the American Statistical Association* 92:315–21.

Andersen, Erling B. 1988. "Comparison of Latent Structure Models." Pp. 207–29 in *Latent Trait and Latent Class Models*, edited by R. Langeheine and J. Rost. New York: Plenum Press.

Anderson, Carolyn J., and Ulf Böckenholt. In press. "Graphical Regression Models for Polytomous Variables." *Psychometrika*.

Becker, Mark P. 1989. "Models for the Analysis of Association in Multivariate Contingency Tables." *Journal of the American Statistical Association* 84:1014–19.

Clogg, Clifford C. 1982. "Some Models for the Analysis of Association in Multiway Cross-classifications Having Ordered Categories." *Journal of the American Statistical Association* 77:803–15.

Clogg, Clifford C., and Edward Shihadeh. 1994. *Statistical Models for Ordinal Variables*. Thousand Oaks, CA: Sage.

Coleman, James S. 1964. *Introduction to Mathematical Sociology*. Glencoe, IL: Free Press.

Davis, James A., and Tom W. Smith. 1996. *General Social Surveys 1972–1996: Cumulative Codebook*. Chicago, IL: National Opinion Research Center.

Edwards, David. 1995. *Introduction to Graphical Modelling*. New York: Springer-Verlag.

Eliason, Scott R. 1995. "Modeling Manifest and Latent Dimensions of Association in Two-way Cross-classifications." *Sociological Methods and Research* 24:30–67.

Goodman, Leo A. 1978. *Analyzing Qualitative/Categorical Data: Log-linear Models and Latent Structure Analysis*. London: Addison-Wesley.

———. 1979. "Simple Models for the Analysis of Association in Cross-classifications Having Ordered Categories." *Journal of the American Statistical Association* 74:537–52.

———. 1985. "The Analysis of Cross-classified Data Having Ordered and/or Unordered Categories: Association Models, Correlation Models, and Asymmetry Models for Contingency Tables with or without Missing Entries." *Annals of Statistics* 13:10–69.

Krzanowski, Wojtek J. 1980. "Mixtures of Continuous and Categorical Variables in Discriminant Analysis." *Biometrics* 36:493–99.

———. 1983. "Distance Between Populations Using Mixed Continuous and Categorical Variables." *Biometrika* 70:235–43.

———. 1988. *Principles of Multivariate Analysis*. New York: Oxford Press.

Langeheine, Rolf. 1988. "New Developments in Latent Class Theory." Pp. 77–108 in *Latent Trait and Latent Class Models*, edited by Rolf Langeheine and Jürgen Rost. New York: Plenum Press.

Lauritzen, Steffen L. 1996. *Graphical Models*. New York: Oxford University Press.

Lauritzen, Steffen L., and Nancy Wermuth. 1989. "Graphical Models for Associations Between Variables, Some of Which are Qualitative and Some Quantitative." *The Annals of Statistics* 17:31–57.

Olkin, Ingram, and R.F. Tate. 1960. "Multivariate Correlation Models with Mixed Discrete and Continuous Variables." *The Annals of Mathematical Statistics* 32: 448–65.

Ritov, Yaacov, and Zvi Gilula. 1991. "The Order-restricted *RC* Model for Ordered Contingency Tables: Estimation and Testing for Fit." *Annals of Statistics* 19: 2090–101.

Vermunt, Jeroen K. 1997. ℓ_{EM}: A General Program for the Analysis of Categorical Data. The Netherlands: Tilburg University. Internet: http://cwis.kub.nl/~fsw_1/mto/.

———. 1998. "*RC* Association Models with Ordered Row and Column Scores: Estimation and Testing." Presented at the 21st Biennial Conference of the Society for Multivariate Analysis in the Behavioral Sciences, Leuven, Belgium, July 15–17, 1998.

Wermuth, Nancy, and Steffen L. Lauritzen. 1990. "Discussion of Papers by Edwards, Wermuth, and Lauritzen." *Journal of the Royal Statistical Society* ser. B, 52:51–72.

Whittaker, Joe. 1989. "Discussion of Paper by van der Heijden, de Falguerolles, and de Leeuw." *Applied Statistics* 38:278–79.

———. 1990. *Graphical Models in Applied Mathematical Multivariate Statistics*. New York: Wiley.