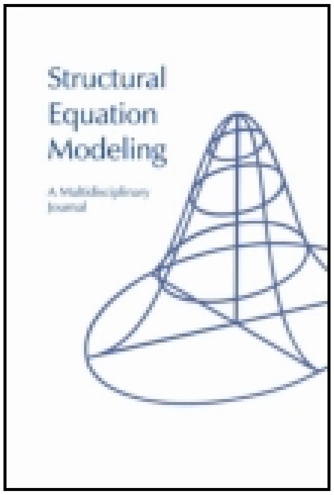

Structural Equation Modeling: A Multidisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hsem20

Residual Associations in Latent Class and Latent Transition Analysis

Tihomir Asparouhov[a] & Bengt Muthén[a]
[a] Muthén & Muthén, Los Angeles, CA
Published online: 08 Oct 2014.

Structural Equation Modeling: A Multidisciplinary Journal, 22: 169–177, 2015

ISSN: 1070-5511 print / 1532-8007 online
DOI: 10.1080/10705511.2014.935844

Residual Associations in Latent Class and Latent Transition Analysis

Tihomir Asparouhov and Bengt Muthén
Muthén & Muthén, Los Angeles, CA

This article explores a method for modeling associations among binary and ordered categorical variables. The method has the advantage that maximum-likelihood estimation can be used in multivariate models without numerical integration because the observed data log-likelihood has an explicit form. The association model is especially useful with mixture models to handle violations of the local independence assumption. Applications to latent class and latent transition analysis are presented.

Keywords: latent class analysis, latent transition analysis, residual associations, residual correlations, spurious classes

This article explores a method for modeling associations among binary and ordered categorical variables. The method has the advantage that maximum likelihood (ML) estimation can be used in multivariate models without numerical integration because the observed data log-likelihood has an explicit form. The association model is especially useful with mixture models to handle violations of the local independence assumption.

Typically in latent class analysis (LCA) all indicators are assumed to be independent conditional on the latent class. Muthén (1984) considered tetrachoric and polychoric correlations in multivariate modeling via the weighted least squares method, but this method cannot be used to estimate mixture models. Qu, Tan, and Kutner (1996) used continuous latent variables in mixture modeling to capture a residual correlation. This approach is generalized in Uebersax (1999). The problem with this approach is that it is not scalable to a large number of variables. In a model with a larger number of latent class indicators, it might be necessary to add multiple residual correlations, which will result in the addition of multiple latent variables. That in turn leads to high-dimensional numerical integration and a loss of precision even when the numerical integration is done with a Monte Carlo integration method.

The main purpose of the association parameters in this article is to be able to add additional correlations between indicators beyond what the main LCA model explains and to prevent such minor residual correlations from interfering with the main modeling focus regarding latent classes. This is a scalable approach that can easily accommodate any number of association parameters; that is, the association parameters do not introduce additional computational burden. We also make the point that the association parameters can be interpreted the same way that residual correlation parameters are interpreted and thus make this new parameterizations easily understandable.

The next section presents the proposed approach. We then consider the relationship between the uniform association parameter and the polychoric correlation parameter. The following section describes an LCA with uniform associations simulation study. We then describe a latent transition model with uniform associations and conduct a simulation study. Different methods for discovering residual associations among latent class indicators are described next. We then illustrate the new methodology with a real data example before the conclusion. All analyses are carried out using M*plus* Version 7.2. Scripts are available at www.statmodel.com.

THE RESIDUAL ASSOCIATION APPROACH

The residual association model proposed here is the uniform association model defined in Goodman (1979). This model is a restricted log-linear contingency table model. The model generalizes naturally to mixture modeling with within-class

Correspondence should be addressed to Tihomir Asparouhov, Muthén & Muthén, 3463 Stoner Avenue, Los Angeles, CA 90066. E-mail: tihomir@statmodel.com

association. As a contrast, consider first the standard log-linear model. Suppose that there are two ordered categorical variables U_1 and U_2 with observed categories 1,...,l_1 and 1,...,l_2. The standard log-linear model is given by

$$P(U_1 = a_1, U_2 = a_2) = \frac{Exp\left(\tau_{1,a_1} + \tau_{2,a_2} + \beta_{a_1 a_2}\right)}{\sum_{i,j} Exp(\tau_i + \tau_j + \beta_{ij})} \quad (1)$$

where $\tau_{1,l_1} = \tau_{2,l_2} = \beta_{l_1 j} = \beta_{il_2} = 0$ for identification purposes. In the LCA context, log-linear modeling of residual association has been considered in Hagenaars (1988) and Magidson and Vermunt (2004). For the importance of considering residual associations, see also Berzofsky, Biemer, and Kalsbeek (2014) and Van Smeden, Naaktgeboren, Reitsma, Moons, and de Groot (2013).

The log-linear model is a fully saturated model when considering two variables and the number of free parameters is $l_1 l_2 - 1$. The parameters β_{ij} represent the deviation from independence of the two variables and there are $(l_1 - 1)$ $(l_2 - 1)$ such parameters. The interpretation of these parameters is, however, not as natural as the polychoric correlation parameter. In addition, the power to discover nonindependence will be lower for the log-linear model simply because the polychoric correlation model is more parsimonious. Suppose that we are using the likelihood ratio test to test the independence model. If both the polychoric and the log-linear models hold, the test statistic value will be the same and the degrees of freedom will be larger for the log-linear model; that is, $(l_1 - 1)$ $(l_2 - 1)$ degrees of freedom for the log-linear model versus 1 *df* for the polychoric correlation model. Thus the power to reject the independence model will be lower for the log-linear model.

The uniform association model resolves the preceding problems by restricting the log-linear model to

$$\beta_{ij} = \beta \times i \times j. \quad (2)$$

Thus the uniform association model for two variables is given by

$$P(U_1 = a_1, U_2 = a_2) = \frac{Exp\left(\tau_{1,a_1} + \tau_{2,a_2} + \beta a_1 a_2\right)}{\sum_{i,j} Exp(\tau_i + \tau_j + \beta ij)}.$$

Note now that the nonindependence of U_1 and U_2 is modeled entirely by a single coefficient β. We call this coefficient the association of U_1 and U_2. If this coefficient is 0 the variables are independent and thus the association coefficient is similar that way to the polychoric correlation. Becker (1989) showed that under certain conditions the association coefficient is approximately $\rho/(1 - \rho^2)$ where ρ is the polychoric correlation. This approximate relationship can be reversed and using β we can approximate the residual correlation as $(\sqrt{1 + 4\beta^2} - 1)/(2\beta)$. Note also that we still have the identifying constraints $\tau_{1,l_1} = \tau_{2,l_2} = 0$, and the constraints $\beta_{l1j} = \beta_{il_2} = 0$ are now replaced by Equation 2. If the variables U_1 and U_2 are both binary, then the uniform association model is equivalent to the log-linear model as they are both saturated.

Another advantage of the uniform association model over the log-linear model is the fact that this model uses the ordered nature of the variables. The log odds for U_1 over two consecutive categories, conditional on U_2 is a linear function of U_2. Thus higher values of U_2 are associated with higher values of U_1 when the association parameter is positive.

An advantage of the uniform association model over the polychoric correlation model is missing data modeling. The weighted least squares estimation of the polychoric correlation model does not support missing at random (MAR) missing data and it generally gives unbiased estimates only when the missing data is missing completely at random (MCAR). The ML estimation of the association model guarantees unbiased estimates even when the missing data is MAR.

It is easy to extend the association model to a multivariate model with more than two variables. Suppose that U_1,...,U_r are ordered categorical variables and the observed categories for U_i are 1,...,l_i. The uniform association model is given by

$$P(U_1 = a_1, U_2 = a_2, \ldots, U_r = a_r) = \frac{Exp\left(\sum_i \tau_{i,a_i} + \sum_{i<j} \beta_{ij} a_i a_j\right)}{\sum_{a_1,a_2,\ldots,a_r} Exp\left(\sum_i \tau_{i,a_i} + \sum_{i<j} \beta_{ij} a_i a_j\right)} \quad (3)$$

where β_{ij} represents the association coefficient between U_i and U_j and can be thought of as the association equivalent of the polychoric correlation between U_i and U_j. Not all of these associations need to be present in the model. Some of the association parameters can be zero. Note here that if the variables U_1,...,U_r can be split in two groups with no association between them, the two groups will be independent. For example, if the two groups are U_1,..., U_k and U_{k+1},..., U_k and all the associations $\beta_{ij} = 0$ when $i \le k$ and $j > k$, then

$$P(U_1 = a_1, U_2 = a_2, \ldots, U_r = a_r) =$$

$$P\,(U_1 = a_1, \ldots, U_k = a_k)\, P(U_{k+1} = a_{k+1}, \ldots, U_r = a_r).$$

This property of the multivariate association model usually allows us to reduce the multivariate model to small groups of independent models, which improves computational efficiency.

The uniform association model naturally extends to mixture modeling and in particular to LCA and latent transition analysis (LTA) models with residual associations. If C represents a latent class variable measured by the observed variables U_i the LCA association model is given by

$$P(U_1 = a_1, U_2 = a_2, \ldots, U_r = a_r|C) = \frac{Exp\left(\sum_i \tau_{i,a_i,c} + \sum_{i<j} \beta_{ij,c} a_i a_j\right)}{\sum_{a_1,a_2,\ldots,a_r} Exp\left(\sum_i \tau_{i,a_i,c} + \sum_{i<j} \beta_{ij,c} a_i a_j\right)} \quad (4)$$

where $\beta_{ij,c}$ are class-specific residual associations. One can selectively add residual associations to the LCA model if they appear to be significant. This approach can be particularly useful in LCA where certain latent class indicators might have higher associations or correlations than explained by the latent class variable of the LCA model. Such residual associations if left out of the model will most likely lead to spurious class formations (see Asparouhov & Muthén, 2011). Thus in a practical application where standard class enumeration criteria such as Bayesian information criterion lead to many more classes than the analyst can interpret, the LCA association model can be used to eliminate spurious class formations that are due to residual indicator associations.

The τ parameters in Equation 4 are not the same τ parameters that usually come from a probit or logistic link function. Those τ parameters will be different and they do depend on the association parameters. If the association parameters are all 0 then the τ parameters will be the same as if the latent class indicators are nominal indicators. The best way to understand the impact of the τ parameters is to look at the class-specific marginal estimated indicator distributions on the probability scale.

The Appendix gives an outline of the ML estimation of the mixture model with the uniform residual associations. In the case when all the indicator variables are binary, the association modeling is equivalent to the local dependence LCA model discussed in Hagenaars (1988) and Magidson and Vermunt (2004). For binary indicator variables the LCA model with residual tetrachoric correlations can also be estimated with Bayesian methods (see Asparouhov & Muthén, 2011).

THE CONNECTION BETWEEN THE ASSOCIATION PARAMETER AND THE POLYCHORIC CORRELATION PARAMETER

This section illustrates the connection between the association parameter and the polychoric correlation parameter using simulated data. We generate ordered categorical data with five categories using a bivariate probit model where the thresholds for both variables are –1.5, –0.5, 0.5, and 1.5, respectively. We vary the polychoric correlation and compare the estimates of the polychoric correlation and the association parameter. We generate a large sample of size $N = 10^5$ so that variation across samples is eliminated and the asymptotic estimates are obtained. The results of this simulation are presented in Table 1. The results indicate that the connection between the polychoric correlation and the uniform association is very strong and larger values of the correlation are equivalent to larger values of the association parameter. The relationship is not one to one and the association parameter is not restricted to be less than 1. The Becker's approximation is quite good for smaller values but it appears to be underestimating the polychoric parameters when the values are large. It is important to note here that this evaluation is simply an example. The connection between polychoric correlation and the uniform association in general depends on the number of categories as well as the threshold values and in other examples might not be similar to the results in Table 1. We can, however, always expect that the general pattern will be preserved. A simulation study with negative polychoric correlations looks identical to the results in Table 1 with all values having a negative sign.

TABLE 1
Comparing Polychoric Correlation and Uniform Association

Polychoric Correlation	*Uniform Association*	*Becker's Approximation*
ρ	β	$(\sqrt{1+4\beta^2}-1)/(2\beta)$
0.10	0.09	0.09
0.30	0.29	0.27
0.50	0.56	0.45
0.70	1.05	0.63
0.90	2.46	0.82

Next we conduct a simulation study using the same model generation but now we generate 100 samples of size $N =$ 1,000. We compute the average Pearson chi-square statistic over these 100 samples to evaluate the ability of the model to fit the bivariate distribution. With 24 *df* in the data and 9 estimated parameters (1 association parameter and 2 × 4 univariate distribution parameters) the Pearson chi-square has 15 *df*. Average test values near 15 or lower mean that the bivariate distribution table was fitted well. We use three bivariate models. The first one is the polychoric correlation model, which is identical to the model used to generate the data. The second model is the uniform association model. The third model is the Qu et al. (1996) model (QTK), which uses a logit link function and a normally distributed latent variable within a logistic regression to model the correlation between the observed variables. The QTK method in this situation is simply the Samejima (1969) graded response model.

Table 2 contains the average Pearson chi-square statistic for the three methods and varying polychoric correlation. It is clear that all three methods fit the data well. Thus we conclude that the uniform association method, which is the only scalable method, works just as well as other standard methods for fitting bivariate distributions and residual covariation between ordered categorical variables. Note that in Table 2 as the polychoric correlation increases the Pearson test statistic values decrease on average. This is explained by the fact that when the correlation between the two variables is bigger,

TABLE 2
Comparing the Average Pearson Chi-Square Values for the Polychoric Correlation Model, the Uniform Association, and the QTK Model

ρ	*Polychoric Correlation*	*Uniform Association*	*QTK*
0.10	15.3	15.3	15.4
0.30	15.0	15.0	15.5
0.50	13.8	14.1	14.6
0.70	13.2	13.9	13.4
0.90	9.9	8.5	11.6

Note. QTK = Qu, Tan, & Kutner (1996) model.

some cells of the bivariate distribution will be empty and that decreases the degrees of freedom and in turn affects the test statistic distribution.

LCA SIMULATION STUDIES

In this section we present some LCA simulation studies to evaluate the performance of the ML estimation when estimating the LCA model with residual associations. In the first simulation we consider a model where the associations are held equal across class and in the second simulation we consider an example where the associations are class specific. We consider an LCA model with two equal-sized classes, 10 indicator variables, and three categories. The threshold parameters given in Equation 4 are as follows: $\tau_{i,1,1} = -1$, $\tau_{i,2,1} = 0$, $\tau_{i,1,2} = 0$, $\tau_{i,2,2} = 1$. We also introduce for our first simulation the association parameters $\beta_{1,2,c} = \beta_{1,6,c} = \beta_{2,7,c} = \beta_{3,8,c} = \beta_{4,9,c} = \beta_{5,10,c} = 0.3$ for both classes $c = 1$ and $c = 2$. We generate 100 data sets of sample size 2,000 using the LCA association model and we analyze the data using the same model holding the association parameters equal across classes. We add to the LCA model only the six nonzero associations listed earlier.

In the second simulation we use $\beta_{1,2,2} = \beta_{1,6,1} = \beta_{2,7,1} = \beta_{3,8,1} = \beta_{4,9,1} = \beta_{5,10,1} = 0.3$; that is, the associations are not the same between the two classes. Class-specific associations are created again both for the generation of the data and for the estimation. Because the estimated model and the generating model are identical in the two simulations, we expect to see unbiased estimates and 95% coverage. We also introduce MAR missing data in the simulation studies. The probability that U_i is observed is $1/(1 + Exp(-1)) = 0.73$ for $i = 3, ..., 10$, U_2 is always observed, and the probability of U_1 to be observed is $1/(1+Exp(-1-U_2))$. This method of generating missing data yields MAR missing data, rather than simply MCAR, because the probability that U_1 is missing depends on U_2.

The results of the simulation are presented in Table 3. In both simulations, class-invariant and class-specific uniform association parameters are estimated well. The bias is negligible and the coverage is near the nominal levels of 95%.

TABLE 3
Absolute Bias and Coverage for the Uniform Association Parameters in Latent Class Analysis

Parameter	*Simulation 1*	*Simulation 2*
$\beta_{1,2}$	.01 (.97)	.01 (.96)
$\beta_{1,6}$	.01 (.94)	.01 (.91)
$\beta_{2,7}$	.01 (.96)	.02 (.94)
$\beta_{3,8}$	.01 (.95)	.01 (.94)
$\beta_{4,9}$	.00 (.96)	.01 (.95)
$\beta_{5,10}$	.01 (.96)	.01 (.96)

LTA SIMULATION STUDIES

In LTA, typically the same instrument is used to measure a latent class over several time points. The goal of LTA is to evaluate how the latent class changes over time. When the same item is administered over time to the same individual it is common to observe residual correlation that goes beyond what the item is supposed to measure. This is due to personal perceptions to particular questions, and personal biases and interpretation of particular items. In this section we explore the consequences of ignoring these item-specific residual correlations.

We generate 100 samples of size $N = 1{,}000$ using an LTA model with two time points. At each time point we have two latent classes. The latent class variable is measured by five binary indicators. We denote the latent class variable at Time points 1 and 2 by C_1 and C_2. The bivariate distribution for C_1 and C_2 is generated using the following parameters:

$$P_{11} = P(C_1 = 1, C_2 = 1) = 0.31$$

$$P_{21} = P(C_1 = 1, C_2 = 2) = 0.19$$

$$P_{12} = P(C_1 = 2, C_2 = 1) = 0.25$$

$$P_{22} = P(C_1 = 2, C_2 = 2) = 0.25$$

Once the latent class variables are generated, we use the following LTA model with uniform associations to generate the indicator variables. Denote the indicator variables i at time point j by U_{ij}.

$$\begin{aligned} &P(U_{11} = a_{11}, \ldots, U_{52} = a_{52} | C_1, C_2) = \\ &\prod_i P(U_{i1} = a_{i1}, U_{i2} = a_{i2} | C_1, C_2) = \\ &\prod_i \frac{Exp(\tau_{i,a_{i1},c_1} + \tau_{i,a_{i2},c_2} + \beta_i a_{i1} a_{i2})}{\sum_{a_{i1},a_{i2}} Exp(\tau_{i,a_{i1},c_1} + \tau_{i,a_{i2},c_2} + \beta_i a_{i1} a_{i2})} = \\ &\frac{Exp(\sum_j \sum_i \tau_{i,a_{ij},cj} + \sum_i \beta_i a_{i1} a_{i2})}{\sum_{a_{11},\ldots,a_{52}} Exp(\sum_j \sum_i \tau_{i,a_{ij},cj} + \sum_i \beta_i a_{i1} a_{i2})}. \end{aligned} \quad (5)$$

TABLE 4
Average Estimate and Coverage for the Class Allocation and Transition Probabilities in Latent Transition Analysis (LTA)

Parameter	*True Value*	*LTA With Association*	*LTA Without Association*
p_{11}	.31	.31 (.97)	.34 (.90)
p_{21}	.19	.19 (.96)	.16 (.56)
p_{12}	.25	.25 (.96)	.21 (.52)
p_{22}	.25	.25 (.95)	.29 (.71)
q_1	.62	.62 (.94)	.68 (.63)
q_2	.50	.50 (.95)	.58 (.51)

In this uniform association model, the marginal distribution of $[U_{ij}|C_j]$ is the same across time points. The threshold parameters are time invariant. The uniform association β_i is the residual association between the same indicator at the two time points U_{i1} and U_{i2}. We generate the data according to the preceding model using the following parameter values $\tau_{i,1,1} = 1$ and $\tau_{i,1,2} = -1$ and $\beta_i = 0.3$. We estimate two LTA models, both holding the conditional distribution $[U_{ij}|C_j]$ invariant over time; that is, by holding the τ parameters equal across time. The first LTA model includes the uniform association and the second model does not.

We report the estimated class allocation probabilities p_{ij} in Table 4. We also report the results for the transition probability q_i where

$$q_i = P(C_2 = i|C_1 = i).$$

The results clearly show that if the residual associations are not accounted for, the LTA results are biased. The LTA with the uniform association yields unbiased estimates and good coverage. The standard LTA without the uniform associations underestimated the number of individuals that change latent class. The number of individuals that remained in the same class was overestimated by 7% on average. The coverage for the class allocation parameters for the standard LTA model is also quite poor. This result is natural and expected. When the residual similarities between the indicators are not accounted for, the standard LTA will attempt to explain it by additional correlation between the latent class variables.

METHODS TO DISCOVER RESIDUAL CORRELATIONS

Two methods are discussed in this section that can be used to find unaccounted residual correlation in LCA. The first method is based on finding misfit in the bivariate distribution via the bivariate Pearson test statistic. The second method is based on directly estimating an LCA model with a large number of uniform associations, possibly all associations. Both methods have advantages and disadvantages that we briefly discuss. The two methods also have different statistical power to discover residual associations. The formal evaluation of the power is beyond the scope of this article.

To illustrate the two methods we use a generated data set of size $N = 5{,}000$ using a two-class LCA model with residual associations. We generate the data as in Simulation 1 earlier with the exception that no missing data are generated; that is, there are 10 latent class indicator variables and there are no missing data in any of them. We generate a single data set and we apply the two methods to determine which association should be added to the standard LCA model.

Bivariate Pearson Testing

Consider first the bivariate Pearson method. For each pair of indicator variables U_i and U_j we compute the Pearson test statistic

$$T_{ij} = \sum_{a_1,a_2} \frac{(E_{ija_1a_2} - O_{ija_1a_2})^2}{E_{ija_1a_2}}$$

where $E_{ija_1a_2}$ is the model-estimated number of observations for which $U_i = a_1$ and $U_j = a_2$ and $O_{ija_1a_2}$ is the corresponding observed quantity.[1] This statistic is not a chi-square statistic because the estimated and the observed quantities are a part of a bigger model and thus the testing setup is formally not the same as the standard Pearson test, which evaluates the entire contingency table rather than just a bivariate table. However, the test statistic is still a good indicator for a residual association as this example illustrates. As an approximate degrees of freedom value for this test statistic, we would suggest $l_il_j - l_i - l_j + 1$, because there are $l_il_j - 1$ degrees of freedom in the bivariate contingency table and there are $l_i - 1 + l_j - 1$ univariate distribution parameters. In our case because $l_i = l_j = 3$, the approximate degrees of freedom for this test statistic would be 4 with an upper 5% quintile of 9.45; that is, any value above 9.45 can be considered as an indication of a possible residual association. Note, however, that this approach needs to account for multiple testing; that is, we can expect that just by chance at least one in 20 of the test statistics will be above that cutoff value and thus a higher cutoff value is a better choice. In a practical situation, it is best if the Pearson test statistics are ordered in descending order and only the top few are considered. That way the uncertainty of the distribution of the Pearson test statistic will be avoided. Oberski, Kollenburg, and Vermunt (2013) used bootstrap methods to determine the distribution of the bivariate Pearson test statistic.

Another problem with the Pearson test statistic is that it is not a reliable source of information in the presence of missing data. This is because we are comparing the observed univariate and bivariate values to the model estimated values. In the presence of missing data, the observed values

[1]This test is obtained using the M*plus* TECH10 option.

TABLE 5
Pearson Statistic for Bivariate Fit in Latent Class Analysis With Class-Invariant Associations

Parameter	*Pearson*
$\beta_{3,8}$	72
$\beta_{4,9}$	60
$\beta_{2,7}$	48
$\beta_{5,10}$	38
$\beta_{1,2}$	28
$\beta_{1,6}$	20
$\beta_{4,6}$	12
$\beta_{5,7}$	11

are inferior to the estimated values because they are not based on the full information contained in the entire data as are the estimated values. Thus misfit in the bivariate Pearson chi-square statistic might be due to MAR missing data rather than an omitted residual association; that is, the value of the statistic might be large because the observed values have selection bias.

The bivariate Pearson method is very easy to use and it only requires the estimation of the standard LCA model. For our generated data set, Table 5 shows all association parameters with Pearson statistic above 10 in descending order. The top six most significant associations come out to be exactly the true associations used to generate the data. If we examine the Pearson test statistic after we include those six associations in the LCA model, the largest Pearson test statistic value is now 11, and we can conclude that the added association parameters have resolved the bivariate misfit.

Yet another drawback of this method is that it does not distinguish between a class-specific association and a class-invariant association. The Pearson statistic is a measure of bivariate fit for all the classes together. If an association is positive in one class and negative in another, it is unlikely that the Pearson statistic will detect that association at all. If, however, an association is positive in one class and zero in another, the Pearson statistic can detect such an association although with diminished power. A data set generated as in Simulation 2 earlier where the associations are class specific yields smaller but similar Pearson test values. Table 6 contains the test values bigger than 10 in that case. Four out of the six associations were detected. Presumably lack of power made the other two associations go undetected. In both cases, the Pearson test statistic did not erroneously suggest any associations that are not in the data.

Including All Uniform Associations

The second method we discuss here is based on directly estimating the LCA model with all uniform associations included in the model. This method is feasible and can directly detect significant associations. However, this method has drawbacks as well. Including all association parameters

TABLE 6
Pearson Statistic for Bivariate Fit in Latent Class Analysis With Class-Specific Associations

Parameter	*Pearson*
$\beta_{1,6}$	27
$\beta_{1,2}$	22
$\beta_{4,9}$	20
$\beta_{3,8}$	14

TABLE 7
Associations *T* Statistic for Latent Class Analysis With All Class-Invariant Associations

Parameter	*Association*	*T Value*
$\beta_{3,8}$	0.39	8.7
$\beta_{2,7}$	0.39	6.7
$\beta_{5,10}$	0.28	5.3
$\beta_{4,9}$	0.32	5.0
$\beta_{1,2}$	0.35	3.9

in the model reduces the power to detect significance. Another drawback is that if there are a large number of class indicators, the computation can become slow, in particular when there are missing data. In our example of 10 indicators with no missing data, the estimation with all 45 class-invariant associations included in the model took 6 min to estimate. For comparison purposes, the LCA analysis with just the true six associations takes 1 sec to estimate. Thus, this is a more computationally intensive method than the Pearson statistic. Unlike the Pearson statistic, the LCA model with all associations included yields reliable results even in the presence of missing data.

Using our generated data set we estimate the LCA with all class-invariant associations included and report in Table 7 all associations with T-statistic values above 2 in descending order. Here five out of the six true associations were detected and again no spurious associations were detected.

In principle the LCA model can be estimated with all class-specific associations or with all class-invariant associations. The model with all class-specific associations is identified in principle. When the ordered variable has 10 categories or more the estimation of the model becomes very similar to a latent profile analysis where all variables are treated as normally distributed variables. It is well known that all correlations can be included as class-specific correlations in a latent profile analysis. Also, it was pointed out in Asparouhov and Muthén (2011) that in the case where all variables are binary the model with all class-specific correlations can be estimated with the Bayes estimator. However, unless the sample size is very large and the number of indicators is small, including all class-specific associations will yield a model with many local solutions that will most likely hinder this method's usability. In our generated example, after including all class-specific associations, the LCA model did not have two equal-sized classes even when using good starting values. This means that the LCA with the all class-specific uniform association is so different from the original LCA model that the significance of association cannot be trusted to apply in the original LCA model. Thus we can recommend using the second method with all class-invariant associations only.

These association detection tools can be thought of as data mining tools. Ultimately whether an association parameter is included in the LCA model should be decided by the LRT based on the model with and without that association, by the BIC criterion, or by the T test when the association is included in the LCA model. To test if a particular association is class-specific or class-invariant, one can use the LRT test or the T test for the difference between the class-specific associations, or in the case of more than two classes, the Wald test can be used to test simultaneously the equality across all classes. Additional association detection tools and power analysis are discussed for the case of binary items in Oberski et al. (2013).

REAL DATA ILLUSTRATION

In this section we use a real data example to illustrate the advantages of the LCA model with uniform associations. The data we consider consist of 17 antisocial behavior items obtained from the National Longitudinal Survey of Youth (NLSY). A sample of $n = 7{,}326$ subjects ages 16 to 23 is used. The items concern the frequency of various behaviors during the past year. For the present purpose, these items are dichotomized and scored as 0 or 1, with 0 representing never in the last year. The items are damaged property, fighting, shoplifting, stole less than \$50, stole more than \$50, use of force, seriously threaten, intent to injure, use marijuana, use other drugs, sold marijuana, sold hard drugs, "con" someone, take auto, broken into building, held stolen goods, gambling operation. We consider a latent class analysis for the 17 antisocial behavior items.

Table 8 contains the BIC for the LCA model with three, four, five, and six latent classes. As is often the case, BIC does not show a decrease followed by an increase as is needed for using the minimum BIC as a guide to the number of classes. The five-class solution has a clear substantive interpretation, whereas the six-class solution merely has two slight variations on one of the classes in the five-class solution. For each of these LCA models, we also count the number of pairs with Pearson test statistic > 30. Such bivariate test values can be considered severe violations of model fit. The number of such a degree of misfit stabilizes at five classes. Instead of adding additional classes so that these residual associations are accounted for, one can simply add a few residual associations, thereby keeping the number of classes as low as possible. Instead of 18 extra parameters when adding a class, a few residual association parameters can be added. This enhances the chances of finding a solution with a best minimum BIC.

TABLE 8
Latent Class Analysis for the Antisocial Behavior data

Model	*Log-Likelihood*	*BIC*	*Pearson > 30*	*No. of Parameters*
three-class LCA	−41713	83898	15	53
four-class LCA	−41007	82647	8	71
five-class LCA	−40808	82409	5	89
six-class LCA	−40658	82269	4	107
five-class LCA-UA	−40498	81842	0	95

Note. BIC = Bayesian information criterion; LCA = latent class analysis; LCA–UA = latent class analysis model with uniform associations.

TABLE 9
Bivariate Associations in 5-Latent Class Analysis for the Antisocial Behavior Data

Item	*Item*	*Pearson*
THREAT	INJURE	87.462
DRUG	SOLDPOT	60.218
SOLDPOT	SOLDDRUG	57.812
DRUG	SOLDDRUG	37.042
FIGHT	THREAT	31.398

Using the five-class solution we explore adding association parameters instead of more classes. Table 9 contains the five pairs of items with Pearson statistic > 30 in the five-class LCA. We notice that a group of three items (DRUG, SOLDPOT, SOLDDRUG) all have a residual association among each other. The remaining two residual associations also involve the same item THREAT. To form a complete group we also consider the additional association between FIGHT and INJURE and we form another block of three items (FIGHT,INJURE,THREAT) with all residual associations in the group. In total we add six residual associations. The results for the estimated five-class LCA model with uniform associations (LCA–UA) are also reported in Table 8. The BIC shows that the the five-class LCA–UA provides the best fit to the data among the models we considered. The likelihood improvement due to the six association parameters is much greater than that of the added sixth class. The Pearson test statistic in the LCA–UA model shows that no pairs of variables display a severe bivariate misfit; that is, all test statistics are smaller than 30. The uniform association modeling approach avoids adding spurious classes in LCA to account for violations of the local independence assumption.

Further exploration that goes beyond the purpose of this illustration can illuminate the data analysis and the measurement instrument itself. Analysis should be conducted for each association to see if the association is statistically significant in all classes, and if the association is the same in all classes if it is significant. In classes where an association is not statistically significant, the association parameter can be fixed to 0 and in the remaining classes the association parameter can be class specific if the differences between the association parameters are statistically significant. If certain items are highly correlated in all classes, one can go further and question the need for a particular item or perhaps revise the item so that it extracts more information or combine the highly correlated items to form a single item that will represent the sum of highly correlated items.

CONCLUSION

The uniform association modeling approach can be very useful in LCA and LTA. It can prevent model misspecification, the addition of spurious classes, and violations of the local independence assumption. It can also eliminate the need for more computationally intensive models with many latent factors used to capture residual associations.

The uniform association approach also provides an easily interpretable parameterization due to the fact that the association parameters behave so similar to the well-understood correlation parameters. Computationally the LCA–UA is straightforward. This is valuable in those situations where many random starting values are used to search for latent class solutions. The computation does not involve numerical integration and is essentially similar in computational work to the estimation of the standard LCA model.

There are currently some limitations to this modeling as implemented in M*plus* Version 7.2. Direct effects from covariates to latent class indicators cannot be included in LCA–UA models together with continuous latent factors measured by the latent class indicators. These limitations, however, might be resolved in the near future.

REFERENCES

Asparouhov, T., & Muthén, B. (2011). Using Bayesian priors for more flexible latent class analysis. In *Proceedings of the 2011 Joint Statistical Meeting, Miami Beach, FL.*

Becker, M. P. (1989). On the bivariate normal distribution and association models for ordinal categorical data. *Statistics & Probability Letters*, 8, 435–440.

Berzofsky, M. E., Biemer, P. P., & Kalsbeek, W. D. (2014). Local dependence in latent class analysis of rare and sensitive events. *Sociological Methods Research*, *43*, 137–170.

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

Hagenaars, J. A. (1988). Latent structure models with direct effects between indicators: Local dependence models. *Sociological Methods & Research*, *16*, 379–405.

Magidson, J., & Vermunt, J. K. (2004). Latent class models. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 175–198). Thousand Oaks, CA: Sage.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115–132.

Muthén, B. (2001). M*plus technical appendices*. Retrieved from http://statmodel.com/download/techappen.pdf

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*, 463–469.

Oberski, D., Kollenburg, G., & Vermunt, J. (2013). A Monte Carlo evaluation of three methods to detect local dependence in binary data latent class models. *Advances in Data Analysis and Classification*, *7*, 267–279.

Qu, T., Tan, M., & Kutner, M. H. (1996). Random-effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, *52*, 797–810.

Samejima, F. (1969). Estimation of a latent ability using a response pattern of graded scores. *Psychometrika Monographs*, *34*(Suppl. 4).

Uebersax J. (1999). Probit latent class analysis with dichotomous or ordered category measures: Conditional independence/dependence models. *Applied Psychological Measurement*, *23*, 283–297.

Van Smeden, M., Naaktgeboren, C. A., Reitsma, J. B., Moons, K. G., & de Groot, J. A. (2013). Latent class models in diagnostic studies when there is no reference standard—A systematic review. *American Journal of Epidemiology*, *179*, 423–431.

APPENDIX

In this appendix we provide some details on the ML estimation of the LCA–UA model. As a first step we describe the ML estimation for the uniform association model with a single class; that is, without mixture modeling. We can rewrite Equation 3 as follows:

$$P(U_1 = a_1, U_2 = a_2, \ldots, U_r = a_r) = \frac{Exp(\mu(a_1, \ldots, a_r))}{\sum_{a_1,a_2,\ldots,a_r} Exp(\mu(a_1, \ldots, a_r))} \tag{A.1}$$

where

$$\mu(a_1, \ldots, a_r) = \sum_i \tau_{i,a_i} + \sum_{i<j} \beta_{ij} a_i a_j.$$

The log-likelihood function F is given by

$$F = \sum_{i,a_i} n_{i,a_i} \tau_{i,a_i} + \sum_{i<j,a_i,a_j} n_{i,j,a_i,a_j} \beta_{ij} a_i a_j - n \log\left(\sum_{a_1,a_2,\ldots,a_r} Exp(\mu(a_1, \ldots, a_r))\right)$$

where n_{i,a_i} is the number of observations for which $U_i = a_i$; n_{i,j,a_i,a_j} is the number of observations for which $U_i = a_j$ and $U_j = a_j$; and n is the total number of observations. To maximize F we need to compute the first derivatives of F with respect to the parameters τ_{i,a_i} and β_{ij} and then use a general maximization algorithm such as the quasi-Newton method that requires only first derivatives evaluation. The derivatives are computed as follows

$$\frac{\partial F}{\partial \tau_{i,a_i}} = n_{i,a_i} - nP(U_i = a_i)$$

where $P(U_i = a_i)$ is the marginal probability implied by the current parameter estimates and can be computed as follows, for $i = 1$:

$$P(U_1 = a_1) = \frac{\sum_{a_2,\ldots,a_r} Exp(\mu(a_1, \ldots, a_r))}{\sum_{a_1,a_2,\ldots,a_r} Exp(\mu(a_1, \ldots, a_r))}.$$

The derivatives with respect to the association parameters are computed as follows:

$$\frac{\partial F}{\partial \beta_{ij}} = \sum_{a_i,a_j} a_i a_j n_{i,j,a_i,a_j} - n \sum_{a_i,a_j} a_i a_j P(U_i = a_i, U_j = a_j)$$

where $P(U_i = a_i, U_j = a_j)$ is again the marginal probability that $U_i = a_i$ and $U_j = a_j$ implied by the current parameter estimates. The standard errors of the parameter estimates can be computed using the first derivatives (see Muthén, 2001), or using the second derivatives of the log-likelihood that are computed similarly.

To generalize the preceding estimation to the case of the LCA–UA model, we follow the EM algorithm described in Muthén and Shedden (1999). The computation of the posterior class probabilities in the E-step is the same as in Muthén and Shedden (1999) with the exception that now the class-specific indicator distribution is computed via Equation 4. The M-step is computed as described previously in the one-class model with the modification that n_{i,a_i} and n_{i,j,a_i,a_j} are now the class-specific quantities derived from the posterior class probabilities computed in the E-step.