

Cross-Lagged Panel Modeling with Categorical Outcomes

Bengt Muthén, Tihomir Asparouhov
& Katie Witkiewitz *

September 26, 2023

*We thank Noah Hastings who provided support with tables and figures.

Abstract

To date, cross-lagged panel modeling has been studied only for continuous outcomes. This paper presents methods that are suitable also when there are categorical outcomes. Modeling, testing, identification, and estimation are discussed for the case of binary and ordinal outcomes. A two-part ordinal model is proposed for ordinal variables with strong floor effects often seen in applications. An example considers the interaction between stress and alcohol use in an alcohol treatment study. Extensions to multiple-group analysis and modeling in the presence of trends are discussed.

Keywords: binary and ordinal variables; panel data; random intercept; RI-CLPM; stress and drinking; alcohol treatment study

1 Introduction

The cross-lagged panel model (CLPM) and its random intercept counterpart RI-CLPM are popular models for investigating longitudinal relationships between two or more variables where the variables at each time point are regressed on themselves and each other at previous time points. For an overview and a discussion of the merits of CLPM and RI-CLPM, see, e.g., Hamaker (2023). To date, however, the literature covers only continuous variables. This paper presents methods that are suitable also when there are categorical variables.

The categorical case needs special considerations in terms of modeling and estimation. Maximum likelihood estimation is generally not feasible but Bayesian and weighted least squares methods can be used. The paper demonstrates analysis methods that work well in practice for both binary and ordinal variables as well as combinations of binary, ordinal, and continuous variables.

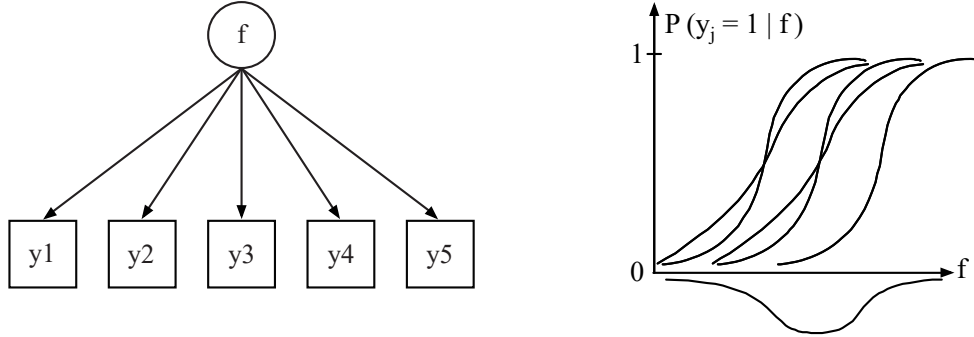
Section 2 discusses modeling, testing, identification, and estimation matters for the case of a binary variable, followed by simulations for the binary univariate case as well as the binary bivariate RI-CLPM case. Section 3 presents applications of analyses with a binary variable using alcohol data from a large randomized treatment study. Section 4 considers ordinal variables and presents a two-part ordinal model suitable for variables that have strong floor effects. Section 5 continues the alcohol treatment example using an ordinal alcohol risk variable. Section 6 presents extensions to the analysis of multiple groups as well as models that allow trends. Section 7 concludes. Mplus scripts for key analyses are shown in the Appendix.

2 Binary, univariate outcome: Modeling and testing concepts

The left part of Figure 1 shows a model with a factor f influencing five variables y_1 - y_5 . With continuous observed variables y , the arrows represent linear regressions with slopes referred to as factor loading. For binary observed variables, the arrows represent non-linear regressions as shown in the right part of the figure. The regressions represent the probability of $y=1$ as opposed to 0 as a function of the value of the factor. For example, the items may represent incorrect/correct responses to five different math test where the factor is referred to as ability or achievement. As the ability factor value increases, the probability of answering the item correctly increases. For a given ability value, the five regression curves show that the right-most curve represents the most difficult item in that the probability of answering it correctly is the lowest. The curves also differ in how well they discriminate between ability values where steeper curves represent larger probability differences between lower and higher ability values. Parameterization in terms of difficulty and discrimination is used in Item Response Theory (IRT) applications whereas factor analysis applications use the parameterization of item threshold and factor loading. The threshold formulation is useful for the generalization to ordinal outcomes. Typically, a normally distributed factor is assumed and the curves are represented by logistic or probit functions.

An equivalent formulation of the one-factor model is shown in Figure 2. This uses a continuous latent response variable y^* underlying each binary observed variable. The

Figure 1: 1-factor model (a circle represents a latent variable and a square represents an observed variables)



y^* variables have ordinary linear regressions on the factor where the short arrows below the y^* circles represent residuals. The regression coefficients are the factor loadings. When y^* exceeds a threshold, $y = 1$ is observed and otherwise $y=0$. Continuing the math test example, the y^* value induced by the factor may be just above the threshold or far above it, thereby capturing the idea that the skill needed to solve the item correctly could be measured in a finer gradation.

In this paper, a multivariate normal distribution is assumed for the y^* variables. This is equivalent to specifying a normal distribution for the factor f together with normally distributed residuals. This is because a sum of normal distributions is also normal. In terms of Figure 1, a normal factor together with probit regressions for y on f also leads to normal y^* variables. Specifying a logistic regression instead does not lead to y^* normality. The correlations between the normal y^* variables are referred to as tetrachoric when the observed variables are binary.

The y^* concept also covers the ordinal case shown in Figure 3. This shows an example of two 5-category variables represented by two y^* variables, each of which has four thresholds. The thresholds are not equally spaced over the y^* distribution, causing strong ceiling effects. There is a linear relationship between the y^* variables as follows from normality which motivates the use of regular correlations among the y^* variables (in the ordinal case, the correlations are referred to as polychoric and polyserial if one variable is continuous).

The assumption of multivariate normality for the y^* variables does not necessarily fit every data set and it is important to test the assumption as will be discussed shortly. Extensions of the model for the ordinal case is often necessary to achieve good model fit as will be discussed in Section 4.1.

So far, the observed variables in Figure 1 and Figure 2 have been viewed as different variables, that is, corresponding to a cross-sectional analysis. In that case, the focus is on the relationships between f and the y 's. This paper, however, considers longitudinal analysis where the same variable is measured at several time points. The primary focus is on the relationship among the y 's. In the longitudinal case, the factor of Figure 2 is viewed as a random intercept i where the factor loadings are all fixed at 1. The strength of correlation among the y^* variables at different time points is constant and

Figure 2: 1-factor model using a latent response variable representation

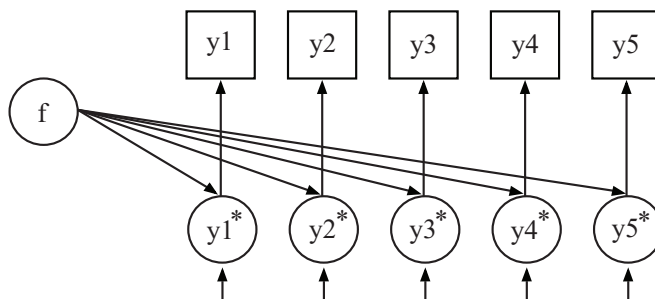


Figure 3: Ordinal model using a latent response variable representation

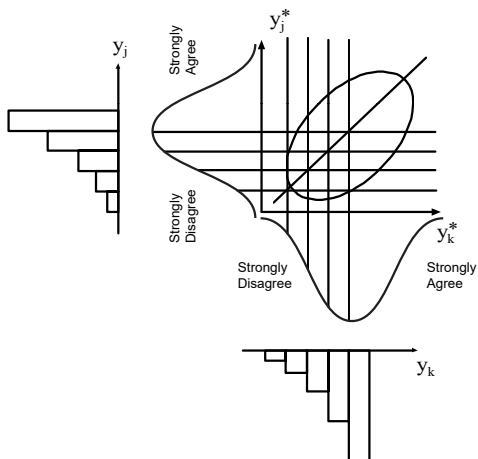
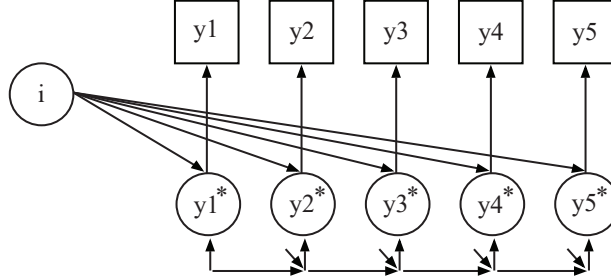


Figure 4: 1-factor (random intercept) model with auto-regressive residuals



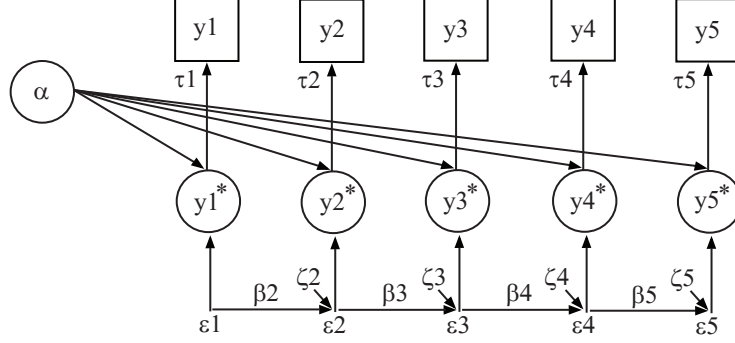
captured by the random intercept variance. With a trend, one or more random slope latent variables can be added.

In repeated measurement modeling for continuous outcomes it is well-established that correlation among the residuals needs to be allowed for in order for the model to fit well (see, e.g., Chi & Reinsel, 1989). The random intercept alone cannot capture all the correlation among the y^* 's and thereby not among the y 's. This residual correlation is conveniently translated to the multivariate probit framework. Figure 4 shows a model with linear auto-regressions of lag 1 for the y^* residuals. This figure represents a categorical counterpart to the univariate part of a continuous-outcome RI-CLPM (Hamaker et al., 2015) where the relationships among the residuals capture the dynamic within part of the model. The identification and estimation of this model is discussed in Section 2.1.

Model fit assessment for categorical outcomes can be done in several different ways. One approach is analogous to what is used in structural equation modeling, where fit to covariances or correlations is considered. For the multivariate probit model, this amounts to fit of correlations among the latent response variables y^* underlying each observed categorical y variable. This was studied in Muthén et al. (1997) using chi-square testing based on a weighed least-squares estimator (WLSMV) for a multivariate probit model. The Muthén et al. (1997) WLSMV chi-square works well when the number of variables is not large and the sample size is not small which makes it particularly suitable for cross-lagged panel modeling. Checking of model fit based on chi-square for a multivariate probit model is also possible with Bayesian estimation as discussed in Asparouhov and Muthén (2021a, b) using posterior predictive checking that produces a posterior predictive p-value (PPP). Using any fit statistic, the general posterior predictive checking approach is to compute the fit statistic for the observed data, generate a fit statistic distribution based on generated data from the estimated model, and find the proportion of cases where the latter is larger than the former. Based on the same overall chi-square as used with WLSMV, the Bayes approach, however, has low power for binary outcomes and is less powerful than the WLSMV chi-square test (Asparouhov & Muthén, 2021a). The Bayes approach is more powerful for ordinal variables.

A second approach considers the fit to the data in the form of response patterns, that is, a frequency table for all variables. A model may fit the y^* correlations well but not the frequency table. Even a just-identified y^* model that includes all possible correlations may not fit the frequency table. This was discussed in Muthén (1993).

Figure 5: Random intercept model with auto-regressive residuals



With categorical variables, the model can be tested against data using the standard Pearson and likelihood-ratio chi-square frequency table tests. Summing over the cells of the table, these two tests are expressed as:

$$Pearson : \sum_j (o_j - e_j)^2 / e_j \quad (1)$$

$$Likelihood\ ratio : 2 \sum_j o_j \log(o_j / e_j) \quad (2)$$

There are, however, typically too many frequency table cells with many cells having estimated frequencies close to zero, invalidating the tests. For example, with 8 binary variables there are $2^8 = 256$ possible response patterns, where many patterns are often not observed (zero cells in the frequency table) leading to the two tests disagreeing strongly and becoming useless. A practical approach is to consider fit to the most frequent response patterns, e.g., the twenty most frequent. There are, however, alternative frequency table checks where the tables are collapsed into univariate and bivariate tables which ensures higher frequencies. In particular, bivariate frequency checking is a useful way to find model misspecification. Asparouhov and Muthén (2022) presents significance testing of standardized residuals for both response patterns and bivariate frequency tables.¹

2.1 Binary univariate case: Identification and estimation

Consider again the binary random intercept probit model shown in Figure 5 for $T = 5$. It can be expressed as follows for individual i and time point t . For the y_t^* continuous latent response variable at time t with threshold parameter τ_t , $y_t^* > \tau_t$ implies $y_t = 1$

¹In Mplus, TECH10 for WLSMV and Bayes give standardized residuals for response patterns, uni- and bi-variate frequency tables, and Bayes PPP for Pearson fit to uni- and bi-variate tables.

while otherwise $y_t = 0$. The model specifies the linear relations

$$y_{it}^* = \alpha_i + \epsilon_{it}, \quad (3)$$

$$\epsilon_{it} = \beta_t \epsilon_{it-1} + \zeta_{it}; \quad t = 2, \dots, T, \quad (4)$$

$$\epsilon_{i1} = \zeta_{i1}, \quad (5)$$

where $\alpha_i \sim N(0, \psi)$ represents the random intercept, ϵ_{it} represents the residual for y_{it}^* , β_t represents the auto-regression, and $\zeta_t \sim N(0, \theta_t)$ are the residuals in the auto-regressions.

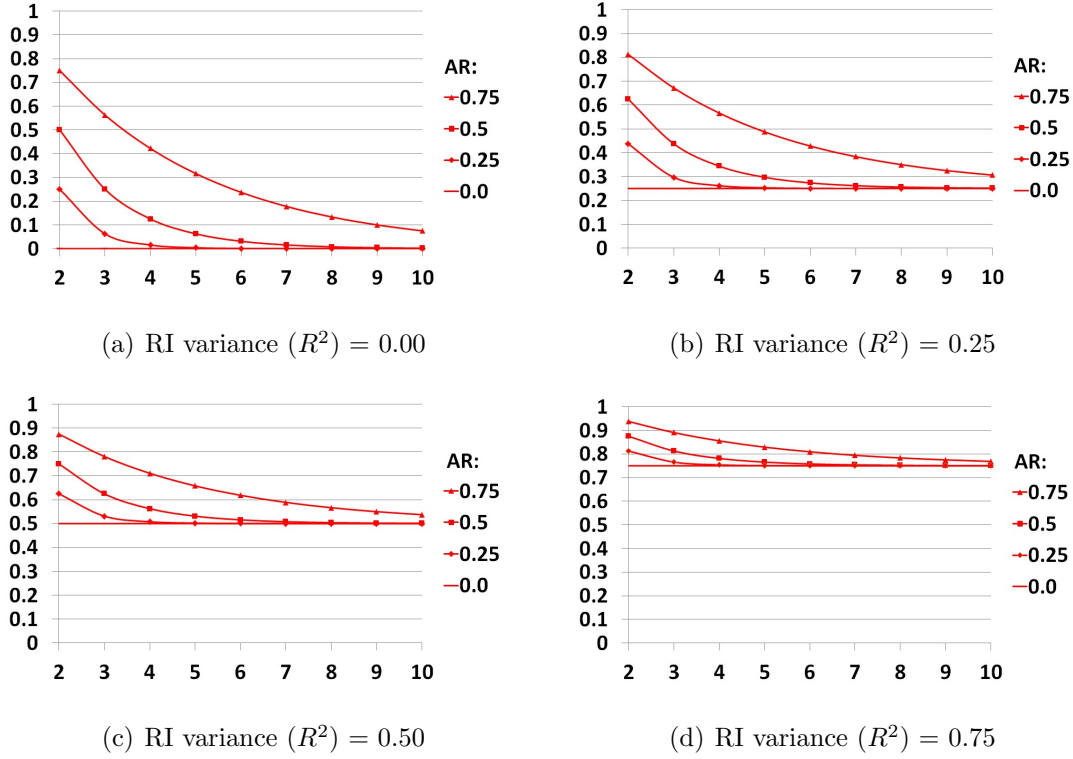
A major distinction between this model and the corresponding model for continuous outcomes is that the variances for the ζ residuals are not all identified. This can be portrayed as the loss of information due to not observing the y^* variables directly but only in a discretized fashion. For a binary variable, the mean π and variance $\pi(1 - \pi)$ are mathematically related so that the sample mean and variance do not provide information about two separate parameters as in the continuous case. A maximum of $T - 1$ $V(\zeta_{it})$ variances can be identified as will be discussed below. Typically, however, it is difficult to estimate $T - 1$ variances without incurring large standard errors for model parameters. Estimation faces an empirical identification issue where the number of identifiable variances depends on the data in terms of correlations across time, number of time points, and sample size. Fixing the first residual variance at 1 while estimating the remaining ones would account for the remaining ζ variances possibly being smaller since the remaining ones are residuals in the regressions among the ϵ residuals. In practice, however, the explained variance in the ϵ regressions is small so that the variances for the ζ_2, \dots, ζ_T residuals are not that much smaller than the residual variance at the first time point. This means that it often makes little difference to estimates of other parameters when letting the ζ_2, \dots, ζ_T variances be estimated. Fixing all residual variances at 1 is often reasonable as the default. A first step to relax this default model would be to free the first residual variance but even this variance may obtain a large standard error.

Identification issues for the binary random intercept model with auto-regressive residual are as follows. Consider first the impact of random intercept variance and auto-regression on the y^* correlations over time. Figure 6 shows four panels which differ by the magnitude of the random intercept variance. Within each panel four curves are shown differing by the magnitude of auto-regression. The curves show the correlation between y^* at $t = 1$ and y^* at $t = 2, 3, \dots, T$ for $T = 10$. The curves are computed by the formula $\text{Corr}(y_1^*, y_t^*) = \psi + \beta^{t-1}(1 - \psi)$ where ψ is the random intercept variance, β is the constant auto-regression among the residuals, and y^* variances are all 1. Due to the unit y^* variances, ψ is the same as the R-square of y^* explained by the random intercept. The formula shows that as t increases, β^{t-1} decreases due to $|\beta| < 1$. This means that as t increases, the second term on the right-hand side of the formula decreases and the correlation decreases down towards the asymptote of ψ , the random intercept variance. The larger ψ is, the higher the asymptote. Also, the higher the auto-regression, the slower the decline in correlations over time. For instance, panel (c) of Figure 6 shows the case of random intercept variance 0.5 where with auto-regression of 0.25, the correlation between the first two time points is a little above 0.6 and declines to the asymptote of 0.5 at approximately $t = 4$.

While the threshold parameters are trivially identified in terms of the proportion $y=1$ for the outcomes, the key model parameters ψ , β_t , and residual variances θ_t need

to be identified in terms of correlations among the y^* variables. Figure 6 suggests how this identification is accomplished. This can be viewed as choosing the estimates of the ψ , β_t , θ_t parameters to fit the curve of the sample correlations at different time distances.

Figure 6: RI and AR1 impact on y^* correlations across time (T=10)



Muthén and Asparouhov (2002), see also Muthén (1996), showed that growth modeling with categorical outcomes and no auto-correlated residuals can identify $T - 1$ residual variances in addition to the random effect variances (a simulation with correlated residuals was presented in Muthén, 1996). Random intercept modeling is a special case of such modeling. A growth model for binary outcomes needs $T \geq 4$ while the random intercept model needs $T \geq 3$. Figure 6 shows that in the case of random intercept variance $\psi = 0.5$ and auto-regression 0.25, the auto-regression itself gives zero correlation contribution at $t \approx 4$, that is, a time distance of at least 3. Three time points spaced at least 3 time points apart would therefore have no auto correlation, leading to the identification of ψ and $T - 1$ θ_t 's based on Muthén and Asparouhov (2002).

The actual identification expressions can be presented as follows. Consider again the case of $T = 10$ and an auto-regression of 0.25 so that the auto-regression itself gives zero contribution at $t \approx 4$. Fixing the residual variance $\theta_1 = 1$ and knowing the random intercept variance ψ , time 1 correlations identify the remaining $T - 1$ θ_t 's when considering the correlations for longer time distances where the auto-regressive contributions are zero:

$$Corr(y_1, y_t) = \psi / (\sqrt{\psi + 1} \sqrt{\psi + \theta_t}). \quad (6)$$

Consider again the example of the auto-regressive contribution to the correlation being zero for $t \geq 4$, that is, at a time distance of at least 3. With ψ known, the correlation between $t = 1$ and $t = 4$ identifies θ_4 , the correlation between $t = 1$ and $t = 5$ identifies θ_5 , etc. up to the correlation between $t = 1$ and $t = 10$ identifying θ_{10} . In all cases, the time distance is at least 3. It remains to identify θ_2 and θ_3 for which the time distance is less than 3. But knowing for instance θ_{10} from the previous reasoning, the correlation between $t = 2$ and $t = 10$ identifies θ_2 and the correlation between $t = 3$ and $t = 10$ identifies θ_3 . The T-1 auto-regression parameters of β are then identified from among the remaining correlations.

Estimation of the univariate random intercept probit model may be carried out by maximum likelihood (ML), weighted least-squares (WLSMV), and Bayes. The ML estimator, however, needs to use numerical integration over the T+1 latent variables and is therefore not feasible for a typical number of time points. WLSMV is a fast estimator which works well when there is little missing data. It can also give information about the empirical identification status in that it presents the condition number of the estimated information matrix.² The WLSMV uses a convenient residual specification where the residuals can be referred to directly as latent variables; see Asparouhov and Muthén (2023).³ This means that the residuals can be regressed on each other as is needed for the auto regressions and for the cross-lagged regressions in the bivariate case. In contrast, the weighted least squares estimation in Muthén (1983, 1984, 1996) could estimate only correlations among residuals. WLSMV is, however, disadvantaged because it does not handle MAR missingness like the full-information ML and Bayes estimators. The Bayes estimator combines a practical approach with full-information estimation. The Bayes approach to be used in the application sections uses an efficient algorithm together with the residual specification (see Asparouhov & Muthén, 2023). In some cases, non-symmetric confidence/credibility intervals are needed. With Bayes, this is obtained automatically while with WLSMV, bootstrapping is needed. Because of the arbitrary scale of the y^* latent response variables, it is useful to present estimates in a standardized metric where the y^* variances are 1.

Table 1 shows the number of parameters τ_t , ψ , β_t , the number of sample statistics, and degrees of freedom for the univariate random effect probit model with fixed residual variances θ_t . Here, degrees of freedom refers to the number of restrictions imposed on the sample statistics of univariate proportions and correlations. This is the degrees of freedom of the chi-square test of fit for the weighted least-squares estimator (WLSMV). $T = 3$ is the minimum number of time points required for identifying the model. It should be noted, however, that $T = 3$ is a bare minimum for this type of analysis with categorical outcomes because this generally provides little information to distinguish between correlation due to the random intercept versus due to autocorrelation. More time points are strongly recommended.

2.2 Binary bivariate case: RI-CLPM

The generalization of the binary univariate outcome model to the bivariate case is shown in Figure 7. This is the binary counterpart to RI-CLPM (Hamaker et al., 2015). The bivariate model offers no further identification complications beyond the univariate

²In Mplus, this is the ratio of smallest to largest eigenvalue of the estimated information matrix.

³This is the hat notation in Mplus. The Theta parameterization is used.

Table 1: Number of parameters, sample statistics, and degrees of freedom for the binary outcome univariate random effect probit model with fixed residual variances

T	# parameters	# sample statistics	DF
3	3+1+2=6	6	0
4	4+1+3=8	10	2
5	5+1+4=10	15	5
10	10+1+9=20	55	35

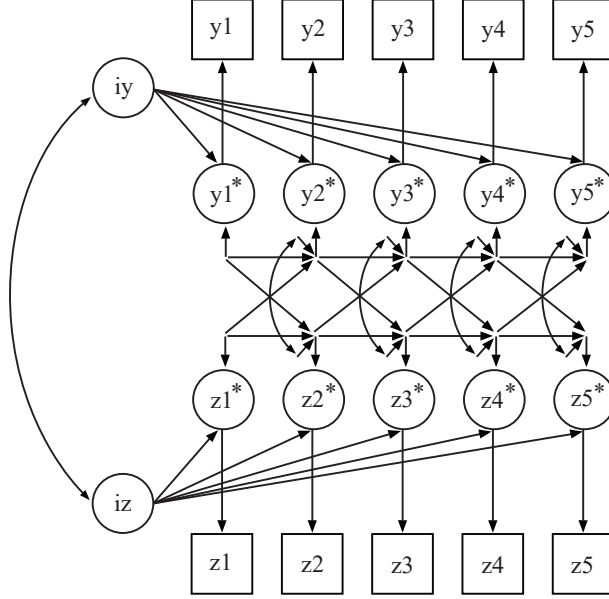
case. Each univariate part follows the identification rules just discussed. The cross-lagged parameters are identified in terms of the correlations between the y^* and z^* variables. Estimation can be performed by Bayes or WLSMV.

2.3 Binary outcome simulations: Univariate and bivariate cases

Simulations for the case of a univariate binary outcome are shown in Table 2 and Table 3 for the case of $T = 5$. Population values are based on analyses with suicidal ideation and substance abuse data (Ialongo, 2022). Data are generated with time-varying thresholds representing binary outcomes of low prevalence with $P(y_1 = 1)$ ranging from 0.12 to 0.20. The low-prevalence binary case represents lower-end categorical information, so that better performance can be expected for binary variables with more even split and for ordinal variables. Time-varying thresholds and time-varying auto-regressions are specified in the analyses. Residual variances are fixed at 1. For simplicity, there is no missing data. 500 replications are carried out using both WLSMV and Bayes estimation. With Bayes, 2,000 draws (iterations) are recorded, having skipped every 10th iteration for better standard error estimation. WLSMV uses first- and second-order information from univariate proportions and correlations whereas Bayes uses full information. With binary outcomes and $T = 5$, WLSMV uses information from 15 sample statistics (5 univariate proportions and 10 tetrachoric correlations). The $2^5 = 32$ response patterns represent the potential raw data that is used by the full-information estimation by Bayes. This means that Bayes uses about twice as many sample statistics as WLSMV which is expected to reduce the variability of the estimates. Many of the response pattern frequencies are, however, low which implies that WLSMV use key parts of the available information so that the reduction in variability by Bayes may not be large.

WLSMV results in Table 2 show that at $N = 500$, somewhat biased parameter estimates and standard errors are obtained. In particular, the random intercept variance is overestimated and its standard error underestimated. These biases may result in inclusion of a random intercept when it is not needed. The model has 5 degrees of freedom. The WLS chi-square testing results are good with a mean of 4.86 and a 5% reject proportions of 0.042. For $N = 1000$, all results are good. Bayes results in Table 3

Figure 7: Bivariate RI-CLPM for categorical outcomes



show a similar picture. For $N = 500$, the Bayes standard errors perform better than WLSMV and are on the whole smaller as expected by a full-information estimator as compared to the limited-information estimation by WLSMV.

Simulations for the bivariate binary outcome case of RI-CLPM with $T=5$ are shown in Table 4 for the WLSMV estimator and in Table 5 for the Bayes estimator. The population parameter values are again chosen from the Ialongo (2022) study with one variable having the same univariate parameters as those of the univariate simulation and the other with similar values. Time-varying thresholds and auto-regressions are again specified in the analyses. The cross-lagged effects are specified as time varying. For simplicity, there is no missing data. With Bayes, 5,000 draws (iterations) are recorded, having skipped every 10th iteration for better standard error estimation. 500 replications are carried out. With two binary outcomes and $T = 5$, WLSMV uses information from 55 sample statistics (10 univariate proportions and 45 tetrachoric correlations). The $2^{10} = 1024$ response patterns represent the potential raw data that is used by the full-information estimation by Bayes. This means that Bayes uses about 20 times as many sample statistics as WLSMV and this is expected to make important reduction in the variability of the estimates. The tables show results for only the new parameters of cross-lagged effects.

For WLSMV, $N=500$ is clearly insufficient as is seen in the parameter estimate bias and standard error bias. $N=1000$ shows an improvement and $N=2000$ shows acceptable results. With $N=2000$, the power to detect the $Z7^{\wedge} ON Y6^{\wedge}$ effect is estimated as 0.858 (see the last column). The model has 34 parameters and 21 degrees of freedom. The WLSMV chi-square testing of model fit is performing well with chi-square mean and 5% rejection proportions for $N=1000/N=2000$ of 20.3/20.5 and 0.05/0.04.

For Bayes, the results are acceptable already at $N=500$ and excellent at $N=1000$.

Table 2: Monte Carlo results for univariate binary outcome for $T = 5$ using WLSMV

	Population	ESTIMATE		S. E.		95%	% Sig
		Average	Std. Dev.	Average	M. S. E.	Cover	Coeff
N = 500							
Z5^ ON Z4^	0.122	0.0890	0.1762	0.1588	0.0320	0.924	0.118
Z6^ ON Z5^	0.089	0.0754	0.1795	0.1715	0.0323	0.938	0.078
Z7^ ON Z6^	0.166	0.1556	0.1940	0.1750	0.0377	0.936	0.162
Z8^ ON Z7^	0.126	0.1245	0.1788	0.1723	0.0319	0.944	0.138
Thresholds							
Z4\$1	1.282	1.3066	0.1311	0.1212	0.0177	0.932	1.000
Z5\$1	1.438	1.4659	0.1341	0.1269	0.0187	0.936	1.000
Z6\$1	1.663	1.7010	0.1522	0.1363	0.0245	0.926	1.000
Z7\$1	1.786	1.8233	0.1515	0.1412	0.0243	0.930	1.000
Z8\$1	1.879	1.9205	0.1510	0.1466	0.0245	0.946	1.000
Variances							
IZ	1.536	1.6280	0.3284	0.2934	0.1161	0.920	1.000
N = 1000							
Z5^ ON Z4^	0.122	0.1060	0.1136	0.1104	0.0131	0.942	0.166
Z6^ ON Z5^	0.089	0.0834	0.1236	0.1214	0.0153	0.944	0.098
Z7^ ON Z6^	0.166	0.1582	0.1256	0.1239	0.0158	0.954	0.264
Z8^ ON Z7^	0.126	0.1271	0.1203	0.1219	0.0145	0.956	0.198
Thresholds							
Z4\$1	1.282	1.2992	0.0846	0.0848	0.0074	0.960	1.000
Z5\$1	1.438	1.4551	0.0872	0.0882	0.0079	0.956	1.000
Z6\$1	1.663	1.6882	0.1012	0.0946	0.0108	0.922	1.000
Z7\$1	1.786	1.8074	0.0969	0.0977	0.0098	0.948	1.000
Z8\$1	1.879	1.9037	0.1006	0.1014	0.0107	0.960	1.000
Variances							
IZ	1.536	1.5953	0.2059	0.2018	0.0458	0.956	1.000

Table 3: Monte Carlo results for univariate binary outcome for $T = 5$ using Bayes

	Population	ESTIMATE		S. E.		95%	% Sig
		Average	Std. Dev.	Average	M. S. E.	Cover	Coeff
N = 500							
Z5^ ON Z4^	0.122	0.0967	0.1546	0.1575	0.0245	0.956	0.102
Z6^ ON Z5^	0.089	0.0669	0.1571	0.1671	0.0251	0.960	0.062
Z7^ ON Z6^	0.166	0.1571	0.1749	0.1717	0.0306	0.936	0.158
Z8^ ON Z7^	0.126	0.1347	0.1760	0.1731	0.0310	0.938	0.156
Thresholds							
Z4\$1	1.282	1.3044	0.1190	0.1195	0.0146	0.942	1.000
Z5\$1	1.438	1.4800	0.1266	0.1259	0.0177	0.932	1.000
Z6\$1	1.663	1.7088	0.1361	0.1345	0.0206	0.950	1.000
Z7\$1	1.786	1.8283	0.1409	0.1393	0.0216	0.938	1.000
Z8\$1	1.879	1.9298	0.1381	0.1457	0.0217	0.952	1.000
Variances							
IZ	1.536	1.6598	0.2836	0.2962	0.0956	0.940	1.000
N = 1000							
Z5^ ON Z4^	0.122	0.1072	0.1101	0.1091	0.0123	0.946	0.172
Z6^ ON Z5^	0.089	0.0841	0.1130	0.1167	0.0128	0.954	0.098
Z7^ ON Z6^	0.166	0.1585	0.1278	0.1206	0.0164	0.936	0.288
Z8^ ON Z7^	0.126	0.1307	0.1271	0.1204	0.0161	0.928	0.216
Thresholds							
Z4\$1	1.282	1.2940	0.0866	0.0833	0.0076	0.952	1.000
Z5\$1	1.438	1.4598	0.0869	0.0869	0.0080	0.954	1.000
Z6\$1	1.663	1.6874	0.0924	0.0928	0.0091	0.950	1.000
Z7\$1	1.786	1.8067	0.0947	0.0960	0.0094	0.946	1.000
Z8\$1	1.879	1.9037	0.0940	0.0998	0.0094	0.952	1.000
Variances							
IZ	1.536	1.5962	0.1844	0.1986	0.0375	0.944	1.000

Table 4: Monte Carlo results for bivariate binary RI-CLPM, $T = 5$, WLSMV

	Population	ESTIMATE		S. E.		95%	% Sig
		Average	Std. Dev.	Average	M. S. E.	Cover	Coeff
N = 500							
Y7^ ON Z6^	0.213	0.1987	1.1535	0.5677	1.3280	0.941	0.134
Z7^ ON Y6^	0.375	0.4620	1.7628	0.7869	3.1086	0.899	0.312
N = 1000							
Y7^ ON Z6^	0.213	0.2199	0.2962	0.1865	0.0876	0.918	0.258
Z7^ ON Y6^	0.375	0.3882	0.2549	0.1876	0.0650	0.896	0.584
N = 2000							
Y7^ ON Z6^	0.213	0.2116	0.1389	0.1239	0.0193	0.930	0.422
Z7^ ON Y6^	0.375	0.3732	0.1383	0.1254	0.0191	0.924	0.858

For $N=2000$, the power to detect the $Z7^{\wedge}$ ON $Y6^{\wedge}$ effect is estimated as 0.944. As expected, Bayes has lower variability in the estimates than WLSMV. The advantage of the full-information Bayes estimator versus the limited-information WLSMV estimator is clear from these results.

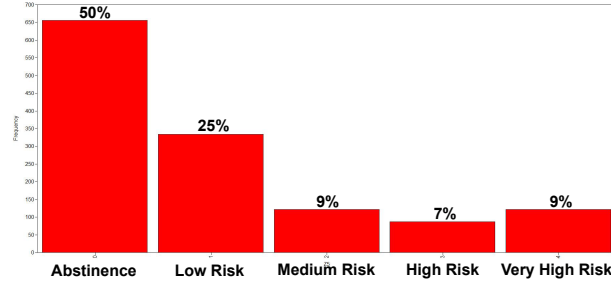
3 Binary outcome example

Data from the COMBINE Study of Alcohol Use Disorder are used to illustrate the techniques for categorical outcomes. COMBINE is a 16-week, multisite randomized double-blind clinical trial comparing treatments of alcohol dependence (Anton et al., 2006). The sample size is 1,383. The measurement occasions to be considered here are: Baseline, week 1, week 2, week 4, week 6, week 8, week 10, week 12, week 16. There are also follow-up measurement occasions up to week 52. For this illustration, the $T = 8$ time points of the treatment are used, week 1 - week 16. The focus is on the relationship between perceived stress and alcohol use during the trial. There is a robust literature examining associations between alcohol and stress using preclinical models with non-human animals, human laboratory studies, and intensive longitudinal studies (see Armeli et al., 2000; Becker, 2017; Sinha, 2022), but few studies have examined bidirectional effects during treatment among individuals with alcohol use disorder. The stress variable is based on a 4-item, brief version of The Perceived Stress Scale with

Table 5: Monte Carlo results for bivariate binary RI-CLPM, $T = 5$, Bayes

	Population	ESTIMATE		S. E.		95%	% Sig
		Average	Std. Dev.	Average	M. S. E.	Cover	Coeff
N = 500							
Y7^ ON Z6^	0.213	0.2409	0.2234	0.2629	0.0506	0.962	0.188
Z7^ ON Y6^	0.375	0.3788	0.2398	0.2645	0.0574	0.956	0.394
N = 1000							
Y7^ ON Z6^	0.213	0.2348	0.1645	0.1705	0.0275	0.952	0.338
Z7^ ON Y6^	0.375	0.3803	0.1713	0.1731	0.0293	0.940	0.682
N = 2000							
Y7^ ON Z6^	0.213	0.2244	0.1102	0.1130	0.0122	0.954	0.550
Z7^ ON Y6^	0.375	0.3798	0.1140	0.1160	0.0130	0.950	0.944

Figure 8: Distribution of the alcohol risk variable at week 4



scores of 0 to 16 (McHugh et al., 2013). The analyses will use two different measures of alcohol use. Alcohol Risk is measured as a 5-category variable: Abstinence, low risk, medium risk, high risk, very high risk. These are WHO-defined drinking risk levels based on amount of alcohol consumed. The Alcohol Risk variable is also used to define a binary variable of Abstinence versus not by combining the four highest categories. The analyses will treat the stress variable as continuous using linear regressions. This is not appropriate for Alcohol Risk which as shown in Figure 8 has a strong floor effect that would bias results from a linear model. The binary Abstinence variable is analyzed first, whereas analyses using ordinal models for Alcohol Risk are presented later.

3.1 CLPM and RI-CLPM analyses using the binary abstinence variable

This section presented results of the $T = 8$ analysis of the binary abstinence variable. 8% have missing data at all eight time points and are deleted, resulting in a sample size of 1,375. There is rather little attrition. At the last time point of week 16, 93% remain in the sample. The proportion non-abstinent varies between 0.45 and 0.51. The analyses will use both the limited information WLSMV estimator and the full-information Bayes estimator but large differences in results are not expected due to the low degree of missing data. With categorical data, the raw data can be represented by the response patterns observed in the sample. With binary outcomes and $T = 8$, there are $2^8 = 256$ possible patterns. In these data, 234 patterns have non-zero frequency, 125 patterns have frequency greater than 1, and 20 patterns have frequency greater than 10. The 234 observed response patterns constitute the full information in the raw data which is used by the Bayes estimator. In contrast, the WLSMV estimator uses only the 36 first- and second-order sample statistics corresponding to the proportions and correlations among the 8 variables.

The 20 most frequent response patterns are shown in Table 6. It is seen that 23% of the sample has the response pattern of all zeros, that is, individuals who are abstinent at all eight time points. The estimated frequencies for WLSMV and Bayes in Table 6 refer to the unrestricted binary probit model where no restrictions are placed on the thresholds or correlations among the continuous latent response variables. This tests if a probit model is suitable for the data in the first place before adding restrictions on the correlations. The unrestricted probit model has 36 parameters (8 thresholds

Table 6: Response pattern frequencies for abstinence outcome

Pattern	Percentage	Observed Frequency	Estimated Frequency		Stand'd Residual Z-score	
			WLSMV	BAYES	WLSMV	BAYES
00000000	22.69	312.00	291.16	289.78	1.28	1.37
11111111	19.85	273.00	291.67	291.80	-1.17	-1.18
11111110	2.76	38.00	26.23	25.47	1.86	1.98
00000001	2.47	34.00	31.02	33.42	0.49	0.10
01111111	1.60	27.00	28.62	28.88	-0.29	-0.33
00011111	1.60	22.00	15.89	14.00	1.26	1.66
00111111	1.38	19.00	16.52	18.91	0.54	0.02
11111011	1.09	15.00	12.11	10.59	0.72	1.10
10000000	1.09	15.00	12.23	12.11	0.68	0.72
11110000	1.02	14.00	11.98	11.35	0.52	0.68
00010000	1.02	14.00	12.11	16.52	0.48	-0.59
00000111	0.95	13.00	11.98	11.60	0.27	0.37
11111101	0.95	13.00	10.84	10.72	0.57	0.61
11101111	0.87	12.00	11.10	10.09	0.25	0.53
00000100	0.87	12.00	10.72	8.70	0.35	0.92
00001000	0.87	12.00	11.73	11.35	0.07	0.18
11000000	0.87	12.00	10.34	12.36	0.46	-0.10
11111100	0.87	12.00	9.58	10.21	0.67	0.49
11110111	0.87	12.00	11.85	11.98	0.04	0.01
11011111	0.80	11.00	8.20	8.57	0.81	0.70

and 28 correlations) whereas an unrestricted frequency table model has $2^8 - 1 = 255$ parameters. In other words, the unrestricted probit model is a very parsimonious representation of the data. Table 6 shows that this model fits reasonably well with no significant standardized residuals among the 20 most frequent patterns for WLSMV and only 1 significant standardized residual for Bayes (z-score = 1.98).

Table 7 shows univariate analyses of the binary abstinence variable using the Bayes estimator.⁴ The first model is the just mentioned unrestricted probit model. As described in Section 2, the Bayes estimator provides a posterior predictive p-value (PPP) where $PPP > 0.05$ is often used as a descriptive measure of acceptable fit and PPP around 0.5 is considered excellent. The unrestricted probit model, model 1, gets a PPP of 0.520. PPP is, however, always around 0.5 for a model that is just-identified like model 1. Although Bayes uses more information than first- and second-order moments by using the further information in the raw data, the PPP model testing is based on chi-square which still concerns fit to the first- and second-order moments so the model is still just-identified. In other words, the H0 and H1 models are the same.

⁴Mplus input for select models are shown in Appendix.

Table 7: Bayes results for univariate analysis of abstinence ($N = 1375$, $T = 8$)

	Model	# par's	PPP	# Sig. Residuals		Comment
				Resp Pattern	Bivar	
1.	Unrestricted	36	0.520	1*	0	Good fit
2.	AR1	15	0.082	2	4	Ok fit
3.	AR2	21	0.474	0	0	Good fit
4.	RI-AR1	16	0.189	2	4	Good fit
5.	RI-AR2	22	0.472	0	0	Good fit

* 3rd most frequent pattern with 38 observations and Z-score=1.98.

WLSMV chi-square testing has zero degrees of freedom and can therefore also not be used to test the unrestricted model 1.

As mentioned in Section 2, testing the model against data can be done by checking the fit to the response patterns and the bivariate frequency tables. In the current analyses, standardized residual fit for the 20 most frequent response patterns are used as one part in assessing overall model fit. In addition, fit to the bivariate frequency tables is considered. There are $8(8 - 1)/2 = 28$ bivariate frequency tables and since each table has 4 cells, there are 112 cells total available for testing of standardized residuals. Making the crude approximation of independent tests in the cells, a Type I error of 5% for the 112 test, or 6 tests, are expected to be significant when the model is correct. This number will be used as a threshold for a descriptive fit assessment. The unrestricted model of Table 7 shows that only 1 response pattern, the third most frequent pattern with frequency 38, has a significant misfit in terms of the standardized residual and the z-score is only 1.98. None of the bivariate standardized residuals show misfit. The overall assessment is that this model fits the data well which means that testing of restrictions on the correlations as in models 2 - 5 is appropriate.

Models 2 and 3 of Table 7 are models without the random intercept and using auto-regression with lags 1 and 2, respectively. From the improvement in fit, it is clear that a lag of 2 is motivated. Models 4 and 5 use a random intercept, where again there is a preference for using a lag of 2. Model 5 with lag 2 obtains a small variance of 0.089 for the random intercept with Bayesian credibility interval [0.001 0.375]. This indicates that there is not a large trait component for the tendency to report abstinence or not over the 8 weeks. Letting the first residual variance be freely estimated as discussed in Section 2.1 does not improve fit and gives a large standard error for the residual variance.

Turning to the analysis of primary interest, Table 8 shows results for bivariate analysis of the binary abstinence variable and the continuous stress variable. For these two variables, the Bayes PPP and the WLSMV chi-square refer to the full bivariate model, whereas the number of significant residuals refer to the binary abstinence variable only. The random intercept model for the bivariate case was shown in Figure 7, except that the continuous latent response variables are directly observed for the stress variable. The cross-lagged effects are lag 1 for all models. Time invariance is not imposed for

any of the parameters. This model is referred to as RI-CLPM. The CLPM models 1, and 2 that do not have random intercepts fit poorly. The RI-CLPM model 3 with auto-regressions of lag 1 also fits poorly, whereas the RI-CLPM model 4 with auto-regressions of lag 2 fits well as assessed by both Bayes and WLSMV. Models 5 - 8 will be discussed in the next section.

For model 4, the abstinence random intercept variance is now somewhat larger than in the univariate analysis, with Bayes estimate 0.191 and CI [0.081 0.413]. As mentioned earlier, it is useful to present estimates in a standardized metric given the arbitrary scale of the latent response variables. The random intercept variance estimate for abstinence translates to small R-square values for the latent response variables at the different time points, with values between 0.03 and 0.16. In contrast, the random intercept variance for the stress outcome gives high R-square values for the latent response variables, ranging from 0.53 to 0.56. For the residual auto regressions, the abstinence R-square values are in the range of 0.7 to 0.8 whereas the stress R-square values are lower, ranging from 0.10 to 0.23. The WLSMV estimates are similar. Freeing the first residual variance for abstinence as discussed in Section 2.1, does not lead to an estimate significantly different from the default of 1 for either estimator. The significance of the cross-lagged effects is found to be the same when having this residual variance fixed or free.

The interesting key finding of model 4 is that the cross-lagged estimates show no significant influence of stress on non-abstinence whereas all seven cross-lagged effects of non-abstinence on stress are significant with standardized values ranging from 0.17 to 0.28. These lagged effects suggest that failing to stay abstinent during the trial causes increased stress. The lag of 1 for the cross-lagged effects translates to 2 weeks, except for the second time point which is 1 week after the first and the eighth time point which is 4 weeks after the seventh.

3.2 Binary outcome: Contemporaneous and reciprocal modeling alternatives

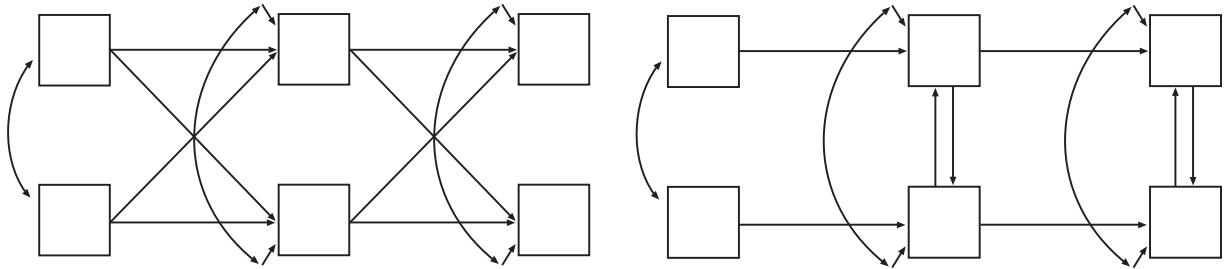
Conclusions drawn from the CLPM and RI-CLPM models have been challenged in Muthén and Asparouhov (2023). They pointed out that there are several competing models that are equivalent or nearly equivalent in terms of model fit but have different interpretation. They argued for also examining models that allow contemporaneous (lag 0) effects instead of or in addition to cross-lagged effects. While this challenge was made in the context of continuous outcomes, the same principles hold also with categorical outcomes. Figure 9 displays four key models for $T = 3$ shown as equivalent in Muthén and Asparouhov (2023). For simplicity, no random intercepts are included. Model (a) is the regular CLPM, models (b) and (c) use reciprocal, lag 0 effects but differ in whether cross-lagged effects are included. Model (d) has a single-direction lag 0 effect and cross-lagged effects. Estimates of reciprocal effects in models (b) and (c) often find a significant lag 0 effect in only one direction, thereby giving support for model (d).

The reciprocal model (b) is referred to as RI-RLPM (random intercept reciprocal lagged panel model) and the reciprocal model (c) is referred to as RI-RCLPM (random intercept reciprocal cross-lagged panel model). Bayes estimation is not available in Mplus for the reciprocal models but they can be estimated using WLSMV. As pointed

Table 8: Results for bivariate analysis of stress and abstinence ($N = 1375$, $T = 8$)

Model	# par's	PPP/ χ^2	# Sig. Residuals		Comment
			Resp Pattern	Bivar	
1. CLPM1, Bayes	60	0.000	1	19	Poor fit
2. CLPM2, Bayes	72	0.000	0	0	Poor fit
3. RI-CLPM1, Bayes	63	0.016	2	6	Poor fit
4. RI-CLPM2, Bayes	75	0.283	0	0	Good fit
RI-CLPM2, WLSMV	75	$\chi^2(69)=83$ ($p=.1218$)	0	0	Good fit
5. RI-RCLPM, WLSMV	70	$\chi^2(74)=84$ ($p=.1899$)	0	0	Good fit
6. RI-RLPM, WLSMV	63	$\chi^2(81)=87$ ($p=.2999$)	0	0	Good fit
7. Single-direction lag 0, WLSMV	69	$\chi^2(75)=84$ ($p=.2143$)	0	0	Good fit
8. Single-direction lag 0, WLSMV No cross-lagged effects	55	$\chi^2(89)=87$ ($p=.5519$)	0	0	Good fit

Figure 9: Four equivalent panel models for $T = 3$



(a) CLPM, lag1 cross-lags

(b) Reciprocal lag0, no cross-lags

(c) Reciprocal lag0, lag1 cross-lags,
no residual covariances

(d) Single-direction lag0, lag1 cross-lags,
no residual covariances

out in Muthén and Asparouhov (2023), for these models it is important to allow for non-symmetric confidence intervals which can be obtained in the WLSMV context using bootstrapping. The RI-RCLPM also needs to apply parameter constraints to obtain admissible parameter estimates. The analysis uses a restriction applied to the reciprocal effects held time invariant as described in Muthén and Asparouhov (2023) to avoid dual solutions and negative R-square (this is referred to as restriction a). The results are presented as model 5 in Table 8, showing that the model fits well and is more parsimonious than the RI-CLPM. Model 6 is the RI-RLPM which does not include cross-lagged effects (model type (b)) and uses time-invariant lag 0 effects. This model also fits well and is more parsimonious than model 5. Using bootstrapping, model 6 shows insignificant lag 0 effects of stress on non-abstinence but significant effects of non-abstinence on stress with standardized estimates ranging from 0.18 to 0.25 (the standardized effects vary despite invariant lag 0 effects due to varying latent response variable variances). Based on this finding, model 7 uses the single-direction, time-invariant lag 0 model (model type (d)) which also fits well. The model 7 lag 0 effect of non-abstinence on stress has standardized effects ranging from 0.22 to 0.29. The single-direction lag 0 model which instead uses the reverse effect of stress on non-abstinence fits the same and has a significant lag 0 effect but the standardized effects are much smaller, ranging from 0.06 to 0.08 (not shown). Model 7 does not have any significant cross-lagged effects and they are excluded in model 8 which is the most parsimonious of the eight models. The model 8 lag 0 effects of non-abstinence on stress have standardized effects ranging from 0.18 to 0.24. Freeing the first residual variance for the abstinence variable as discussed in Section 2.1 gave a larger standard error, did not result in an improvement in fit, and did not change the magnitude of the standardized lag 0 effect estimates.

The set of analyses in Table 8 indicate that there is an effect of non-abstinence on stress rather than the other way around. In line with the conclusions of Muthén and Asparouhov (2023), the time lag for the effect is, however, difficult to establish. The more traditional model 4 states that the effect has lag 1 which is mostly a time interval of two weeks in these data. In contrast, model 8 states that the effect is contemporaneous. Although model 8 is more parsimonious, both model 4 and model 8 fit the data well. There is not a strong statistical argument for choosing between the models. The models are not nested due to the differences of including residual covariances or not and including lag 0 effects or not. Therefore chi-square difference testing cannot be applied. It may be disappointing to not be able to determine the lag of the effect, but that is the nature of the design of data collection. More importantly, the direction of the effect has been determined.

4 Ordinal outcome

This section turns to the case of ordinal outcomes such as for the alcohol risk outcome in Figure 8. The unrestricted probit model is often rejected for ordinal outcomes when there is a strong floor effect. When applying the unrestricted probit model to the alcohol risk variable, five response patterns have significant standardized residuals with an especially strong misestimation of the most common response pattern of abstinence at all eight time points where the observed frequency 312 obtains the estimate 267. The total number of bivariate cells is $25 \times 8(8 - 1)/2 = 700$ of which 273 cells show

significant standardized residuals which is far greater than the 35 suggested by the 5% threshold. It is clear that an alternative model is needed for this variable.

The bivariate probit model for two ordinal variables was shown in Figure 3. With C categories, the number of parameters in the model is $C1-1+C2-1+1$ (2 sets of thresholds + 1 polychoric correlation) = $C1+C2-1$ parameters. For $C1=C2=5$, this adds to 9 parameters. The unrestricted multinomial model for the two variables has $C1*C2-1$ parameters which for $C=5$ adds to 24 parameters. This is the model that is tested against in the bivariate frequency tests and where model misfit is often found. However, intermediate models are possible. This paper uses a model which has $C1+C2$ parameters, that is, adding one parameter to the unrestricted probit model for two variables. For $C=5$, it has 10 parameters. Despite adding only a single parameter, this model fits the data considerably better. The model will be referred to as the two-part ordinal model.

4.1 Two-part ordinal model

The two-part ordinal model is inspired by two-part regression modeling of semicontinuous outcomes proposed by Duan et al. (1983) and two-part growth modeling with semicontinuous outcomes in Olsen and Schafer (2001); see also two-part growth mixture modeling in Muthén (2001). For ordinal outcomes, the model draws on the two-part regression analysis with an ordinal outcome that was used in Muthén, Muthén and Asparouhov (2016). The model is suitable for outcomes that have a strong floor effect as seen for the alcohol risk variable. The idea of the model is shown in Figure 10 for the case of a 4-wave growth model with random intercept and random slope growth factors. A 5-category ordinal variable is split into an ordinal part p (positive categories) for individuals who are above the floor value and a binary part b defined by being at the floor value ($b=0$) or above it ($b=1$). The ordinal outcome is missing when the binary outcome is zero. A strength of the two-part model is that the two parts can have different relations to covariates as indicated by the x variable in the figure. For example, a treatment dummy variable can have different influence on the two parts. Each part uses a probit model with continuous latent response variables specified for the binary variable and the ordinal variable. For the ordinal part, there are $C-2$ thresholds and for the binary part there is one.

The two-part model of Figure 10 corresponds to a single outcome, where the ordinal and binary parts are correlated only via their random intercepts. The univariate two-part ordinal model with random intercepts can be expressed as follows for the binary part b and ordinal part p in terms of the corresponding latent response variables,

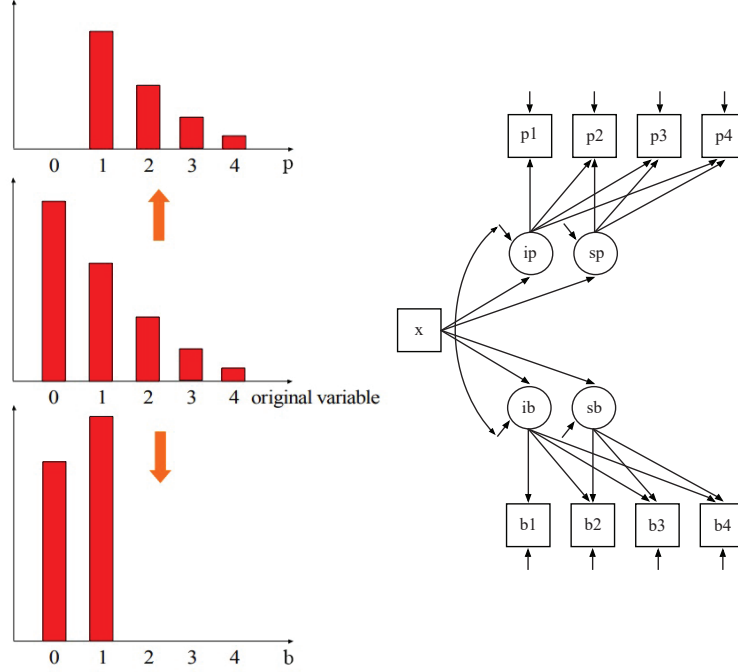
$$b_{it}^* = \alpha_{bi} + \epsilon_{bit}, \quad (7)$$

$$p_{it}^* = \alpha_{pi} + \epsilon_{pit}, \quad (8)$$

where $\alpha_{bi} \sim (N, 0, \psi_b)$ and $\alpha_{pi} \sim (N, 0, \psi_p)$ denote random intercepts that are correlated, and the normally distributed ϵ residuals typically have auto-regressions as before in Equation (4). Here, b_{it}^* has a single threshold τ_{bt} for each timepoint t while for an outcome with C categories, p_{it}^* has $C - 2$ thresholds for each time point t , where for the ordinal variable p_{it} observed in category c ,

$$p_{it} = c \Leftrightarrow \tau_{c-1,t} \leq p_{it}^* < \tau_{c,t}. \quad (9)$$

Figure 10: Two-part model



As before, however, a first step is to test the fit of the unrestricted multivariate normal probit model. For the unrestricted two-part ordinal model, there are no random intercepts and correlations are instead considered between the latent response variables for the two parts themselves. The exception is correlation between the two parts at the same time point which are not well determined given that the ordinal part is never observed when the binary part is zero; these concurrent correlations are fixed at zero in the estimation of the unrestricted model.

For a bivariate RI-CLPM of ordinal outcomes, there can in principle be four model parts, an ordinal and a binary for each of the two outcomes. The four processes would be correlated via their random intercepts only. In the current application, a two-part model is considered for the alcohol outcome but is not needed for the continuous stress outcome. This model therefore has three correlated random intercepts.

The two-part ordinal model is estimated by Bayes which properly handles the missing data on the ordinal outcome when the binary outcome is zero. ML would also handle the missing data correctly, but as before, ML involves numerical integration with too many dimensions to be practical. Because WLSMV is a limited-information estimator, it cannot handle this special missing data situation properly.

5 Ordinal outcome example

The univariate analysis of the $T = 8$ alcohol risk outcomes involves eight ordinal and eight binary variables when applying the two-part ordinal model. Table 9 shows univariate analysis results using both the regular ordinal probit model and the two-

Table 9: Univariate analysis of 5-category alcohol risk using regular and two-part ordinal models (N=1375, T=8)

Model	# par's	PPP/ χ^2	# Significant Resp Pattern*	Residuals Bivar	Comment
Regular ordinal probit					
1. Unrestricted	60	0.498	5 (312)	273 (39%)	Poor fit
2. AR2	45	0.151	5 (312)	277	Poor fit
3. RI-AR2	46	0.135	5 (312)	274	Poor fit
Two-part ordinal probit					
4. Unrestricted	144	0.472	1 (12)	29 (4%)	Good fit
5. AR2	58	0.145	2 (46, 13)	106 (15%)	Poor fit
6. RI-AR2	61	0.228	1 (12)	52 (7%)	OK fit

* Observed frequency in parentheses.

part ordinal probit model. Model 1 is the unrestricted regular ordinal probit model. This model is not appropriate for the data in that it has strong indications of misfit with a significant standardized residual for 5 response patterns including the most frequent response pattern of 312 with individuals being in the abstinence category at all time points. Model 1 also has 273 significant standardized residuals for the bivariate frequency tables which is 39% of all 700 bivariate cells.⁵ The more restricted models 2 and 3 are consequently also misfitting. These results show the need for the two-part ordinal model.

Turning to the two-part ordinal probit models, model 4 is the unrestricted model and, unlike model 1, model 4 shows a good fit, making it possible to proceed with the more restrictive models 5 and 6. Model 5 is the lag 2 model without random intercepts for the two processes and it fits poorly. Adding the random intercepts of model 6, the three extra variance-covariance parameters produce a much better fit. Model 6 has sizeable variances for the two random intercepts. The random intercept R-squares range from 0.3 to 0.6 with slightly higher values for the binary part. In contrast, model 3 has ignorable random intercept variance and R-squares about 0.04.

The bivariate analysis of stress and alcohol risk with a two-part ordinal representation involves eight continuous, eight ordinal, and eight binary variables. The RI-AR2 model is used for both stress and alcohol risk. The three random intercepts are correlated and lag 1 cross-lagged effects are allowed among all three sets of variables. Concurrent residual correlations are allowed between the continuous variables on the one hand and the ordinal and binary variables on the other hand. Concurrent residual correlations are not included between the ordinal and binary variables due to lack of

⁵The bivariate frequency testing of the two-part ordinal model using TECH10 was introduced in Mplus in Version 8.8.

information since the ordinal variable is missing for a binary variable of zero.⁶ This RI-CLPM has 137 parameters. It obtains a good fit with $PPP = 0.181$, 1 significant standardized residual for the response pattern with frequency 12, and 52 (7%) significant standardized residuals for the bivariate frequency tables.

The RI-CLPM with a two-part ordinal representation of alcohol risk has 5 out of 7 significant cross-lagged effects for the binary part influencing stress and 6 out of 7 significant cross-lagged effects for the ordinal part influencing stress. The standardized effects range from 0.08 to 0.22 for the binary part and 0.16 to 0.29 for the ordinal part. For the cross-lagged effects of stress influencing the two parts of alcohol risk, 1 out of 7 effects are significant for the binary part and zero for the ordinal part. The conclusion is that increased alcohol risk has a lagged positive effect on stress, with a larger effect for the ordinal part than the binary part. There is almost no evidence of lagged effects from stress to alcohol risk.

Reciprocal (lag 0) effect modeling using the two-part ordinal model can currently not be estimated using Bayes (WLSMV cannot handle two-part modeling as mentioned earlier). Single-direction lag 0 two-part ordinal modeling is, however, possible using Bayes. The lag0 effects are held time invariant. This model has 123 parameters and is therefore more parsimonious than the two-part ordinal RI-CLPM with 137 parameters. As expected, the model fits about the same in the two directions and has similar fit as the RI-CLPM. The effects from alcohol risk to stress are larger than from stress to alcohol risk for both the binary and ordinal parts. In standardized terms, the effects of binary risk on stress range from 0.18 to 0.22 and the effects from ordinal risk to stress range from 0.27 to 0.39. The effects from stress to binary risk range from 0.12 to 0.15 and the effects from stress to ordinal risk range from 0.14 to 0.19. These analyses do not give a clearcut conclusion of the direction of influence between the two outcomes but the effects are stronger in the alcohol to stress direction. As in the binary case, it is not possible to determine if the effects have lag 1 or lag 0.

6 Extensions

Multiple-group analysis makes it possible to study group differences in parameters of all the models discussed. Because group membership is typically time invariant, the influence of groups on the time-invariant random intercepts is of key interest, particularly group differences in their means. For instance, the zero means of the random intercepts in the two-part ordinal model of Equations (7) and (8) can instead be estimated while holding thresholds invariant across the groups. By fixing the random intercept means to zero for one group, the thresholds of that group are identified by that group's proportions, and this in turn identifies the random intercept means for the other groups through their observed proportions. Group differences in other parameters may also be studied such as cross-lagged or contemporaneous effects.

The COMBINE example has a special interest in group differences in the random intercept means of the alcohol risk variable. This double-blind randomized clinical trial of alcohol use disorder treatment has nine groups, one placebo group and eight groups with different combinations of medication and therapy. Each group consists of approximately 150 subjects.

⁶Mplus uses the Gibbs random walk algorithm for this Bayesian analysis.

Multiple-group two-part ordinal analysis of the nine groups adds the $(9 - 1)3 = 24$ parameters of the means of the three random intercepts to the previous section's bivariate two-part ordinal models for stress and alcohol risk. The random intercept means for the placebo group are fixed at zero as a comparison. The analyses use both the RI-CLPM version and the single-direction lag 0 models of the previous section. The models are estimated by Bayes. The RI-CLPM version uses 161 parameters and obtains $PPP = 0.296$ which is an improvement over the previous analysis with $PPP = 0.181$. The maximum number of significant bivariate cells for any of the nine groups is only 6 with a total of 23. The single-direction lag 0 model has 147 parameters with time-invariant lag 0 effects. As before, similar fit for lag 0 regression in both directions are obtained and the fits are similar to that of the RI-CLPM.

Because all three models fit well, the choice between the three models is not expected to matter in terms of estimating the treatment effects. The results show that neither the choice between cross-lagged versus lag 0 effect modeling, nor the choice of the direction of the lag 0 effect matters in terms of which groups have random intercept means significantly different from the zero value of the placebo group. Here, a negative mean represents a beneficial treatment outcome. No group has significant random intercept means for the stress outcome. Four groups have significant negative means for the random intercept of the binary part. Two of these four groups also have significant negative means for the random intercept of the ordinal part. The four groups with significant effects for abstinence are:

- Naltrexone
- Naltrexone + acamprosate
- Placebo + behavioral intervention
- Naltrexone + acamprosate + behavioral intervention

These treatments were also found to be the most effective in the analyses of the COMBINE study. The second and third treatment listed were the ones found to also decrease the risk of a higher degree of alcohol risk and can therefore be said to be the most successful treatments.

It is also possible to allow group differences in other model parameters such as the effect of alcohol on stress. The groups may be estimated with these parameters unconstrained and then tested for group invariance. A Wald chi-square test of invariance can be carried out based on Bayes estimates as described in Asparouhov and Muthén (2021b). Using the single-direction model with time-invariant lag 0 effect for both the binary and ordinal alcohol risk parts on stress, group invariance was rejected on the 5% level with $\chi^2(16) = 28.37$ ($p = 0.0285$).

A further extension is to add a growth model. For example, linear growth for the two-part ordinal model is shown in Figure 10 where random slopes sp and sb are added to the random intercepts. In the bivariate analysis of stress and alcohol risk using the two-part ordinal model for alcohol risk, there would be three growth models. Auto-regressions among the three variables can be added. A multiple-group version of this model would allow the means of the random slopes to also vary across groups. In a model with no random slope, the random intercept refers to the level at all time points, in this case during the eight time points following the start of treatment. Including a random slope, a parameterization can be chosen so that the random intercept refers to the status at the last time point and treatment effects evaluated for those random

intercept means. Analysis using this model did not change the above findings of which treatments had significant effects.

7 Conclusions

This paper presented modeling, testing, identification, and estimation for the case of binary and ordinal variables in cross-lagged panel modeling. Simulations showed that estimation with both Bayes and weighted least squares methods worked well given a sufficient sample size and a sufficient number of time points. A larger sample size and more time points are required than for continuous variables. A two-part ordinal model was proposed for ordinal variables with strong floor effects. Using a randomized study of alcohol treatment, the methods were used to examine the interaction between stress and alcohol use. Extensions to multiple-group analysis and modeling in the presence of trends were discussed.

The substantive results of the current study are consistent with preclinical data indicating that alcohol consumption increases subsequent stress, and that stress does not strongly predict lagged alcohol consumption. The time scale of weeks may be a limitation, given that some intensive longitudinal studies with heavy social drinkers have shown within day and day-to-day effects of stress on alcohol consumption (Armeli et al., 2000; Wemm et al., 2022). Future research should test the association between stress and alcohol use among individuals with alcohol use disorder using intensive longitudinal data collected during treatment.

Categorical variables also appear in panel studies as factor indicators. Random effects modeling of such a multiple-indicator case was studied in Muthén (1983, 1984). However, in the multiple-indicator case, the CLPM and RI-CLPM modeling still considers a continuous outcome, namely the factor, so that CLPM and RI-CLPM analysis can draw on the continuous-variable modeling of Mulder and Hamaker (2020).

Further methods for the joint longitudinal analysis of several categorical variables are discussed in Muthén and Asparouhov (2022). They include bivariate latent transition analysis and analysis with distal outcomes. With intensive longitudinal data where there are many time points spaced closely in time, dynamic structural equation modeling (DSEM; Asparouhov et al., 2018; Hamaker et al., 2023) is available for categorical outcomes. DSEM analyses are intended for data with $T \geq 20$ and are not suitable for the small number of time points in panel data that is considered here.

References

- [1] Anton, R.F. and others (2006). Combined pharamacotherapies and behavioral interventions for alcohol dependence. The COMBINE study: A randomized controlled trial. *Journal of the American Medical Association*, 295, No. 17, 2003 - 2017.
- [2] Armeli, S., Carney, M.A., Tennen, H., Affleck, G. & O'Neill, T.P. (2000). Stress and alcohol use: A daily process examination of the tressor-vulnerability model. *Journal of Personality and Social Psychology*.
- [3] Asparouhov, T., Hamaker, E.L. & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25:3, 359-388, DOI: 10.1080/10705511.2017.1406803
- [4] Asparouhov, T. & Muthén, B. (2021a). Bayesian analysis of latent variable models using Mplus. Version 5, September 18, 2021. <https://www.statmodel.com/download/BayesAdvantages18.pdf>
- [5] Asparouhov, T. & Muthén, B. (2021b). Advances in Bayesian model fit evaluation for structural equation models, *Structural Equation Modeling: A Multidisciplinary Journal*, 28:1, 1-14, DOI: 10.1080/10705511.2020.1764360
- [6] Asparouhov, T. & Muthén, B. (2022). Assessing model fit for SEM models with categorical variables via contingency tables. Technical Report. <https://www.statmodel.com/download/Tech10.pdf>
- [7] Asparouhov, T. & Muthén, B. (2023). Residual structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 30, 1-31. DOI: 10.1080/10705511.2022.2074422
- [8] Becker, H.C. (2017). Influence of stress associated with chronic alcohol exposure on drinking. *Neuropharmacology*, Aug 1;122:115-126. doi: 10.1016/j.neuropharm.2017.04.028. Epub 2017 Apr 19.
- [9] Chi, E.M. & Reinsel, G.C. (1989). Model for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association*, 84, 406, 452-459.
- [10] Duan, N., Manning, W.G. Morris, C.N. & Newhouse, J.P. (1983). A comparison of alternative models for the demand for medical care. *Journal of Business & Economic Statistics*, 2, 283-289.
- [11] Hamaker, E. L., Kuiper, R. M., Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20, 102-116.
- [12] Hamaker, E.L., Asparouhov, T. & Muthén, B. (2023). Dynamic structural equation modeling as a combination of time series modeling, multilevel modeling, and structural equation modeling. Chapter 31, pp. 576-596 in: *The Handbook of Structural Equation Modeling* (2nd edition); Rick H. Hoyle (Ed.); Publisher: Guilford Press.
- [13] Hamaker, E. L. (2023). The within-between dispute in cross-lagged panel research and how to move forward. Forthcoming in *Psychological Methods*.
- [14] Ialongo, N. (2022). Personal communication.
- [15] McHugh et al. (2013). Positive affect and stress reactivity in alcohol-dependent outpatients. *J. Studies in Alcohol and Drugs*.

- [16] Mulder, J.D. & Hamaker, E.L. (2020). Three extensions of the Random Intercept Cross-Lagged Panel Model. *Structural Equation Modeling: A Multidisciplinary Journal*, DOI: 10.1080/10705511.2020.1784738
- [17] Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, 22, 48-65.
- [18] Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115-132.
- [19] Muthén, B. (1993). Goodness of fit with categorical and other non-normal variables. In K.A. Bollen, & J.S. Long (Eds), *Testing Structural Equation Models* (pp. 205-243). Newbury Park, CA: Sage. http://www.statmodel.com/bmuthen/articles/Article_045.pdf
- [20] Muthén, B. (1996). Growth modeling with binary responses. In A.V. Eye & C. Clogg (eds.). *Categorical variables in developmental research: Methods of analysis* (pp. 37-54). San Diego, CA: Academic Press. http://www.statmodel.com/bmuthen/articles/Article_064.pdf
- [21] Muthén, B., du Toit, S.H.C., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Unpublished technical report. http://www.statmodel.com/download/Article_075.pdf
- [22] Muthén, B. (2001). Two-part growth mixture modeling. Unpublished report. <http://www.statmodel.com/download/2PGMM.pdf>
- [23] Muthén, B. & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus Web Note No. 4*. <https://www.statmodel.com/download/webnotes/CatMGLong.pdf>
- [24] Muthén, B., Muthén, L, & Asparouhov, T. (2016). Regression and mediation analysis using Mplus. Los Angeles, CA: Muthén & Muthén.
- [25] Muthén, B. & Asparouhov, T. (2022). Using Mplus to do cross-lagged modeling of panel data, part 2: Categorical variables. *Mplus Web Talk No. 4, Part 2*. <http://www.statmodel.com/Webtalk4P2.shtml>
- [26] Olsen, M.K. & Schafer, J.L. (2001). A two-part random effects model for semi-continuous longitudinal data. *Journal of the American Statistical Association*, 96, 730-745.
- [27] Sinha, R. (2022). Alcohol's negative emotional side: The role of stress neurobiology in alcohol use disorder. *Alcohol Research. Current Reviews*. Volume 42, issue 127.
- [28] Wemm, S. E., Tennen, H., Sinha, R., & Seo, D. (2022). Daily stress predicts later drinking initiation via craving in heavier social drinkers: A prospective in-field daily diary study. *Journal of Psychopathology and Clinical Science*, 131(7), 780–792. <https://doi.org/10.1037/abn0000771>

8 Appendix

Mplus input for key analyses.

Table 1: Univariate analysis of abstinence using unrestricted model

	USEVARIABLES = z1-z8; CATEGORICAL = z1-z8;
DEFINE:	CUT z1-z8 (0.5);
ANALYSIS:	ESTIMATOR = BAYES; BITERATIONS = (5000); THIN = 10; PROCESSORS = 8;
MODEL:	z1-z8 WITH z1-z8;
OUTPUT:	STANDARDIZED RESIDUAL TECH8 TECH10;
PLOT:	TYPE = PLOT3;

Table 2: Univariate analysis of abstinence using RI-AR2

	USEVARIABLES = z1-z8; CATEGORICAL = z1-z8;
DEFINE:	CUT z1-z8 (0.5);
ANALYSIS:	ESTIMATOR = BAYES; BITERATIONS = (5000); THIN = 10; PROCESSORS = 8;
MODEL:	i BY z1-z8@1; z2^-z8^ PON z1^-z7^; z3^-z8^ PON z1^-z6^;
OUTPUT:	STANDARDIZED RESIDUAL TECH8 TECH10;
PLOT:	TYPE = PLOT3;

Table 3: Bivariate analysis of stress and abstinence using RI-CLPM2

```

USEVARIABLES = y2-y9 z2-z9; ! y is stress, z is alcohol risk
CATEGORICAL = z2-z9;

DEFINE:      CUT z2-z9(0.5);

ANALYSIS:    ESTIMATOR = BAYES;
              BITERATIONS = (5000);
              PROCESSORS = 8;
              THIN = 10;

MODEL:       iz BY z2-z9@1;
              z3^-z9^ PON on z2^-z8^;
              z4^-z9^ PON z2^-z7^;
              iy BY y2-y9@1;
              y3^-y9^ PON y2^-y8^;
              y4^-y9^ PON y2^-y7^;

              y3^-y9^ PON z2^-z8^;
              z3^-z9^ PON y2^-y8^;

              y2^-y9^ PWITH z2^-z9^;

OUTPUT:      STANDARDIZED RESIDUAL TECH8 TECH10;

PLOT:        TYPE = PLOT3;

```


Table 4: Bivariate analysis of stress and abstinence using RI-RCLPM

```

USEVARIABLES = y2-y9 z2-z9; ! y is stress, z is alcohol risk
CATEGORICAL = z2-z9;

DEFINE:      CUT z2-z9(0.5);

ANALYSIS:    ESTIMATOR = WLSMV;
              PARAMETERIZATION = THETA;
              STARTS = 20;
              BOOTSTRAP = 500;

MODEL:       iz BY z2-z9@1;
              z3^-z9^ PON z2^-z8^;
              z4^-z9^ PON z2^-z7^;
              iy by y2-y9@1;
              y3^-y9^ PON y2^-y8^;
              y4^-y9^ PON y2^-y7^;

              y3^-y9^ PON z2^-z8^;
              z3^-z9^ PON y2^-y8^;

              y3^-y9^ PON z3^-z9^ (ryz);
              z3^-z9^ PON y3^-y9^ (rzy);

              y2^ WITH z2^;

MODEL CONSTRAINT: ! Restriction (a):
                  0<ryz*rzy;
                  0<1-ryz*rzy;

OUTPUT:      STANDARDIZED RESIDUAL TECH1 TECH10;

              CINTERVAL(BOOTSTRAP);

PLOT:        TYPE = PLOT3;

```

Table 5: Bivariate analysis of stress and abstinence using RI-RLPM

```

USEVARIABLES = y2-y9 z2-z9; ! y is stress, z is alcohol risk
CATEGORICAL = z2-z9;

DEFINE:      CUT z2-z9(0.5);

ANALYSIS:    ESTIMATOR = WLSMV;
              PARAMETERIZATION = THETA;
              STARTS = 20;
              BOOTSTRAP = 500;

MODEL:       iz BY z2-z9@1;
              z3^-z9^ PON z2^-z8^;
              z4^-z9^ PON z2^-z7^;
              iy BY y2-y9@1;
              y3^-y9^ PON y2^-y8^;
              y4^-y9^ PON y2^-y7^;

              ! y3^-y9^ PON z2^-z8^;
              ! z3^-z9^ PON y2^-y8^;

              y3^-y9^ PON z3^-z9^ (ryz);
              z3^-z9^ PON y3^-y9^ (rzy);

              y2^ WITH z2^;

              y3^-y9^ PWITH z3^-z9^;

OUTPUT:      STANDARDIZED RESIDUAL TECH1 TECH10;

              CINTERVAL(BOOTSTRAP);

PLOT:        TYPE = PLOT3;

```

Table 6: Bivariate analysis of stress and abstinence using single-direction lag 0

```

USEVARIABLES = y2-y9 z2-z9; ! y is stress, z is alcohol risk
CATEGORICAL = z2-z9;

DEFINE:      CUT z2-z9(0.5);

ANALYSIS:    ESTIMATOR = WLSMV;
              PARAMETERIZATION = THETA;
              STARTS = 20;

MODEL:       iz BY z2-z9@1;
              z3^-z9^ PON z2^-z8^;
              z4^-z9^ PON z2^-z7^;
              iy by y2-y9@1;
              y3^-y9^ PON y2^-y8^;
              y4^-y9^ PON y2^-y7^;

              y3^-y9^ PON z2^-z8^;
              z3^-z9^ PON y2^-y8^;

              ! y2^-y9^ PWITH z2^-z9^;

              y2^ WITH z2^;
              y3^-y9^ PON z3^-z9^ (lag0);

OUTPUT:      STANDARDIZED RESIDUAL TECH1 TECH10;

PLOT:        TYPE = PLOT3;

```

Table 7: Univariate analysis of alcohol risk using unrestricted regular ordinal probit

```

USEVARIABLES = z1-z8;
CATEGORICAL = z1-z8;

ANALYSIS:    ESTIMATOR = BAYES;
              BITERATIONS = (5000);
              THIN = 10;
              PROCESSORS = 8;

MODEL:       z1-z8 WITH z1-z8;

OUTPUT:      STANDARDIZED RESIDUAL TECH8 TECH10;

PLOT:        TYPE = PLOT3;

```

Table 8: Univariate analysis of alcohol risk using regular ordinal probit RI-AR2

	USEVARIABLES = z1-z8; CATEGORICAL = z1-z8;
ANALYSIS:	ESTIMATOR = BAYES; BITERATIONS = (5000); THIN = 10; PROCESSORS = 8;
MODEL:	i BY z1-z8@1; z2^-z8^ PON z1^-z7^; z3^-z8^ PON z1^-z6^;
OUTPUT: PLOT:	STANDARDIZED RESIDUAL TECH8 TECH10; TYPE = PLOT3;

Table 9: Univariate analysis of alcohol risk using unrestricted two-part ordinal

	USEVARIABLES = u1-u8 p1-p8; CATEGORICAL = u1-u8 p1-p8;
DATA TWOPART:	NAMES = z1-z8; BINARY = u1-u8; CONTINUOUS = p1-p8; CUTPOINT = 0; TRANSFORM = NONE;
ANALYSIS:	ESTIMATOR = BAYES; BITERATIONS = (10000); THIN = 10; PROCESSORS = 8;
MODEL:	u1-u8 WITH u1-u8; p1-p8 WITH p1-p8; p1^-p8^ ON u1^-u8^; p1^-p8^ PON u1^-u8^@0;
OUTPUT:	STANDARDIZED RESIDUAL TECH8 TECH10;
PLOT:	TYPE = PLOT3;

Table 10: Univariate analysis of alcohol risk using two-part ordinal RI-AR2

	USEVARIABLES = u1-u8 p1-p8; CATEGORICAL = u1-u8 p1-p8;
DATA TWOPART:	NAMES = z1-z8; BINARY = u1-u8; CONTINUOUS = p1-p8; CUTPOINT = 0; TRANSFORM = NONE;
ANALYSIS:	ESTIMATOR = BAYES; BITERATIONS = (10000); THIN = 10; PROCESSORS = 8;
MODEL:	ib BY u1-u8@1; ip BY p1-p8@1; u2^-u8^ PON u1^-u7^; u3^-u8^ PON u1^-u6^; p2^-p8^ PON p1^-p7^; p3^-p8^ PON p1^-p6^; ! u1-u5 PWITH p1-p5;
OUTPUT:	STANDARDIZED RESIDUAL TECH8 TECH10;
PLOT:	TYPE = PLOT3;

Table 11: Bivariate analysis of stress and alcohol risk using two-part ordinal RI-CLPM

```

USEVARIABLES = y1-y8 u1-u8 p1-p8;
CATEGORICAL = u1-u8 p1-p8;

DATA TWOPART:  NAMES = z1-z8;
                BINARY = u1-u8;
                CONTINUOUS = p1-p8;
                CUTPOINT = 0;
                TRANSFORM = none;

ANALYSIS:      ESTIMATOR = BAYES;
                BITERATIONS = (10000);
                ! FBITERATIONS = 200;
                THIN = 10;
                PROCESSORS = 8;
                ALGORITHM = GIBBS(RW);

MODEL:         iy BY y1-y8@1;
                ib BY u1-u8@1;
                ip BY p1-p8@1;

                ! univariates:
                y2^-y8^ PON y1^-y7^;
                y3^-y8^ PON y1^-y6^;

                u2^-u8^ PON u1^-u7^;
                u3^-u8^ PON u1^-u6^;

                p2^-p8^ PON p1^-p7^;
                p3^-p8^ PON p1^-p6^;

                ! bivariate:
                y2^-y8^ PON u1^-u7^;
                y2^-y8^ PON p1^-p7^;

                u2^-u8^ PON y1^-y7^;
                p2^-p8^ PON y1^-y7^;

                ! covariances:
                y1^-y8^ PWITH u1^-u8^;
                y1^-y8^ PWITH p1^-p8^;

OUTPUT:        STANDARDIZED RESIDUAL TECH8 TECH10;
PLOT:          TYPE = PLOT3;

```

Table 12: Bivariate analysis of stress and alcohol risk treatment effects using two-part ordinal RI-CLPM

```

USEVARIABLES = y1-y8 u1-u8 p1-p8;
CATEGORICAL = u1-u8 p1-p8;
CLASSES = c(9);
KNOWNCLASS = c(cCell = 1-9);

DATA TWOPART:
  NAMES = z1-z8;
  BINARY = u1-u8;
  CONTINUOUS = p1-p8;
  CUTPOINT = 0;
  TRANSFORM = NONE;

ANALYSIS:
  TYPE = MIXTURE;
  ESTIMATOR = BAYES;
  BITERATIONS = (5000);
  THIN = 10;
  PROCESSORS = 8;
  ALGORITHM = GIBBS(RW);

MODEL:
  %OVERALL%
  iy BY y1-y8@1;
  ib BY u1-u8@1;
  ip BY p1-p8@1;

  ! univariates:
  y2^-y8^ PON y1^-y7^;
  y3^-y8^ PON y1^-y6^;
  u2^-u8^ PON u1^-u7^;
  u3^-u8^ PON u1^-u6^;
  p2^-p8^ PON p1^-p7^;
  p3^-p8^ PON p1^-p6^;

  ! bivariate:
  y2^-y8^ PON u1^-u7^;
  y2^-y8^ PON p1^-p7^;
  u2^-u8^ PON y1^-y7^;
  p2^-p8^ PON y1^-y7^;

  ! covariances:
  y1^-y8^ PWITH u1^-u8^;
  y1^-y8^ PWITH p1^-p8^;

  %c#1%
  [iy@0 ib@0 ip@0];
  %c#9%
  [iy ib ip];

OUTPUT:
  STANDARDIZED RESIDUAL TECH8 TECH10;

PLOT:
  TYPE = PLOT3;

```