Log-multiplicative association models as latent variable models for nominal and/or ordinal data

Anderson, C.; Vermunt, J.K.

Published in:
Sociological Methodology

Publication date:
2000

Document Version
Peer reviewed version

Citation for published version (APA):
Anderson, C., & Vermunt, J. K. (2000). Log-multiplicative association models as latent variable models for nominal and/or ordinal data. *Sociological Methodology*, *30*, 81-122.

Log-Multiplicative Association Models as Latent Variable Models for Nominal and/or Ordinal Data *

Carolyn J. Anderson
and
Jeroen K. Vermunt

July 8, 1999

*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, MC-708, Champaign, IL, 61820 or Jeroen K. Vermunt, Deparment 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.

Abstract

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-multipicative models as latent variable models for discrete variables are presented in this paper. The models are based on two major assumptions: (1) Observed variables are conditionally independent given the unobserved variables, and (2) the conditional distribution of the unobserved variables is multivariate normal. From these assumptions, special cases of a general model are discussed. The models have many desirable properties, including having schematic or graphical representations of the system of observed and unobserved variables from which the log-multiplicative models can be read, providing estimates of the means, variances and covariances of the latent variables given values on the observed variables, and the models can be fit by marginal maximum likelihood estimation without the use of multiple, numerical integrations. 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 & Smith, 1996) are modeled, and in the other example, panel data (Coleman, 1964) are analyzed.

Keywords: Log-linear models, graphical models, $RC(M)$ association model, conditional Gaussian distribution, marginal maximum likelihood estimation.

1 Introduction

Associations in multivariate categorical data are often due to collapsing over other variables. When the variables collapsed over are continuous and are either unobserved or not directly measurable, models that represent the observed associations in terms of the unobserved or latent variables greatly facilitate the description and interpretation of the multiple, observed associations. Furthermore, such models allow one to study the underlying structural relationships between the unobserved variables, because the observed variables act as indicators of the unobserved ones. Ideally, the latent variable models should permit researchers to transform their specific theories and hypotheses about the associations between both the observed and unobserved variables into statistical models, which in turn can be readily fit to observed data.

The proposed latent variable models belong to a family of "location models" for discrete and continuous variables (Olkin & Tate, 1960; Afifi & Elashoff, 1969; Kraznowski, 1980, 1983, 1988). The models presented here differ from previously discussed location models 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. Adding the assumption that the distribution of the continuous variables given the discrete ones is multivariate normal, the model implied for the observed data is a log-multiplicative association model, which is an extension of a log-linear model.

In log-multiplicative models, associations between 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 & Shihadeh, 1994; Goodman, 1979, 1985).

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

The models developed here have many desirable properties, including having schematic or graphical representations. Not only are the graphs useful pictorial representations of theories about phenomenon, the corresponding log-multiplicative model can be read from the graph. In many cases, estimates of the conditional means, variances and covariances of the latent variables are by-products of the estimation of the log-multiplicative model parameters. Furthermore, the models can be fit by marginal maximum likelihood estimation without the use of multiple, numerical integrations.

In Section 2, the general latent variable model is proposed and the corresponding log-multiplicative model is derived. In Section 3, we discuss special cases of the general model and develop theory that permits log-multiplicative models to be used as latent variable models for a variety of underlying structural models. In Section 4, the estimation of log-multiplicative models is dicussed with a special emphasis on estimating such models with restrictions. In Section 5, two examples are presented to illustrate some of the advantageous aspects of these models.

2 Latent Variable Model for Discrete Observations

Our first assumption is that observed, discrete variables are conditionally independent given the latent, continuous variables. While we do not discuss models that have conditional dependencies between some of the observed variables, such models can be derived by treating the observed variables that are conditionally dependent as a single variable whose levels correspond to combinations of the levels of the individual, conditionally dependent variables.

For our second assumption, consider the table formed by cross-classifying observations according to the discrete (observed) variables. Individuals within the same cell of a table typically differ in terms of their values on the continuous variables, but on average have similar values. To model these individual differences, we assume that within cells of the table, the continuous variables are multivariate normal where the means differ across cells of the table and the covariance matrix may also differ. The conditional multivariate normal assumption for the continuous variables is made in "location models" for discrete and continuous variables (Olkin & Tate, 1960; Afifi & Elashoff, 1969; Kraznowski, 1980, 1983, 1988); however, in our case, the continuous variables are unobserved. The joint distribution of the discrete and continuous variables is a conditional Gaussian distribution (Lauritzen & Wermuth, 1989; Lauritzen, 1996).

With the conditional Gaussian assumption, the marginal distribution of the discrete variables is multinomial and 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 latter case, 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.

Let $\mathcal{A}$ and $\boldsymbol{\Theta}$ denote sets of I observed (discrete) and M latent (continuous) random variables, respectively, and the vectors $\mathbf{a} = (a_{1(j_1)}, \ldots, a_{I(j_I)})'$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)'$ denote realizations of the variables in $\mathcal{A}$ and $\boldsymbol{\Theta}$. The levels of the discrete variable A_i are indexed by j_i where $j_i = 1, \ldots, J_i$. The joint distribution of the observed and latent variables is

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

where $g(\mathbf{a})$ is a function of $\mathbf{a}$, $\mathbf{h}(\mathbf{a})$ is an $(M \times 1)$ vector valued function of $\mathbf{a}$, and $\boldsymbol{\Sigma}(\mathbf{a})$ is the $(M \times M)$ covariance matrix of $\boldsymbol{\Theta}$ which may be a function of $\mathbf{a}$ (Lauritzen & Wermuth, 1989; Lauritzen, 1996).

The dependencies among the discrete variables after controlling for the continuous variables are represented through the function $g(\mathbf{a})$. Since the discrete variables are assumed to be independent given the continuous ones, $g(\mathbf{a})$ is defined as

$$g(\mathbf{a}) = \sum_{i=1}^{I} \lambda_{i(j_i)} \tag{2}$$

where the $\lambda_{i(j_i)}$'s are marginal or main effect terms for the discrete variables.

The dependencies between the discrete and continuous variables are represented through the function $\mathbf{h}(\mathbf{a})$. To keep the model as general as possible, a discrete variable can be related to any continuous variable; therefore, $\mathbf{h}(\mathbf{a})$ is defined as

$$\mathbf{h}(\mathbf{a}) = (\sum_{i=1}^{I} \nu_{i(j_i)1}, \ldots, \sum_{i=1}^{I} \nu_{i(j_i)M})' \tag{3}$$

where $\nu_{i(j_i)m}$ is the category score or scale value for level j_i of variable A_i for latent variable Θ_m. If A_i and Θ_m are related, then $\nu_{i(j_i)m} \neq 0$ for at least one j_i, and if A_i and Θ_m are unrelated, then $\nu_{i(j_i)m} = 0$ for all $j_i = 1, \ldots, J_i$. The scale values may be estimated from data or specified a priori.

The scale values provide information about the conditional means of the latent variables. The means are a linear function of the scale values and the variances and covariances of the latent variables. Specifically, let $\boldsymbol{\mu}(\mathbf{a})$ equal the $(M \times 1)$ mean vector of the latent variables for cell $\mathbf{a}$, then

$$\boldsymbol{\mu}(\mathbf{a}) = \boldsymbol{\Sigma}(\mathbf{a})\mathbf{h}(\mathbf{a}) \tag{4}$$

(Lauritzen, 1996; Lauritzen and Wermuth, 1989). Thus, given estimates of the scale values and the covariance matrix, we can estimate the conditional means of the latent variables.

To obtain a model for the observed data, we integrate equation (1) over $\boldsymbol{\theta}$, which yields

$$P(\mathbf{a}) = (2\pi)^{M/2} |\boldsymbol{\Sigma}(\mathbf{a})|^{1/2} \exp\left[g(\mathbf{a}) + \frac{1}{2}\mathbf{h}(\mathbf{a})'\boldsymbol{\Sigma}(\mathbf{a})\mathbf{h}(\mathbf{a})\right] \tag{5}$$

where $P(\mathbf{a})$ equals the probability of observing $\mathbf{a}$ (Lauritzen & Wermuth 1989; Lauritzen, 1996). Since there are many possible models that may be specified for $\boldsymbol{\Sigma}(\mathbf{a})$, we will emphasize the homogeneous case in which the covariance matrix does not depend on $\mathbf{a}$ (i.e., $\boldsymbol{\Sigma}(\mathbf{a}) = \boldsymbol{\Sigma}$). Much of what is true for the homogeneous models is also true for heterogeneous ones. In our discussions, we point out aspects of the models that differ for the heterogeneous case; however, exactly how they differ depends on the specific model assumed for $\boldsymbol{\Sigma}(\mathbf{a})$. Additionally, we present a detailed example of a heterogeneous model in Section 5.2 where $\boldsymbol{\Sigma}$ depends on the categories of one variable.

Setting $\boldsymbol{\Sigma}(\mathbf{a}) = \boldsymbol{\Sigma}$ and replacing $g(\mathbf{a})$ and $\mathbf{h}(\mathbf{a})$ in equation (5) by their definitions in (2) and (3) yields

$$\begin{aligned} \log(P(\mathbf{a})) &= \lambda + \sum_i \lambda^*_{i(j_i)} + \sum_i \sum_{k>i} \sum_m \sigma_{mm} \nu_{i(j_i)m} \nu_{k(j_k)m} \\ &\quad + \sum_i \sum_{k \neq i} \sum_m \sum_{m'>m} \sigma_{mm'} \nu_{i(j_i)m} \nu_{k(j_k)m'} \end{aligned} \tag{6}$$

where λ is a normalizing constant and $\lambda^*_{i(j_i)} = \lambda_{i(j_i)} + (1/2) \sum_m \sum_{m'} \sigma_{mm'} \nu_{i(j_i)m} \nu_{i(j_i)m'}$. Since the term $(1/2) \sum_m \sum_{m'} \sigma_{mm'} \nu_{i(j_i)m} \nu_{i(j_i)m'}$ is only indexed by i, it gets absorbed into the marginal effect terms. In heterogeneous models, terms such as these are not necessarily absorbed into the marginal effects. An example of this is given in Section 5.2.

Equation (6) is a log-multiplicative association model with bivariate interactions between all pairs of variables. The best fit that can be attained using equation (6) is given by the all 2–way interaction log-linear model (see Becker, 1989). For heterogeneous covariance matrices, some 3–way or higher–way interactions may be present depending on how $\boldsymbol{\Sigma}(\mathbf{a})$ differs over $\mathbf{a}$. For both homogeneous and heterogeneous cases, there is always some log-linear model that provides a baseline for a log-multiplicative model[1].

For $I = 2$ and $M < \min(J_1, J_2) - 1$ uncorrelated latent variables, equation (6) reduces

[1] The converse is also true; that is, 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.

to the $RC(M)$ association model,

$$\log(P_{j_1,j_2}) = \lambda + \lambda^*_{1(j_1)} + \lambda^*_{2(j_2)} + \sum_{m=1}^{M} \phi_m \nu_{1(j_1)m} \nu_{2(j_2)m} \tag{7}$$

where the association parameter ϕ_m is the variance of latent variable m. However, if there are M uncorrelated pairs of correlated latent variables, equation (6) still reduces to the same form as equation (7), except that m indexes pairs of latent variables and ϕ_m is a covariance (see Anderson & Böckenholt, 1999). When there are more that two observed variables, models with uncorrelated and correlated latent variables often imply different log-multiplicative models. The importance of this is that in the social sciences, latent variables are typically correlated.

The only structural restriction in the latent variable model used to derive equation (6) is that the discrete variables are conditionally independent given the continuous ones (and $\mathbf{\Sigma}(\mathbf{a}) = \mathbf{\Sigma}$). Without further restrictions, model (6) is too complex.

3 Restricted Latent Variable Models

Two types of more restrictive (homogeneous) models, which differ in terms of their complexity and the identification constraints needed to estimate their parameters, are presented. In Section 3.1, models are derived for cases where each observed variable is related to only one latent variable (i.e., single indicators), and in Section 3.2, we discuss more complex models where observed variables can be related to multiple latent variables (i.e., multiple indicators). The single indicator models are special cases of the multiple indicator models.

In Sections 3.1 and 3.2, the models are derived algebraically; however, we also present their corresponding graphical representations. In the graphs, the observed discrete variables are represented by squares and the continuous latent variables are represented by circles. Lines or edges that connect variables indicate that variables are related. The absence of a line between two variables indicates that the two variables are conditionally independent given all of the other variables. In Section 3.3, we outline how to read log-multiplicative models from graphs.

3.1 Single Indicators

In Section 3.1.1, log-multiplicative models are derived for cases of one, two and M (correlated) latent variables, and in Section 3.1.2, identification constraints are discussed .

3.1.1 Models with Single Indicators

The simplest structural model is a one common latent variable model. For $I = 4$, the graph of this model is given in Figure 1. Since $M = 1$,

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

Setting $\boldsymbol{\Sigma}(\mathbf{a}) = \boldsymbol{\Sigma}$ and replacing $g(\mathbf{a})$ and $\mathbf{h}(\mathbf{a})$ in equation (5) by their definitions in equations (2) and (8), the log-multiplicative model for the observed data is

$$\log(P(\mathbf{a})) = \lambda + \sum_i \lambda^*_{i(j_i)} + \sum_i \sum_{k>i} \sigma_{11} \nu_{i(j_i)1} \nu_{k(j_k)1}. \tag{9}$$

A slightly more complex structural model is a two latent variable model. Suppose that there are two, correlated latent variables, Θ_1 and Θ_2, and that the observed variables can be partitioned into two mutually exclusive sets $\mathcal{A}_1 = \{A_1, \ldots, A_r\}$ and $\mathcal{A}_2 = \{A_{r+1}, \ldots, A_I\}$ where the variables in $\mathcal{A}_1$ are indicators of Θ_1 and those in $\mathcal{A}_2$ are indicators of Θ_2. For $I = 4$, the graph for an example of this model is given in Figure 2. Given the absence of a direct relationship between Θ_2 and variables in $\mathcal{A}_1$ and between Θ_1 and $\mathcal{A}_2$, $\nu_{i(j_i)1} = 0$ for $A_i \in \mathcal{A}_2$, and $\nu_{i(j_i)2} = 0$ for $A_i \in \mathcal{A}_1$, which leads to

$$\mathbf{h}(\mathbf{a}) = (\sum_{i=1}^{r} \nu_{i(j_i)1}, \sum_{i=r+1}^{I} \nu_{i(j_i)2})'. \tag{10}$$

Setting $\mathbf{\Sigma}(\mathbf{a}) = \mathbf{\Sigma}$ and replacing $g(\mathbf{a})$ and $\mathbf{h}(\mathbf{a})$ in equation (5) by their definitions in equations (2) and (10), the model for the observed data is

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

In the most complex, single indicator model, each latent variable has only one indicator; therefore, $M = I$ and $\nu_{i(j_i)m} = 0$ for all j_i when $i \neq m$. For $I = 4$, the graph of this model is given in Figure 3. For this model,

$$\mathbf{h}(\mathbf{a}) = (\nu_{1(j_i)1}, \ldots, \nu_{I(j_I)I})'. \tag{12}$$

Setting $\mathbf{\Sigma}(\mathbf{a}) = \mathbf{\Sigma}$ and replacing $g(\mathbf{a})$ and $\mathbf{h}(\mathbf{a})$ in equation (5) by their definitions in equations (2) and (12), the model for the observed data is

$$\log(P(\mathbf{a})) = \lambda + \sum_{i} \lambda^*_{i(j_i)} + \sum_{i} \sum_{k>i} \sigma_{ik} \nu_{i(j_i)i} \nu_{k(j_k)k}. \tag{13}$$

Equation (13) is a multivariate generalization of the $RC(1)$ association model, which for three variables is equivalent to models discussed by Clogg (1982; Agresti, 1984). If category scores are known, then equation (13) 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 (13) 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 variables is observed, then the corresponding covariance can be set to zero. If partial associations between all pairs of variables are present, then restrictions on the latent variable model can be imposed that correspond to simpler graphs (e.g., Figures 1 and 2). In these simpler models, equality restrictions are imposed on the association parameters (i.e., variances and/or covariances) across multiplicative terms in the equations.

3.1.2 Identification Constraints

Identification constraints are required to estimate the parameters of 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 a zero (e.g., $\sum_{j_i} \nu_{i(j_i)m} = 0$). We use zero sum constraints in the examples presented in Section 5.

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 and in Section 3.2, it is advantageous to set the scale of the category scores for one variable that is an indicator of the latent variable. For example, if A_1 and Θ_m are related, then $\sum_{j_1} \nu^2_{1(j_1)m} = 1$. The rule adopted here 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 (11), the category scale values of one variable in $\mathcal{A}_1$ and one variable in $\mathcal{A}_2$ need to be scaled. For model (13), we take this rule to the limit and impose scaling constraints on the scale values for all the 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 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. 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 variable (i.e., the ω_{im}'s).

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

$$\log(P(\mathbf{a})) = \lambda + \lambda^*_{1(j_1)} + \lambda^*_{2(j_2)} + \lambda^*_{3(j_3)} + \sum_i \sum_{k>i} \sigma^*_{11}\nu_{i(j_i)1}\nu_{k(j_k)1}. \tag{14}$$

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 (14) is empirically indistinguishable from model (13), 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; Agresti, 1984).

For heterogeneous models, location constraints are required on the marginal effect terms and usually on the scale values. In many cases, the scaling rule will also apply; however, we cannot make global statements without specifying a model for $\mathbf{\Sigma}(\mathbf{a})$.

3.2 Multiple Indicators

Observed variables may be related to more than one latent variable. The major difficulty in using log-multiplicative models as multiple indicator latent variable 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 adopted in Section 3.1 (i.e., location constraints on the marginal effect terms and the scale values and a scaling constraint on the category scores of one variable per latent variable) are adopted here as well. The additional identification constraints (if any) required depend on the complexity of the model.

Let $\mathbf{N}_i$ equal the $(J_i \times M)$ matrix whose columns contain the scale values for the categories of variable A_i. If A_i is not an indicator of a particular latent variable, then the corresponding column of $\mathbf{N}_i$ contains zeros. 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 $\mathbf{N}_i \mathbf{\Sigma} \mathbf{N}_k'$ where $\mathbf{\Sigma}$ is the covariance matrix of the latent variables. For each possible observation, the interaction terms in the model equal the appropriate elements from the matrices in the set

$$\{\mathbf{N}_i \mathbf{\Sigma} \mathbf{N}_k' | i < k\}. \tag{15}$$

Determining the additional constraints needed to identify a model consists of determining whether transformations of the $\mathbf{N}_i$'s and $\mathbf{\Sigma}$ exist that have no effect on the value of the elements of the matrix products in (15). Since the number of possible multiple indicator models is far too large to consider here, we show how constraints required for three of the four models that are used in the examples presented in Section 5 are determined. The fourth model, which has a heterogeneous covariance matrix, is discussed in Section 5.2.2.

In the most complex latent variable model where each observed variable is related to all of the latent variables (i.e., the general model assumed in Section 2), none of the columns of the $\mathbf{N}_i$'s equals $\mathbf{0}$. Given any $(M \times M)$ non-singular matrix T, we can always set $\mathbf{N}_i^* = \mathbf{N}_i\mathbf{T}$ for all i and $\mathbf{\Sigma}^* = \mathbf{T}^{-1}\mathbf{\Sigma}\mathbf{T}^{-1}$ without changing the values of any of the elements of the matrix products in (15). Given this indeterminacy (and for convenience), we can arbitarily set all covariances equal to zero and estimate the M variances. Combining this choice with our other identification constraints pins down a unique solution.

If an observed variable is not an indicator of one 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. Let $\boldsymbol{\nu}_{im}$ equal the $(J_i \times 1)$ vector of scale values $\nu_{i(j_i)m}$. 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 cannot be arbitarily set equal to zero, because $\mathbf{N}_1\mathbf{\Sigma}\mathbf{N}_4' = (\sigma_{12}\boldsymbol{\nu}_{11}\boldsymbol{\nu}_{42})$. 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.

The third example consists of a model for four (or more) observed variables where the observed variables are all related to one common latent variable, and pairs of the discrete variables may also be related to uncorrelated (pair specific) latent variables. In addition to whatever constraints are needed for the common part of the model, the scale values for each discrete variable related to a pair specific latent variable must have a scaling condition imposed on them (just as the scale values for both variables in the RC association model for 2–way tables must have a scaling condition imposed on them).

3.3 Reading Models from Graphs

The way that log-multiplicative models are read 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 edges connecting the observed and latent variables have been labeled by the corresponding scale values. The interaction terms in the log-multiplicative models equal 1/2 times 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 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 indicator graph Figure 4. The log-multiplicative model for this graph is

$$
\begin{aligned}
\log(P(\mathbf{a})) \quad = \quad & \lambda + \lambda^*_{1(j_1)} + \lambda^*_{2(j_2)} + \lambda^*_{3(j_3)} + \lambda^*_{4(j_4)} \\
& + \sigma_{11}[\nu_{1(j_1)1}\nu_{2(j_2)1} + \nu_{1(j_1)1}\nu_{3(j_3)1} + \nu_{2(j_2)1}\nu_{3(j_3)1}] \\
& + \sigma_{22}[\nu_{2(j_2)2}\nu_{3(j_3)2} + \nu_{2(j_2)2}\nu_{4(j_4)2} + \nu_{3(j_3)2}\nu_{4(j_4)2}] \\
& + \sigma_{12}[\nu_{1(j_1)1}\nu_{2(j_2)2} + \nu_{1(j_1)1}\nu_{3(j_3)2} + \nu_{1(j_1)1}\nu_{4(j_4)2} + \nu_{2(j_2)1}\nu_{3(j_3)2} \\
& \quad + \nu_{2(j_2)1}\nu_{4(j_4)2} + \nu_{3(j_3)1}\nu_{2(j_2)2} + \nu_{3(j_3)1}\nu_{4(j_4)2}]
\end{aligned}
\qquad (16)
$$

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 $A_1 \rightarrow \Theta_1 \rightarrow A_1$. The covariance of a variable with itself is the variance, so we multiple $(1/2)\nu^2_{1(j_1)1}$ by the variance of Θ_1. Paths from a variable back to itself that involve a pair of latent variables are only found in multiple indicator models. For example, the directed paths $A_2 \rightarrow \Theta_1 \rightarrow \Theta_2 \rightarrow A_2$ and $A_2 \rightarrow \Theta_2 \rightarrow \Theta_1 \rightarrow A_2$ 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 5.2.2).

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 $A_1 \rightarrow \Theta_1 \rightarrow A_2$, and $A_2 \rightarrow \Theta_1 \rightarrow A_1$. The term $\sigma_{12}\nu_{1(j_1)1}\nu_{2(j_2)2}$ results from the directed paths $A_1 \rightarrow \Theta_1 \rightarrow \Theta_2 \rightarrow A_2$ and $A_2 \rightarrow \Theta_2 \rightarrow \Theta_1 \rightarrow A_1$.

Some heterogeneous models may include extra terms due to $|\mathbf{\Sigma}(\mathbf{a})|^{1/2}$ in equation (5). For homogenous models, $|\mathbf{\Sigma}(\mathbf{a})|^{1/2} = |\mathbf{\Sigma}|^{1/2}$, which is absorbed into the constant λ. In heterogeneous models, depending on how the covariance matrix differs over cells of the table, $|\mathbf{\Sigma}(\mathbf{a})|^{1/2}$ 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 $|\mathbf{\Sigma}(\mathbf{a})|^{1/2}$ is absorbed into the corresponding 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 $|\mathbf{\Sigma}(\mathbf{a})|^{1/2}$ for the single cell and another value of $|\mathbf{\Sigma}(\mathbf{a})|^{1/2}$ for the rest of the table. Only one element of $\mathbf{\Sigma}(\mathbf{a})$ 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).

4 Maximum Likelihood Estimation

The maximum likelihood estimation of the parameters of the log-multiplicative models p-resented in the previous sections is described here. The starting point is the most general, homogeneous latent variable model given in equation (6). The restricted 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 the some of the maximum likelihood equations will 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

$$\begin{aligned}
\frac{\partial \log L}{\partial \lambda^*_{i(j_i)}} &= \sum_{\mathbf{a}|j_i} \left[n(\mathbf{a}) - P(\mathbf{a})\right], \\
\frac{\partial \log L}{\partial \sigma_{mm}} &= \sum_{\mathbf{a}} \sum_{i} \sum_{k>i} \nu_{i(j_i)m} \nu_{k(j_k)m} \left[n(\mathbf{a}) - P(\mathbf{a})\right],
\end{aligned}$$

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

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 uni-dimensional Newton algorithm. This iterative method, which has become more or less the standard method for obtaining ML estimates of log-multiplicative models (see, for instance, Goodman, 1979; Clogg, 1982; Becker, 1989), involves updating one parameter at a time fixing the other parameters at their current value. A uni-dimensional 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 derivates are evaluated at the current values of all model parameters. The relevant second-order derivatives for the parameters appearing in equation (6) are

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

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 the previous section, 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 $\sum_{j_i}(\nu_{i(j_i)m})^2 = 1$).

Only a slight modification of the uni-dimensional 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

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

The provisional values for β_{im1} and β_{im2} can be obtained by regressing the term on the left hand side of equation (17) 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 multi-dimensional 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 log-multiplicative models (e.g., Anderson & Böckenholt, 1998; Clogg & Shihadeh, 1994; Eliason, 1995).

5 Examples

In Section 5.1, we present an example with multicategory items from the General Social Survey, and in Section 5.2, we fit models to Coleman's (1964) panel data for both boys and girls, including a heterogeneous covariance model.

5.1 Data from the 1994 General Social Survey

For this example, we analyze an $(2 \times 4 \times 5 \times 5)$ cross-classification of 899 responses from the 1994 General Social Survey (Davis & Smith, 1996) to the following four items:

A "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).

B "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).

C "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).

D "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).

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 2–way interaction log-linear models were fit. While the all 2–way model fits the data ($G^2 = 117.93$, $df = 136$, $p = .87$), 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, the one latent variable model, Model (c), is unsatisfactory. The two uncorrelated latent variable model where each item is an indicator of both latent variables, Model (d), 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 latent variable and additional 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 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[2], and it's 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 CD association. Table 2 contains the estimated association parameters and

[2]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.

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 C and B are most strongly related to the common latent variable, followed by items D and A. For this model, the conditional mean of the common latent variable equals $\hat{\sigma}_{11} = 11.294$ times the sum of the scale values corresponding to a given response pattern. The order of category scores for the common latent variable corresponds to the order of the response options, except for item D where the scale values for "strongly agree" and "agree" are nearly equal but out of order (i.e., $\hat{\nu}_{D(1)1} = -.135$ and $\hat{\nu}_{D(2)1} = -.145$). The greater the agreement with a statement, the greater the value on the mean of the latent variable.

Relative to the common latent variable, the variable for the extra CD association accounts for inconsistent extreme responses "strongly agree" to item C but "strongly disagree" to item D, 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 D. Item D immediately follows C, while A and B are from two different sections of the survey. Item D differs from the other items in that the traditional roles of husbands and wives are reversed and children are explicitly mentioned.

5.2 The Coleman Panel Data

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 items: their attitude toward (positive, negative) and their self-perception of membership in (yes, no) the leading or popular crowd. While the data for the boys have been analyzed extensively (e.g., Agresti, 1997; Andersen, 1988; Goodman, 1978; Langeheine, 1988; Whittaker, 1990), the data for the girls has not. After analyzing the boys and girls data separately by fitting models to each, and we fit models that include gender as a fifth variable.

5.2.1 Separate Analyses

The fit statistics for models fit separately to the boys and girls data are reported in Table 4. Starting with the boys data, we find that the independence log-linear model fails to fit ($G^2 = 421.68$, $df = 11$, $p < .001$), but the all 2–way interaction log-linear model provides a good fit for the boys ($G^2 = 1.21$, $df = 5$, $p = .94$). Given that the all 2–way model fits, we are justified in fitting homogeneous log-multiplicative models. The simplest model with one common latent variable fails to fit ($G^2 = 243.59$, $df = 7$, $p < .001$). We next fit a multiple indicator, two correlated latent variable model where 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., Figure 4 where B_1 and B_2 correspond to A_3 and A_4, respectively). For identification, the category scores for A_1 and B_2 are scaled and $\sigma_{11} = 1$. This model, Model (d) in Table 4, has the same fit and degrees of freedom as the all 2–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", (i.e., Figure 2 where B_1 and B_2 correspond to A_3

and A_4). The corresponding single indicator, two correlated latent variable model, Model (e), 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 i, yields $G^2 = 5.43$, $df = 8$, and $p = .71$. Lastly, to check whether $\sigma_{12} = 0$, we fit the uncorrelated latent variable version of Model (f); however this model, Model (g), fails to fit the data ($G^2 = 97.52$, $df = 9$, $p < .001$).

Our final model for the boys data, Model (f), is a linear by linear interaction model with restrictions across the association parameters, which can be fit using software that fits generalized linear models. The estimated variances (and standard errors from multi-dimensional Netwon-Raphson[3]) 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., "positive" or "yes") and $-.707$ for $j = 2$ (i.e., "negative" or "no").

For the girls data, we repeat the same analyses performed on the boys data. 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 which was the best one for the boys data, fail to fit the data[4]; however, the lack-of-fit appears to

[3]The estimated standard errors from multi-dimensional Newton-Raphson and from the jackknife of the uni-dimensional Newton procedure are equal to within $\pm.0001$.

[4]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,

be due to one cell. The response pattern A_1 = "positive", B_1 = "yes", A_2 = "negative", and B_2 = "no"' (i.e., the $(2,2,1,1)$ cell) has a relatively large residual.

For the girls, the covariance matrix for the $(2,2,1,1)$ cell may not equal that for all the other response patterns. If so, then as discussed in Section 3.3, we should add a single parameter, τ, to fit the cell perfectly. Re-fitting all the models adding the term $\tau\delta_{2211}$ (where $\delta_{2211} = 1$ for cell $(2,2,1,1)$ and 0 otherwise) greatly improves the fit of Models (d), (e) and (f) for the girls data[5]. 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; 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 $\mathbf{\Sigma}_{boys} = \mathbf{\Sigma}_{girls}$. The estimates of elements of $\mathbf{\Sigma}$ for the girls are slightly larger than those for the boys. The estimates (and standard errors) for the girls are $\hat{\sigma}_{11,girls} = .760(.040)$, $\hat{\sigma}_{22,girls} = 1.586(.052)$ and $\hat{\sigma}_{12,girls} = .138(.014)$.

5.2.2 Combined analysis

Given the results from fitting models separately to the boys and girls data, we expect that A_1 and A_2 are related to an unobserved attitude variable, B_1 and B_2 are related to an

−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.

[5] *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 again point to Model (f) as the best

unobserved membership variable, and scale restrictions can be imposed on the scale values for A_1, A_2, B_1 and B_2. Furthermore, we would like to permit the values on the unobserved variables to differ for boys and girls, and possibly $\mathbf{\Sigma}$ as well. This underlying model is represented by the graph in Figure 5.

To derive the most general log-multiplicative model for the graph in Figure 5, we define $g(\mathbf{a})$ as

$$g(\mathbf{a}) = \lambda + \lambda_{A_1(j)} + \lambda_{A_2(j)} + \lambda_{B_1(j)} + \lambda_{B_2(j)} + \lambda_{G(j)} \tag{17}$$

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

$$\mathbf{h}(\mathbf{a}) = \begin{pmatrix} \nu_{A_1(j)1} + \nu_{A_2(j)1} + \nu_{G(j)1} \\ \nu_{B_1(j)2} + \nu_{B_2(j)2} + \nu_{G(j)2} \end{pmatrix}. \tag{18}$$

The first row in equation (18) 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 define a heterogeneous covariance matrix

$$\mathbf{\Sigma}_{G(j)} = \begin{pmatrix} \sigma_{11G(j)} & \sigma_{12G(j)} \\ \sigma_{12G(j)} & \sigma_{22G(j)} \end{pmatrix} \tag{19}$$

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 (5) by the defintions in equations (17), (18) and (19), respectively, yields

$$\begin{aligned} \log(F(\mathbf{a})) &= \lambda + \lambda_{A_1(j)} + \lambda_{A_2(j)} + \lambda_{B_1(j)} + \lambda_{B_2(j)} + \lambda^*_{G(j)} \\ &\quad + \frac{1}{2}\sigma_{11G(j)}[\nu^2_{A_1(j)1} + \nu^2_{A_2(j)1}] + \frac{1}{2}\sigma_{22G(j)}[\nu^2_{B_1(j)2} + \nu^2_{B_2(j)2}] \end{aligned}$$

$$
\begin{aligned}
&+\sigma_{11G(j)}[\nu_{A_1(j)1}\nu_{A_2(j)1}+\nu_{A_1(j)1}\nu_{G(j)1}+\nu_{A_2(j)1}\nu_{G(j)1}] \qquad (20)\\
&+\sigma_{22G(j)}[\nu_{B_1(j)2}\nu_{B_2(j)2}+\nu_{B_1(j)2}\nu_{G(j)2}+\nu_{B_2(j)2}\nu_{G(j)2}]\\
&+\sigma_{12G(j)}[\nu_{A_1(j)1}\nu_{B_1(j)2}+\nu_{A_1(j)1}\nu_{B_2(j)2}+\nu_{A_2(j)1}\nu_{B_1(j)2}+\nu_{A_2(j)1}\nu_{B_2(j)2}\\
&\qquad +\nu_{A_1(j)1}\nu_{G(j)2}+\nu_{A_2(j)1}\nu_{G(j)2}+\nu_{B_1(j)2}\nu_{G(j)1}+\nu_{B_2(j)2}\nu_{G(j)1}]
\end{aligned}
$$

where $\lambda^*_{G(j)} = \lambda_{G(j)} + \log(|\boldsymbol{\Sigma}_{G(j)}|^{1/2}) + (1/2)\sum_{m=1}^{2}\sum_{m'=1}^{2}\sigma_{mm'G(j)}\nu_{G(j)m}\nu_{G(j)m'}$. While this log-multiplicative model appears quite complex, its interpretation is greatly facilitated by the graph in Figure 5. The model can be read from its graph using the method outlined in Section 3.3.

Given the results from the previous section, rather than estimating $\nu_{A_1(j)1}$, $\nu_{A_2(j)1}$, $\nu_{B_1(j)2}$ and $\nu_{2(j)2}$, we set them equal to $\pm.7071$. 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 (20).

The fit statistics for models with gender as a fifth observed variable are reported in Table 6. While the all 2–way interaction log-linear model is the baseline model for the homogeneous version of equation (20), the log-linear model with all 3–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 (20). Since the all 2–way interaction log-linear model, Model (a) in Table 6, fails to fit, the homogeneous model, Model (c), should also fail to fit. Not only does the homogeneous model fail to fit, but so does the model with an extra parameter to fit the $(2,2,1,1,)$ cell for the girls (i.e., $\tau\delta_{2211,G(j)}$ where $\delta_{2211,girls} = 1$ for the $(2,2,1,1)$ cell for the girls, and 0 otherwise).

Since the log-linear model with the 3–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 $\tau\delta_{2211,G(j)}$ is added to the model, the model fits the data ($G^2 = 19.47$, $df = 19.47$, $p = .30$). 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 model fit separately to the boys and girls data. Using the scale values and estimated covariance matrix we compute estimates of the mean values on the latent attitude and membership variables for cell means using equation (4). 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 (e.g., 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 that for the boys. The slopes of the lines for boys and girls differ, because $\hat{\mathbf{\Sigma}}_{boys} \neq \hat{\mathbf{\Sigma}}_{girls}$. If $\hat{\mathbf{\Sigma}}_{boys} = \hat{\mathbf{\Sigma}}_{girls}$, then the lines for boys and girls would be parallel and any difference between them would be due to 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/her attitude is toward the leading crowd (and visa versa).

6 Discussion

Log-multiplicative models provide a powerful and flexible approach to studying the relationships between nominal and/or ordinal variables in 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. Additional possibilities include adding individual level covariates to the model (Anderson & Böckenholt, 1999) and imposing inequality restrictions on the category scale values (Vermunt, 1998). Areas for future work include studying the relationship between the log-multiplicative models and item response theory models, and further exploration of the use of log-multiplicative models to estimate of individuals' values on latent variables.

References

Afifi, A.A. and Elashoff, R.M. (1969). Multivariate two sample tests with dichotomous and continuous variables. I. The location model. *The Annals of Mathematical Statistics, 40*, 290–298.

Agresti, A. (1984). *Analysis of Ordinal Categorical Data*, NY: Wiley.

Agresti, A. (1997). A model for repeated measurements of a multivariate binary response. *Journal of the American Statistical Association,* 92, 315–321.

Andersen, E.B. (1988), Comparison of latent structure models. In R. Langeheine and J. Rost (Eds) *Latent Trait and Latent Class Models* (pp 207–229), NY: Plenum Press.

Anderson, C.J. and Böckenholt, U. (1999). Graphical regression models for polytomous variables. manuscript under review.

Becker, M.P. (1989), Models for the analysis of association in multivariate contingency tables. *Journal of the American Statistical Association, 84*, 1014–1019.

Clogg, C.C. (1982). Some models for the analysis of association in multiway cross-classifications having ordered categories. *Journal of the American Statistical Association, 77*, 803–815.

Clogg, C.C., and Shihadeh, (1994). *Statistical Models for Ordinal Variable.* Thousand Oaks, CA: Sage.

Coleman, J.S. (1964). *Introduction to Mathematical Sociology.* Glencoe, IL: The Free Press.

Davis, J.A., and Smith,T.W. (1996). *General Social Surveys 1972–1996: Cumulative Codebook.* Chicago, IL: NORC.

Eliason, S.R. (1995). Modeling manifest and latent dimensions of association in two-way cross-classifications. *Sociological Methods and Research, 24*, 30–67.

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

Goodman, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association, 74*, 537–552.

Goodman, L.A. (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. *The Annals of Statistics,* 13, 10–69.

Krzanowski, W.J. (1980). Mixtures of continuous and categorical variables in discriminant analysis. *Biometrics, 36*, 493–499.

Krzanowski, W.J. (1983). Distance between populations using mixed continuous and categorical variables. *Biometrika, 70*, 235–243.

Krzanowski, W.J. (1988). *Principles of Multivariate Analysis.* NY: Oxford.

Langeheine, R. (1988). New developments in latent class theory. In R. Langeheine and J. Rost (Eds) *Latent Trait and Latent Class Models* (pp 77–108), NY: Plenum Press.

Lauritzen, S.L. (1996). *Graphical Models.* NY: Oxford Press.

Lauritzen, S.L., and Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *The Annals of Statistics,* 17, 31–57.

Olkin, I., and Tate, R.F. (1960). Multivariate correlation models with mixed discrete and continuous variables. *The Annals of Mathematical Statistics, 32*, 448–465.

Vermunt, J.K. (1998). *RC* association models with ordered row and column scores: estimation and testing. Paper presented at the 21st Biennial Conference of the Society for Multivariate Analysis in the Behavioral Sciences, Leuven, July 15-17, 1998.

Wermuth, N., and Lauritzen, S.L. (1990). Discussion of papers by Edwards, and Wermuth and Lauritzen. *Journal of the Royal Statistical Society, Series B,* 52, 51–72.

Whittaker, J. (1989). Discussion of paper by van der Heijden, de Falguerolles & de Leeuw, *Applied Statistics,* 38, 278–279.

Whittaker, J. (1990). *Graphical Models in Applied Mathematical Multivariate Statistics.* NY: Wiley.

Table 1: Fit statistics for models fit to four items from the 1994 General Social Survey on whether wives and/or husbands should work outside of the home.

Model		df	G^2	p	D	BIC
(a)	Independence	187	1063.25	$< .01$	.378	−209
(b)	All 2-way 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)	AB, AC, AD, BC, BD & CD	145	143.60	.52	.100	−843
(f)	AB, AC, BC, BD & CD	149	144.30	.59	.102	−869
(g)	AB, AC, BC & CD	155	161.36	.35	.113	−893
(h)	AC, BC & CD	158	164.98	.34	.124	−910
(i)	BC & CD	162	168.76	.34	.127	−933
(j)	CD	168	194.41	.08	.138	−948
(k)	BC	169	240.94	$< .01$	.168	−908
(l)	BD	169	260.16	$< .01$	.177	−899
(m)	AB	172	271.65	$< .01$	.183	−898
(n)	AC	171	275.56	$< .01$	.189	−887
(o)	AD	171	275.61	$< .01$	.189	−887

Table 2: Estimated parameters (and standard errors) from Model (j) in Table 1 fit to four items from the 1994 General Social (see text for the items).

		Response Options				
		strongly agree	agree (approve)[a]	neither agree nor disagree	disagree (disapprove)[a]	strongly disagree
$\hat{\sigma}_{11} = 11.294$	(4.252)					
$\hat{\omega}_{A1} = .077^b$	$\hat{\nu}_{A(j_A)1} =$		.055 (.016)		−.055 (.016)	
$\hat{\omega}_{B1} = .587$	$\hat{\nu}_{B(j_B)1} =$	−.306 (.094)	−.211 (.065)		.066 (.030)	.450 (.138)
$\hat{\omega}_{C1} = 1.00$	$\hat{\nu}_{C(j_C)1} =$	−.679 (.063)	−.305 (.074)	.083 (.050)	.324 (.024)	.577 (.051)
$\hat{\omega}_{D1} = .287$	$\hat{\nu}_{D(j_D)1} =$	−.135 (.048)	−.145 (.034)	.020 (.016)	.064 (.032)	.196 (.049)
$\hat{\sigma}_{22} = 2.642$	(1.094)					
	$\hat{\nu}_{C(j)2} =$	.783 (.137)	−.333 (.186)	−.052 (.163)	−.511 (.112)	.113 (.233)
	$\hat{\nu}_{D(j)2} =$	.108 (.164)	−.471 (.104)	.097 (.138)	−.467 (.119)	.734 (.122)

a. The response options for item A were "approve" and "dispprove".

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

Table 3: Coleman (1964) panel data where A_t and B_t refer to the attitude and membership items at time point t. Fitted values and standardized residuals are from Model (f) in Table 6 (graph in Figure 5 with heterogeneous variance and $\tau\delta_{2211,G(j)}$).

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

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

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

Table 4: Fit statistics for models fit separately to the boys and girls panel data.

Model		Boys Data df	Boys Data G^2	Boys Data p	Girls Data df	Girls Data G^2	Girls Data p	Girls Data with $\tau\delta_{2211}$ df	Girls Data with $\tau\delta_{2211}$ G^2	Girls Data with $\tau\delta_{2211}$ 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	.88	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	$< .01$

Table 5: Estimated parameters from the multiple indicator, two correlated latent variable model (Model (d) in Table 4) fit to the boys data. Note: $\hat{\sigma}_{11} = .520$, $\hat{\sigma}_{12} = .076$, and $\sigma_{22} = 1.00$.

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

Table 6: Fit statistics for models fit to Coleman 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 (Figure 5)					
(c)	Homogeneous $\mathbf{\Sigma}$	21	63.41	$< .001$	-121.47
(d)	Model (c) with $\tau\delta_{2211,G(j)}$	20	60.90	$< .001$	-115.17
(e)	Heterogeneous $\mathbf{\Sigma}_{G(j)}$	18	30.39	.03	-128.08
(f)	Model (e) with $\tau\delta_{2211,G(j)}$	17	19.47	.30	-130.19

Table 7: Estimated parameters from Model (f) in Table 6 fit to the Coleman (1964) panel data with gender as a variable. 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 .7071 for $j = 1$ and $-.7071$ for $j = 2$.

	Value(s)			Value(s)	
Parameter	$j = 1$	$j = 2$	Parameter	$j = 1$	$j = 2$
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	$-.060$
$\sigma_{11,boys}$	.578		$\sigma_{11,girls}$	.757	
$\sigma_{22,boys}$	1.228		$\sigma_{22,girls}$	1.583	
$\sigma_{12,boys}$	.123		$\sigma_{12,girls}$	.138	
			τ	.462	

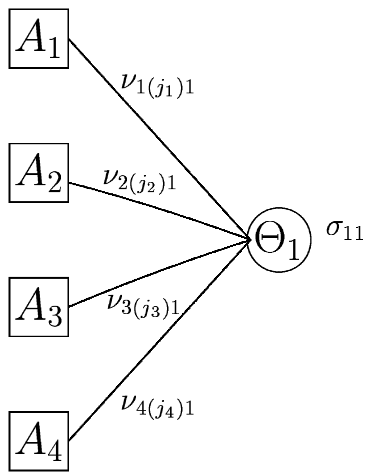

Figure 1: One common latent variable model for four observed variables.

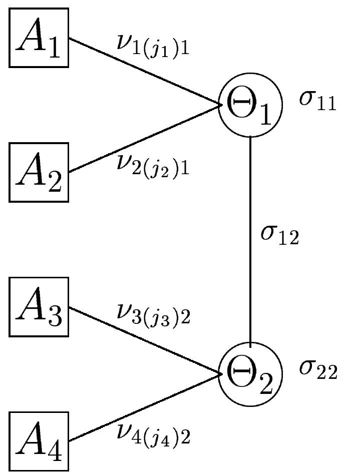

Figure 2: Single indicator, two common correlated latent variable model for four observed variables.

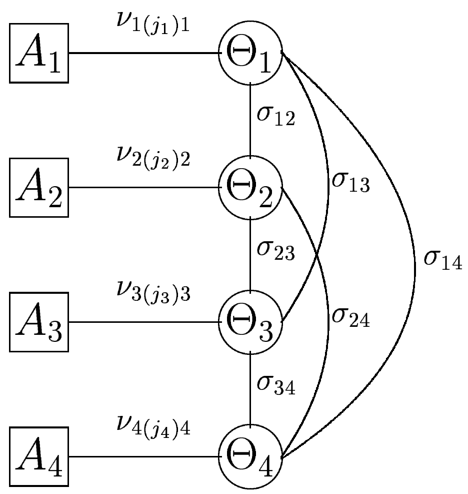

Figure 3: The most complex single indicator model for four variables.

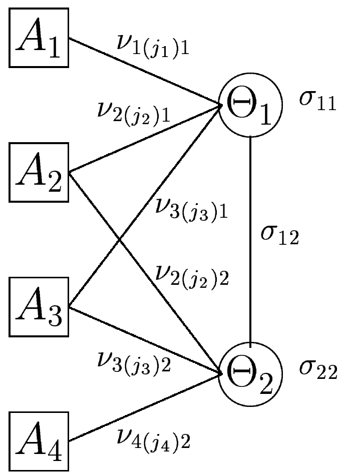

Figure 4: Multiple indicator, two common correlated latent variable model for four observed variables.

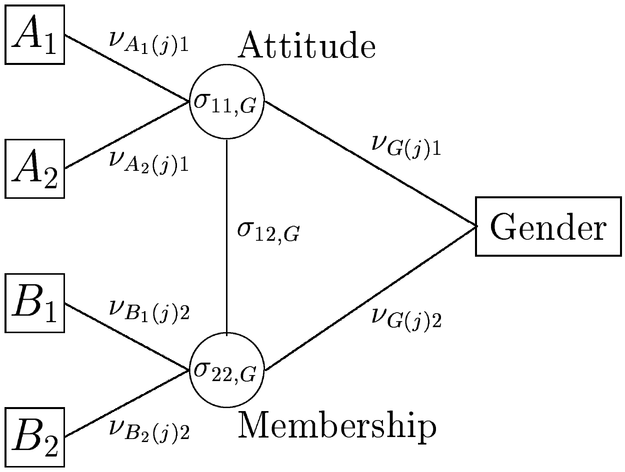

Figure 5: Graph corresponding to log-multiplicative Models (c)–(f) in Table 6 fit to the Coleman panel with gender as the fifth variable.

Figure 6: Plot of estimated attitude (left) and membership (right) means for boys (circles) and girls (dots) using scale values and estimated covariance matrix from Model (f) in Table 6 fit to the Coleman data.