

Residual Structural Equation Models

Tihomir Asparouhov and Bengt Muthén*

Mplus

November 1, 2021

*We are grateful to Ellen Hamaker for helpful comments on the earlier draft of this paper.

1 Introduction

The residual variables in a structural equation model (SEM) can be used to create a secondary structural model. This combination of the primary and the secondary structural models is what we call the Residual Structural Equation Model (RSEM). The RSEM model has been discussed previously in the context of the dynamic structural equation models (DSEM), see Asparouhov et al. (2018) and Asparouhov and Muthén (2020), and it is generally referred to as the residual dynamic structural equation model (RDSEM). The RSEM model has also been used in the context of longitudinal cross-lagged panel models and is the basis for the random intercept cross-lagged panel model (RI-CLPM), see Hamaker et al. (2015), and the latent curve model with structured residuals (LCM-SR), see Curran et al. (2013). In this article, we provide a formal definition for the RSEM model in the context of the standard single-level SEM model with continuous and categorical variables. We describe the implementation of this model in Mplus 8.7 and discuss ML/WLSMV/Bayes model estimation. We illustrate and motivate the use of the RSEM model with several examples. Among these examples are the RI-ARMA(random intercept auto-regressive moving average) model, RI-MEAR(random intercept measurement error auto-regressive) model for longitudinal panel data, and the LCA model with local dependence for ordered categorical variables, which previously have not been available. We also describe an expansion of the RDSEM model which now includes contemporaneous residual modeling in addition to the lagged residual modeling. Model testing for the RSEM model is discussed and a new Pearson posterior predictive p-value (PPP) is introduced that can be used in evaluating model fit for categorical data modeling with Bayesian estimation. The Pearson PPP is particularly useful in evaluating the model fit for Mixture models.

The article is structured as follows. In Section 2 we introduce the general formulation of the model and provide details on the model estimation with the different estimation methods. In Section 3 we provide multiple examples and Mplus simulation studies which include the Mplus input statements as well as output statements. We show how the Mplus hats language for modeling with residuals, previously available only for RDSEM models, is now used to simplify the RSEM model specifications. Section 4 concludes.

2 The general RSEM model

Let Y be a vector of continuous observed dependent variables, η be a vector of continuous latent variables and X be a vector of covariates. The basic SEM model is given by the following two equations

$$Y = \nu + \Lambda\eta + KX + \varepsilon \quad (1)$$

$$\eta = \alpha + B\eta + \Gamma X + \xi, \quad (2)$$

where ν and α are the intercept parameter vectors and Λ , K , B and Γ are the regression parameter matrices. The residual variables in this model are ε and ξ . Denote the vector of all residual variables by $R = (\varepsilon, \xi)$. The RSEM model is then given by the following equation

$$R = B_r R + \zeta, \quad (3)$$

where $\zeta \sim N(0, \Psi)$. The variance covariance matrix Ψ may take any pre-specified form, i.e., it can be diagonal or it can include covariance parameters among the secondary residuals ζ . The RSEM model allows us to explore structural models for the residual variables of a standard SEM model. The RSEM framework can be viewed as a generalization of the unified longitudinal framework discussed in Usami et al. (2019). To accommodate categorical observed variables in this model, we simply replace the categorical variable with the corresponding underlying continuous variables based on the probit link function as it is typically done in the Mplus framework. The threshold parameters that categorize the underlying variables into the observed categories are also included in the model. The underlying continuous variable is used in the above equations instead of the observed categorical variable.

Note that the RSEM model is equivalent to the standard SEM model given in equation (1-2) where the variance covariance matrix of the residual vector R is

$$((I - B_r)^{-1})^T \Psi (I - B_r)^{-1}. \quad (4)$$

Thus the RSEM model can be viewed as a methodology that provides structural form for the variance covariance matrix of the residuals.

Here we also define the extended RSEM model, which is slightly more general than the model defined in equations (1-3). In this model, equation (3) remains the same, while equations (1) and (2) are replaced by

$$Y = \nu + \Lambda\eta + KX + B_{1r}R + \varepsilon \quad (5)$$

$$\eta = \alpha + B\eta + \Gamma X + B_{2r}R + \xi. \quad (6)$$

In the extended RSEM model, the residual variables augment the very same primary SEM model that is used to define them. Currently, the extended RSEM model can be estimated with the ML and WLSMV estimators but not with the Bayes estimator.

The Mplus language that facilitates the structural modeling for the residuals is identical to what has been used for the RDSEM model. For an observed variables Y , the residual variable of Y is referred to as Y^\wedge (Y hat) in the Mplus model statement. Similarly, for a continuous latent variable η , the variable η^\wedge refers to the residual variable of η . The residual variables can be regressed on each other with the usual ON statement or can be correlated with the usual WITH statement.

2.1 Model estimation with the ML and the WLSMV estimators

The maximum-likelihood estimation of the RSEM model simply amounts to treating the residual variables R as an additional set of latent variables. The RSEM model is therefore a standard SEM model, where the loading parameters for the residuals are fixed to 1. For standard SEM models, the model formulation includes actual residual variables, which are now converted to structural latent variables. Therefore the conversion process from RSEM to SEM includes adding new residuals that are zero. Such addition is non-consequential for the ML estimator and is accomplished simply by fixing to 0 the variances of the add-on residuals. This estimation approach is illustrated in the estimation of the RI-CLPM model in Mulder and Hamaker (2021) supplementary materials. The new implementation in Mplus, however, greatly simplifies the model specification for these models as the residual latent variables are automatically created.

Note here that the ML estimator can be used only with continuous dependent variables. That is because equation (4) implies that the residuals are correlated. In the Mplus framework, the residuals of categorical variables can be directly correlated, i.e., not via a latent variable, only with the Bayes and the WLSMV estimators but not with the ML estimator.

The WLSMV estimation of the RSEM model mirrors that of the ML estimator. The residual variables are treated as latent variables. As in the ML estimation, add-on residuals are created with zero variances to replace the

residuals. This step, however, requires the use of the **theta parameterization**, see Muthén and Asparouhov (2002). That is because with the **theta parameterization**, we can set the add-on residual variances to 0. With the **delta parameterization**, the residual variance of a categorical variable is not a model parameter. It is a parameter that depends on all other model parameters and cannot be set directly to any value.

2.2 Model estimation with the Bayes estimator

The Bayes estimator of the RSEM model is somewhat more complicated. Here, it is not possible to use the estimation approach used with the ML and the WLSMV estimators. This is because add-on residuals can not have a fixed residual variance of 0. It is possible to fix the residual variance to a small positive value, such as 0.001, however, such an approach has several drawbacks. First, when using a small but positive add-on residual value, we create an approximate model rather than exact. Second, in practical applications it is difficult to select a good small value that works well for all variables in the model. Third, the Bayesian estimation via the MCMC method converges very slowly due to highly correlated model components updated at different steps in the MCMC. Therefore we have implemented direct Bayesian estimation that does not use add-on residuals. This new estimation method generally follows the estimation methodology described in Asparouhov and Muthén (2010) for standard SEM models. For standard SEM models, all structural parameters and intercepts are updated in the MCMC estimation with a normal conditional distribution, conditional on all other model parameters and latent variables. In the RSEM model that is not the case. The structural model parameters and intercepts in equations (1-2) are estimated in one step and the structural model parameters in equation (3) are estimated in a separate step. The residual variables R in the RSEM Bayesian estimation are not stochastically updated in the MCMC, i.e., they are not treated as latent variables that must be stochastically updated. They are derived quantities computed from all other model components: model parameters, observed and latent variables.

To update the parameters in equations (1-2), conditional on the parameters in (3), we simply reformulate the RSEM model as a SEM model with residual variance covariance given in equation (4). We then apply the approach in Asparouhov and Muthén (2010), which uses a conditional normal distribution (based on conjugate priors) for the updating these parameters.

To update the parameters in equation (3), conditional on the parameters in equations (1-2) and all other quantities, we first compute the residual variables R . These variables are then used for formulating a standard SEM model given in equation (3). Using the Asparouhov and Muthén (2010) method for this SEM model allows us to update the structural parameters in equation (3) as well as the variance covariance matrix Ψ . This approach yields fast, efficient and reliable Bayesian estimation for the RSEM model.

3 Examples

3.1 Growth modeling with autoregressive error structure

Suppose that Y_{it} is an observed variable for individual i at time $t = 1, \dots, T$. A linear growth model is described by the following equation

$$Y_{it} = I_i + S_i \cdot t + \varepsilon_{it}, \quad (7)$$

where I_i and S_i are the normally distributed random intercept and slope. The autoregressive structure for the residuals can be introduced in several different ways. One way to introduce that structure is as follows

$$\varepsilon_{it} \sim N(0, \theta) \quad (8)$$

$$Cor(\varepsilon_{it_1}, \varepsilon_{it_2}) = \rho^{|t_1 - t_2|} \quad (9)$$

where θ and ρ are model parameters. To estimate such a model in Mplus, one can use the **model constraint** command as in Muthén and Muthén (1998-2017) example 6.17. Such an approach is generally limited to at most 50 time points because the model is estimated in a wide format and the size of the manipulated matrices would become too large for larger values of T . The approach is mainly used with only a single variable observed across time. Autoregressive error structure for multivariate models would be difficult to estimate this way because equation (9) would involve matrix power computation which is impractical to implement with the model constraint command. In addition, this approach is typically used when the residual variances and the autoregressive parameters are time invariant. If either of these are not time-invariant the variance covariance matrix implied by the non-invariant versions of equations (8-9) becomes impractical to implement

with the model constraint command. Even with just 10 time points, if the autoregressive parameters and the residual variances are not time invariant, one will need to write out 55 different equations in model constraint.

A different modeling approach that is available in Mplus for growth models with autoregressive error structure is the DSEM and RDSEM modeling frameworks. Examples are discussed in Asparouhov et al. (2018) and Asparouhov and Muthén (2020). Equations (8-9) are replaced most typically by

$$\varepsilon_{it} = \rho_i \varepsilon_{i,t-1} + \zeta_{it} \quad (10)$$

where ρ_i and $\log(\text{Var}(\zeta_{it}))$ are subject-specific normally distributed random variables, see Muthén and Muthén (1998-2017) example 9.37. In the cross-classified DSEM framework, the random effects can also be time and subject specific, see Muthén and Muthén (1998-2017) example 9.39. Such cross-classified modeling for longitudinal data is also discussed in Asparouhov et al. (2018) and Asparouhov and Muthén (2016). These modeling approaches are typically used when the number of time points is larger, i.e., at least 10. However, depending on the model complexity, the number of time points that is required could be much higher. Because the models are estimated in long format, the number of time points that can be modeled is unlimited. In fact, the larger number of time points, the better the estimation is. Because the DSEM and RDSEM models are intended for use with larger values of T , the subject-specific and time-specific parameters are modeled as random effects rather than non-random model parameters. For small number of time points, $T < 10$, even without subject-specific or time-specific autoregressive parameters, the DSEM and RDSEM models tend to have worse performance than the wide-modeling approach due to the added level of uncertainty that comes about from the initial conditions. That refers to the fact that at time $t = 1$, the predictor variable $\varepsilon_{i,0}$ in (10) is an unmeasured latent variable. The influence of the initial condition tends to disappear as T increases, however, for $T < 10$ it can lead to biases in the estimates.

The RSEM model offers a different alternative to the above two approaches. It is a wide modeling approach that focuses on time-specific parameters. Equations (8-9) are replaced by

$$\varepsilon_{it} = \rho_t \varepsilon_{i,t-1} + \zeta_{it} \quad (11)$$

$$\zeta_{it} \sim N(0, \theta_t). \quad (12)$$

Equation (11) is used only for $t > 1$. The initial condition problem that exist with small T in the DSEM and RDSEM models is not present here. We are able to avoid this issue in the RSEM modeling framework because time-invariance is not needed or assumed. In the RSEM model, the parameters are cross-sectionally estimated and there are no subject-specific parameters in the auto-regressive error structure.

One advantage of the RSEM model over the wide-format model (8-9) is that it allows non-invariance of the parameters. Even when the parameters are invariant, however, the RSEM model is more compact and easy to use. Model (11-12) becomes equivalent to (8-9) if the auto-regressive parameters ρ_t are held equal across time, the residual variances parameters θ_t are held equal across time for $t > 1$, and the first time point residual variance is constrained as follows

$$\theta_1 = \theta_2 / (1 - \rho^2). \quad (13)$$

This last constraint must be added to the model constraint statement in the Mplus input file to obtain the equivalence between the two models.

An advantage of the RSEM model over the DSEM/RDSEM model is that the time non-invariance is more flexible. In RSEM, we can estimate partial invariance models where some of the time points have invariant parameters and others do not. Invariance and partial invariance of the autoregressive and residual variance parameters can be tested in Mplus using the **model test** command. Note, however, that there is a difference between the autocorrelation parameter $Cor(\varepsilon_{it}, \varepsilon_{i,t-1})$ and $\rho_t = Cor(\varepsilon_{it}, \varepsilon_{i,t-1}) \sqrt{Var(\varepsilon_{it}) / Var(\varepsilon_{i,t-1})}$. Testing the invariance of ρ_t is not the same as testing the invariance of the correlation parameter. To test the invariance of the correlation parameter, in the model test command, the autocorrelations must be expressed in terms of the model parameters.

Another advantage of the RSEM model over the DSEM/RDSEM model is that it performs better for small T . Furthermore, the RSEM model provides a chi-square test of fit and Bayesian posterior predictive P-value (PPP), based on the comparison of the model with the sample statistics of the multivariate vector Y_1, \dots, Y_T . This is not available in the long-format modeling approach used with the DSEM/RDSEM models.

The disadvantages of the RSEM model as compared to DSEM/RDSEM models is that the model doesn't allow for subject-specific parameters except for the random intercept and slope and that it is limited in terms of how large T can be. Furthermore, the time-varying RSEM model has a

much larger number of parameters as compared to the time-varying DSEM model where the time-specific variations in the auto-regressive parameters are modeled as random effects. Thus the time-varying DSEM model is much more parsimonious when compared to the RSEM model. Note, however, that the RDSEM model currently implemented in Mplus has time-invariant autoregressive structure. Only the DSEM model accommodates time-specific auto-regressive parameters. Since the linear growth model is actually an RDSEM model, rather than a DSEM model, the time-varying DSEM model must be transformed into a time-varying RDSEM model through a non-linear reparameterization. That transformation is described in Asparouhov et al. (2018) for linear and quadratic growth models. This complexity in the time-varying RDSEM estimation can be viewed as another advantage of the RSEM model which achieves this more directly.

All three of the above models are available for categorical data as well. The observed variable Y_{it} is simply replaced by Y_{it}^* . The RSEM model and the model given in equations (8-9) can be estimated with the WLSMV estimator. The Bayesian estimator can be used for all three models with continuous and categorical data.

Figure 1 shows the input file for a simulation study for the RSEM linear growth model for continuous data and with non-invariant autoregressive coefficients ρ_t . We use 7 time points in this simulation study and the times of observations are set to -1.5, -1, -0.5, 0, 0.5, 1, 1.5. Figure 2 shows the results of this simulation study for a subset of the model parameters. The results indicate that the Bayesian model estimation performs well. Figure 2 is a direct extract of the output file produced by Mplus and it allows us to quickly evaluate the quality of the estimation. The estimation is considered good when all of these are satisfied: the point estimates in column 2 show minimal bias as compared to the true values reported in column 1, the standard deviation of the point estimates reported in column 3 and the average standard errors reported in column 4 are relatively small and are comparable to each other (asymptotically the ratio of these two columns should be 1), the mean squared error of the point estimates reported in column 5 is minimal (i.e. close to 0), the coverage of the confidence or credibility intervals reported in column 6 is near to the nominal level of 95%, and the majority of the model parameters are statistically significant in most of the replications (the percentage of replications where the parameter is significant is reported in column 7, i.e., we want to see large values in that column on the percentage scale of 0 to 1).

Figure 1: Input file for RSEM AR(1) linear growth model

```
montecarlo:
    names = y1-y7;
    nobs = 1000;
    nreps = 100;

analysis: estimator=bayes; proc=2;

model population:
y1*1 y2-y7*0.75;
i s | y1@-1.5 y2@-1 y3@-0.5 y4@0 y5@0.5 y6@1 y7@1.5;
i*1 s*0.2; i with s*0.1; [i*2 s*0.3];
y2^-y4^ pon y1^-y3^*0.4;
y5^-y7^ pon y4^-y6^*0.5;

model:
y1*1 y2-y7*0.75 (v1-v7);
i s | y1@-1.5 y2@-1 y3@-0.5 y4@0 y5@0.5 y6@1 y7@1.5;
i*1 s*0.2; i with s*0.1; [i*2 s*0.3];
y2^-y4^ pon y1^-y3^*0.4;
y5^-y7^ pon y4^-y6^*0.5;
```

Figure 2: Output results for RSEM AR(1) linear growth model

MODEL RESULTS		Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Y2^	ON							
Y1^		0.400	0.3827	0.0649	0.0606	0.0045	0.930	1.000
Y3^	ON							
Y2^		0.400	0.4002	0.0420	0.0462	0.0017	0.970	1.000
Y5^	ON							
Y4^		0.500	0.5005	0.0464	0.0434	0.0021	0.950	1.000
I	WITH							
S		0.100	0.0961	0.0347	0.0332	0.0012	0.900	0.790
Means								
I		2.000	1.9960	0.0367	0.0357	0.0014	0.930	1.000
S		0.300	0.3047	0.0219	0.0200	0.0005	0.910	1.000
Variances								
I		1.000	1.0026	0.0704	0.0669	0.0049	0.960	1.000
S		0.200	0.2059	0.0336	0.0284	0.0012	0.890	1.000
Residual Variances								
Y1		1.000	0.9943	0.1320	0.1168	0.0173	0.900	1.000
Y2		0.750	0.7410	0.0508	0.0476	0.0026	0.910	1.000
Y3		0.750	0.7576	0.0403	0.0411	0.0017	0.960	1.000

3.2 RI-CLPM

Suppose that Y_1, \dots, Y_T are a set of outcomes for a variable Y observed at times $1, \dots, T$. We define the RI-AR (random intercept auto-regressive) model as follows

$$Y_t = \nu_t + \eta + \hat{Y}_t \quad (14)$$

$$\hat{Y}_t = \rho_t \hat{Y}_{t-1} + \varepsilon_t. \quad (15)$$

The latent variable η represents the time-invariant person-specific effect and is generally referred to as the random intercept. The parameters ν_t represent the time-specific mean across the population. The latent variable \hat{Y}_t represents the residual, i.e., the time-specific individual-specific deviation from the population mean ν_t and the time-invariant person-specific effect η . Equation (15) represents the auto-regressive part of the model.

The RI-CLPM (random intercept cross-lag panel model) is the bivariate version of the RI-AR model and is described as follows. Suppose that Z_1, \dots, Z_T is a second set of outcomes observed at times $1, \dots, T$. For $t = 1, \dots, T$

$$Y_t = \nu_{1,t} + \eta_Y + \hat{Y}_t \quad (16)$$

$$Z_t = \nu_{2,t} + \eta_Z + \hat{Z}_t. \quad (17)$$

The latent variables η_Y and η_Z represent the time-invariant person-specific effects, i.e., the random intercepts. The parameters $\nu_{j,t}$ represent the time-specific variable-specific mean across the population. The latent variables \hat{Y}_t and \hat{Z}_t represent the residuals. An RSEM model can be used to model the cross-lagged relationships among these residuals as follows. For $t = 2, \dots, T$

$$\hat{Y}_t = \beta_{1,t} \hat{Y}_{t-1} + \beta_{2,t} \hat{Z}_{t-1} + \varepsilon_{1,t} \quad (18)$$

$$\hat{Z}_t = \beta_{3,t} \hat{Y}_{t-1} + \beta_{4,t} \hat{Z}_{t-1} + \varepsilon_{2,t}. \quad (19)$$

The regression parameters $\beta_{j,t}$ represent the bivariate auto-regression and show how the deviations from the means persist over time. Time-specific residual variances can also be estimated $Var(\varepsilon_{j,t}) = \theta_{j,t}$ as well as the contemporaneous covariances $Cov(\varepsilon_{1,t}, \varepsilon_{2,t}) = c_t$.

3.2.1 RI-CLPM with continuous variables

Here we conduct a simulation study to evaluate the performance of the Bayes estimator for the RI-CLPM model with continuous variables using $T = 4$

and sample size of $N = 200$. Figure 3 contains the input file used for this simulation and Figure 4 contains the Mplus output results for a selection of the parameters. The results indicate that the Bayes estimator performs well for this model. Similar results can be obtained with the ML estimator if the Mplus **estimator=ml** option is used in the input file. Note here that in the Mplus language, the residual variance for a variable actually refers to the variance of the secondary residual, i.e., variance specification **y2*1** implies that $Var(\varepsilon_{1,2}) = \theta_{1,2} = 1$. The actual residual variance for Y_2 , which is $Var(\hat{Y}_2)$, is not a model parameter but a derived quantity which can be obtained from equation (18). Figure 5 shows the Mplus input file needed for the RI-CLPM model estimation with a single data set. This input file is substantially simpler than the input file used prior to Mplus 8.7, see Hamaker (2018).

Figure 3: Mplus input file for RI-CLPM simulation study with continuous variables

```
montecarlo: names = y1-y4 z1-z4;
nobs = 200; nreps = 100;

analysis: estimator=bayes; proc=2;

model population:
y1-y4*1 z1-z4*1;
riy by y1-y4@1; riy*0.4;
riz by z1-z4@1; riz*0.5;
y2^ on y1^*0.5 z1^*0.2;
y3^ on y2^*0.3 z2^*0.2;
y4^ on y3^*0.2 z3^*-0.2;
z2^ on z1^*0.5 y1^*0.1;
z3^ on z2^*0.3 y2^*0.3;
z4^ on z3^*0.2 y3^*0.1;
riy with riz*0.2;
y1-y4 pwith z1-z4*0.2;

model:
y1-y4*1 z1-z4*1;
riy by y1-y4@1; riy*0.4;
riz by z1-z4@1; riz*0.5;
y2^ on y1^*0.5 z1^*0.2;
y3^ on y2^*0.3 z2^*0.2;
y4^ on y3^*0.2 z3^*-0.2;
z2^ on z1^*0.5 y1^*0.1;
z3^ on z2^*0.3 y2^*0.3;
z4^ on z3^*0.2 y3^*0.1;
riy with riz*0.2;
y1-y4 pwith z1-z4*0.2;
```

Figure 4: Mplus output results for RI-CLPM simulation study with continuous variables

MODEL RESULTS		Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Y2^	ON							
Y1^		0.500	0.4958	0.1077	0.0943	0.0115	0.950	1.000
Z1^		0.200	0.1926	0.0817	0.0939	0.0067	0.990	0.610
Y3^	ON							
Y2^		0.300	0.2905	0.0817	0.0860	0.0067	0.970	0.910
Z2^		0.200	0.1970	0.0859	0.0868	0.0073	0.950	0.640
Z2^	ON							
Z1^		0.500	0.4821	0.0896	0.0981	0.0083	0.980	1.000
Y1^		0.100	0.0902	0.1031	0.0924	0.0106	0.920	0.200
Z3^	ON							
Z2^		0.300	0.2992	0.0914	0.0920	0.0083	0.940	0.880
Y2^		0.300	0.2968	0.0831	0.0840	0.0068	0.950	0.960
RIY	WITH							
RIZ		0.200	0.2156	0.0851	0.0869	0.0074	0.940	0.750
Y1	WITH							
Z1		0.200	0.1925	0.1057	0.1078	0.0111	0.960	0.460

Figure 5: Mplus RI-CLPM input file

```
variable: names = y1-y4 z1-z4;  
  
data: file is 1.dat;  
  
model:  
  riy by y1-y4@1;  
  riz by z1-z4@1;  
  y2^-y4^ z2^-z4^ pon y1^-y3^ z1^-z3^;  
  y2^-y4^ z2^-z4^ pon z1^-z3^ y1^-y3^;  
  y1-y4 pwith z1-z4;
```


3.2.2 RI-CLPM with categorical variables

The RI-CLPM can be estimated with categorical variables as well. The model will generally require somewhat bigger sample sizes as compared to the continuous variables case. In this simulation we use $N = 500$. Figure 6 contains the input file used for this simulation and Figure 7 contains the Mplus output results for a selection of the parameters. The results indicate that the Bayes estimator performs well in this case as well. The model can also be estimated with the WLSMV estimator using the option **estimator=wlsmv; param=theta;**. The RI-CLPM model in Figure 6 uses binary variables, however, the performance of the Bayes and the WLSMV estimators with ordered categorical variables is similar.

In the continuous case, the residual variance parameters $\theta_{j,t}$ are estimated as free and unequal parameters. In the example given in Figure 6 for the categorical case, the parameters $\theta_{j,t}$ are all fixed to 1 by default. This is along the lines of standard probit regression and is generally needed for identification purposes. In longitudinal studies, however, see for example growth models in Muthén and Asparouhov (2002), it is possible in principle to identify all but one of these residual variance parameters because the scale of the variables is aligned by the random intercept. If the variance parameter for one of the variables in the longitudinal process is fixed to 1, the variance parameters for all other variables in the longitudinal process can in principle be estimated as free parameters. For example, if the residual variance of Y_1 is fixed, the variance of the random intercept is identified from the R^2 of Y_1 explained by the time-invariant component in the model. Then, different values of θ_t for $t > 1$, would imply different random intercept based correlations between Y_1 and Y_t which means that θ_t is identified.

In the model given in Figure 6, one could fix $\theta_{1,1} = \theta_{2,1} = 1$ and estimate the remaining 6 residual variances. This modeling approach, however, is somewhat more advanced and has several caveats. First, the model is prone to empirical non-identification. The identifiability of these additional parameters depends on all other model parameters and for some data sets the parameters cannot be identified. Second, reliable estimation of these parameters requires much larger sample sizes. Third, even with larger sample sizes, the standard errors of these parameters are often quite large and very likely, the residual variance parameters will not be significantly different from 1. This defeats the purpose of estimating these parameters as free model parameters.

In practical situations, one could explore estimating a small selection of these residual variances, while still holding the majority of the residual variances fixed to 1. There are several different strategies that can be employed in deciding which residual variances should be freed. One strategy is to estimate the residual variance for only the first time point. These residual variances may indeed be bigger than the residual variances for remaining time points due to the fact that the first residuals are not regressed on any other variable. In fact, under the assumption of time invariance, where all coefficients are held equal across time, the first time point residual variances must be estimated as free parameters because the model for the first residual is different, see also equation (13).

A different strategy is to estimate the model with all residual variances fixed to 1 and then consider the standardized loadings of the random intercept. Ideally we would want these loadings to be not too different. If a particular variable stands out and has standardized loading for the random intercept that is very different from the rest of the variables, freeing that variable residual variance may yield standardized loadings that are more equitable.

A third strategy is to use the modification indices of the WLSMV estimator and free those residual variances that are most promising. A fourth strategy is to use the Bayes estimator and the BSEM methodology of Muthén and Asparouhov (2012), which is specifically designed to deal with somewhat poorly identified parameters and employs the concept of approximately fixed parameters via stringent prior specifications. This approach can be viewed as the Bayesian equivalent of the modification indices methodology. Prior to Mplus 8.7, the residual variances for categorical variables in the Mplus Bayesian framework were always fixed to 1. This restriction is now eliminated for Bayes and the residuals variances can now be estimated in a properly specified longitudinal or multiple group study, just as this is done with the WLSMV estimator.

Regardless of which strategy is utilized, in practical settings, estimating residual variances for categorical variables requires a more substantial sample size and carefully weighing the pros and cons of such modeling. Adding poorly identified parameters to a model can compromise the inference that can be made from the model as the standard errors of all other model parameters would likely be negatively impacted. On the flip side, fixing residual variances incorrectly to 1 will likely propagate biases in the rest of the model parameters.

There are two other issues that should be addressed here regarding the Bayesian estimation of RSEM models with categorical variables. The first issue is regarding the RI-CLPM model when the RI is actually not present in the model and the outcome variables are binary. In that case, the residual variable \hat{Y}^* and the actual variable Y^* are one and the same. However, due to various complexities in the Mplus algorithm, the autoregressive model specified as $\hat{Y}2 \text{ on } \hat{Y}1$ is estimated differently by Bayes from the model specified as $Y2 \text{ on } Y1$. The difference is related to how the threshold parameters are updated. In the first case, the thresholds are estimated as the negative of the mean of Y^* and that yields the most efficient algorithm. In the second case, the less efficient MH algorithm for updating the thresholds is used. The difference between the algorithms is generally small and in most situations will not be noticeable beyond the speed of the computation. In more complex models, however, such as Mixture models, the difference between the algorithms may become more substantial. Even in that case, however, for larger sample sizes the differences tend to disappear. In all cases, the most efficient algorithm in Mplus is obtained with the language $\hat{Y}2 \text{ on } \hat{Y}1$.

The second important issue for the Bayesian estimation of RSEM is specific to models with ordered categorical variables with thresholds held equal across time. In such situations, Mplus uses the least efficient and only available algorithm for updating the threshold parameters and the latent variables Y^* . This method is referred to as Method 3 in Asparouhov and Muthén (2010) and its applicability is somewhat limited. In practical applications where various complexities arise, the method may not perform well. Therefore, we generally recommended not holding thresholds equal across time for models such as RI-CLPM. Even when the thresholds are time invariant, a more efficient Bayesian estimation can be obtained by estimating time specific thresholds.

Figure 6: Mplus input file for RI-CLPM simulation study with categorical variables

```
montecarlo: names = y1-y4 z1-z4;
nobs = 500; nreps = 100;
categorical=all;
generate=y1-z4(1);

analysis: estimator=bayes; proc=2;

model population:
riy by y1-y4@1; riy*0.4;
riz by z1-z4@1; riz*0.5;
y2^ on y1^*0.5 z1^*0.2;
y3^ on y2^*0.3 z2^*0.2;
y4^ on y3^*0.2 z3^*-0.2;
z2^ on z1^*0.5 y1^*0.1;
z3^ on z2^*0.3 y2^*0.3;
z4^ on z3^*0.2 y3^*0.1;
riy with riz*0.2;
y1-y4 pwith z1-z4*0.2;

model:
riy by y1-y4@1; riy*0.4;
riz by z1-z4@1; riz*0.5;
y2^ on y1^*0.5 z1^*0.2;
y3^ on y2^*0.3 z2^*0.2;
y4^ on y3^*0.2 z3^*-0.2;
z2^ on z1^*0.5 y1^*0.1;
z3^ on z2^*0.3 y2^*0.3;
z4^ on z3^*0.2 y3^*0.1;
riy with riz*0.2;
y1-y4 pwith z1-z4*0.2;
```

Figure 7: Mplus output results for RI-CLPM simulation study with categorical variables

MODEL RESULTS		Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Y2^	ON							
Y1^		0.500	0.5013	0.1370	0.1236	0.0186	0.940	0.990
Z1^		0.200	0.2149	0.1040	0.1155	0.0109	0.980	0.480
Y3^	ON							
Y2^		0.300	0.2731	0.1133	0.1040	0.0134	0.920	0.710
Z2^		0.200	0.2086	0.1055	0.1067	0.0111	0.950	0.520
Z2^	ON							
Z1^		0.500	0.5006	0.1133	0.1300	0.0127	0.950	0.990
Y1^		0.100	0.0984	0.1037	0.1100	0.0106	0.980	0.160
Z3^	ON							
Z2^		0.300	0.2911	0.1072	0.1131	0.0115	0.960	0.750
Y2^		0.300	0.3127	0.1075	0.1064	0.0116	0.970	0.900
RIY	WITH							
RIZ		0.200	0.1999	0.0896	0.0883	0.0080	0.930	0.710
Y1	WITH							
Z1		0.200	0.2201	0.1141	0.1038	0.0133	0.900	0.510

3.3 ARMA models

3.3.1 D-RI-ARMA model

The auto-regressive moving-average (ARMA) model is one of the most commonly used models for analyzing time series data, see Box and Jenkins (1976). If Y_t is an observed variable at time t , the ARMA(1,1) model is given by the following equation

$$Y_t = \alpha + \rho Y_{t-1} + \beta \varepsilon_{t-1} + \varepsilon_t. \quad (20)$$

This model is an extension of the basic auto-regressive model. Through the parameter β , we can predict the dependent variable from the residual in the previous time period. The model is generally equivalent to an AR model with infinite lag and can account for correlations that persist in the data across longer periods of time, without the exponential decay limitation associated with the AR model. Because the model is conceptually the same as an RSEM model, where residuals are used in a path analysis sense, we can adopt the ARMA modeling idea to psychometrics and social science models, see Du Toit and Browne (2001). Suppose that Y_{it} is an observed variable for individual i at time t . We define the D-RI-ARMA (dynamic random intercept auto-regressive moving average) model as follows. For $t = 1$, the model starts with

$$Y_{i1} = \alpha_1 + \eta_i + \varepsilon_{i1}. \quad (21)$$

For $t > 1$ the model is given by the full ARMA expression

$$Y_{it} = \alpha_t + \eta_i + \rho_t Y_{i,t-1} + \beta_t \varepsilon_{i,t-1} + \varepsilon_{it}, \quad (22)$$

where $\varepsilon_{it} \sim N(0, \theta_t)$ and $\eta \sim N(0, \psi)$. This model can be viewed as the univariate version of the GCLM of Zyp hur et al. (2020a). If the moving average parameters $\beta_t = 0$, we obtain a simpler model which we call the D-RI-AR model. Just like in the RI-AR model, all the parameters in the D-RI-ARMA model are time-specific, although they can be estimated as time-invariant. The latent variable η_i takes the role of the random intercept (RI) and is the only subject specific parameter. In the RI-AR model, the random intercept takes the role of the subject specific mean parameter, while in the D-RI-ARMA model, the random intercept is not the subject specific mean parameter. The expected value for $E(Y_{it}|\eta_i)$ depends on the intercept parameters α_t and the autoregressive coefficients ρ_t , see Usami (2021).

The model can easily be extended to a multivariate model, like the RI-CLPM model, by simply using vector and matrix forms in the above equations. In the univariate case, the number of parameters in the D-RI-ARMA model is $4T - 1$, while the total degrees of freedom is $T(T + 3)/2$. Therefore the model would not be identified for $T < 5$. For $T = 5$, $T = 6$ and $T = 7$, the model is too close to the saturated model and most likely will often not be as useful because the standard errors of the parameters will be too large to establish significance. In these situations, however, imposing some time-invariance in the model parameters or fixing some of the insignificant parameters to 0 will likely improve the value of the model. For $T > 7$, we can find enough longitudinal evidence to support this more complex autocorrelation structure with all time-specific parameters. Because the model is estimated in a wide format, it would be time consuming to estimate it if $T > 50$ (or if $P \cdot T > 50$ in the multivariate case). The D-RI-ARMA model is likely to be most useful for T between 7 and 50 when the simpler D-RI-AR or RI-AR models are rejected. In such situations, the added flexibility of the moving-average component would allow us to find better fitting longitudinal models for panel data.

The D-RI-ARMA model can also be viewed as a moving-average extension of the lagged models discussed in Bollen and Brand (2010). In the D-RI-ARMA model, it is possible to not hold all the loadings of the random intercept fixed to 1. Such models are considered in Bollen and Brand (2010). One can argue that the first loading should be free because the first equation is different. This should generally be reserved for $T > 10$, because adding free loadings would not only diminish the degrees of freedom for the model but also will very likely increase the standard errors for the rest of the parameters beyond what is reasonable. Due to such large standard errors, the free loading may in fact become insignificantly different from 1. Freeing some of the loadings should be reserved for those situations when a substantial improvement in model fit is achieved by the additional parameters. Note also that if the first two loadings in the D-RI-ARMA model are free to be estimated, the model becomes unidentified. The parameter β_2 must then be fixed to 0 to obtain an identified model. Thus, if the first loading is free, freeing the second loading will not improve the model fit at all.

3.3.2 Longitudinal evidence

Panel data models, such as D-RI-ARMA and RI-CLPM, are somewhat different from standard SEM models. Similar to the SEM models, these models can be used to make inference for individuals that were not observed. In addition, however, these models attempt to make inference to some extent for time points that were not observed (time points in the future). Therefore one should be mindful regarding the asymptotics with respect to T and not just to N . This is what we refer to as the longitudinal evidence. Longitudinal modeling requires longitudinal evidence. The variance covariance from 4 time points can be regarded as mostly cross-sectional evidence. That is because we have many observations in the sample but very little in term of how these observations evolve across time. When we use such cross-sectional evidence to build longitudinal models, extreme caution should be applied. Simplest models should be preferred when T is small. When simple models do not fit the cross sectional evidence, that may be because the auto-regressive relations are not the same for all individuals rather than requiring a much more complex subject-invariant longitudinal model. Consider this as an example. For any panel data with $T = 7$, the D-RI-ARMA(2,2) model has 0 degrees of freedom and is virtually guaranteed to fit (the cross-sectional evidence of) the data perfectly. This of course doesn't mean that D-RI-ARMA(2,2) is a model useful to predict what will happen at $T = 8$ or that we have determined the underlying nature of the longitudinal process. Longitudinal evidence can be claimed when the standard variance covariance chi-square has a good amount of degrees of freedom (DF) left. We recommend for such models to have the DF be near or greater than the number of model parameters. This way we can avoid over-saturating the model with parameters and over-extending the cross-sectional evidence implications for the longitudinal processes. Weak longitudinal evidence can be strengthened by holding parameters equal across time, by holding parameters equal across neighbouring periods and by removing insignificant parameters from the model to obtain a more parsimonious model.

This problem of empirical underidentification when there are not enough waves of data, is also illustrated by Orth et al. (2021), who analyzed ten different data sets, most of which consisted of only four waves of data. They found that all models, except the simplest AR models, failed to converge for almost all data sets. We regard this as another cautionary tale for the possible misuse of cross-sectional evidence to inform longitudinal models.

When a complex longitudinal model is estimated, a very likely scenario is that some of the residual variances are estimated to be negative. Even if the model fits the data well, the negative residual variances should be taken as possible evidence that the model is over-saturated with parameters. As a result of this over-saturation, a good model fit is found, but the underlying longitudinal process is not feasible. Residual variance parameters can be constrained in Mplus to be positive via the model constraint command. Only when all model parameters are in the admissible parameter space, the fit of the model should be considered.

3.3.3 RI-ARMA model

As discussed in Asparouhov et al. (2018), the ARMA model is equivalent to a measurement error auto-regressive model (MEAR), see also Granger and Morris (1976). If the coefficients in the D-RI-ARMA model are time-invariant, the model becomes similar to the two-level DSEM-MEAR model described in Asparouhov et al. (2018). There are several important differences, however. One difference is that the DSEM-MEAR model equivalence to the ARMA models is subject to parameter restrictions. The MA parameter must be negative for this equivalence to work. Another difference is that the DSEM-MEAR model is estimated in long format, which means that it can be estimated for any T . The D-RI-ARMA model is estimated in wide format, so it can efficiently accommodate only $T < 50$. Yet another difference is the way the two models handle the initial equation at time $t = 1$, which can lead to noticeable differences in the parameter estimates for smaller $T < 10$. The most important difference, however, is how the random intercept is included in the model. In DSEM-MEAR, the random intercept is separate from the ARMA model, i.e., the ARMA model is defined for the residual variable which doesn't include the random intercept.

Here we also introduce a new model, the RI-ARMA (random intercept auto-regressive moving average) model that uses the same approach to the random intercept as the DSEM-MEAR and the RI-CLPM models. This approach can also be viewed as latent centering for the observed variables, see Asparouhov and Muthén (2019). The random intercept part of the model is

$$Y_{it} = \alpha_t + \eta_i + \varepsilon_{it}. \quad (23)$$

The ARMA model is then defined for the residual variables ε_{it} . The model

starts with $t = 2$

$$\varepsilon_{i2} = \rho_2 \varepsilon_{i1} + \hat{\varepsilon}_{i2}. \quad (24)$$

For $t > 2$ the model is given by the full ARMA expression

$$\varepsilon_{it} = \rho_t \varepsilon_{i,t-1} + \beta_t \hat{\varepsilon}_{i,t-1} + \hat{\varepsilon}_{it}. \quad (25)$$

The RI-ARMA model has one parameter less than the D-RI-ARMA model, i.e. $4T - 2$, because β_2 is unidentified (since ε_{i1} and $\hat{\varepsilon}_{i1}$ are one and the same variable). The parameters are as follows: T intercept parameters α_t , T residual variance parameters $\theta_t = Var(\hat{\varepsilon}_{it})$ (assuming $\varepsilon_{i1} = \hat{\varepsilon}_{i1}$), $T - 1$ autoregressive parameters ρ_t for $t > 1$, $T - 2$ moving-average parameters β_t for $t > 2$, and 1 random intercept variance parameter $\psi = Var(\eta_i)$. It should be noted here that if the random intercept loadings are all free parameters, the RI-ARMA model and the D-RI-ARMA model become equivalent and there is a simple reparameterization between the two models. This equivalence hold only in the univariate case and requires the removal of the first moving average parameter from the D-RI-ARMA model which would be unidentified when all the loadings are free, just as it is in the RI-ARMA case. In the multivariate case, an equivalence between the two models can be established if the loadings and the auto-regressive parameters are time invariant. If the loadings and the auto-regressive parameters are time specific, however, the models are not equivalent. Translating the multivariate RI-ARMA model into a multivariate D-RI-ARMA model, we find that the random intercept variables will cross-load onto the other variables, i.e., under the restriction that each variable random intercepts load only on that variable process (diagonal loading matrix), the two models are not equivalent. In a more broader framework, where the time invariant portion of the processes (the random intercepts) can cross-load freely on all the variables as in EFA, the two multivariate models would be equivalent when all parameters are time-specific.

In practical applications, there may be a substantive reason to prefer one of the two models, see Hamaker et al. (2015), Usami (2021) and Orth et al. (2021), but generally the chi-square test of fit can be used to determine which of the two models is best suited for the data.

The new Mplus hats language does not allow us to directly code the RI-ARMA model. Instead it should be done through the creation of the residual latent variables as in Hamaker (2018), which would then be followed by the ARMA model for those latent variables using the hats language. Mplus input

model statements for the RI-ARMA model are illustrated in Figure 8, where the factors f_t correspond to ε_{it} in equation (23).

The RI-ARMA and D-RI-ARMA models are examples of the extended RSEM model given in equations (5-6), rather than the basic RSEM model given in equations (1-3). Here, the residuals are used to augment the primary SEM model. Models with continuous variables can be estimated with the ML estimator, while models with categorical variables can be estimated with the WLSMV estimator. For categorical data, the RI-ARMA and D-RI-ARMA models are defined for Y_{it}^* instead of Y_{it} and the residual variances θ_t are fixed to 1 for identification purposes. With the Bayes estimator, it is possible to estimate the model by introducing latent variables to replace ε_{i1} . Such an approach however will require fixing the residual variance of Y_{it} to small positive value and will result in slow and inefficient estimation. Therefore the Bayesian estimation is not recommended for the RI-ARMA/D-RI-ARMA models at this time. The minimum number of time points needed to estimate the RI-ARMA/D-RI-ARMA models with ordered categorical variables will be slightly higher than it is for continuous variables and with binary variables it will be slightly higher than for ordered categorical.

Figure 8: Mplus RI-ARMA model statement

```
Model:  
i by y1-y7@1;  
f1 by y1@1; y1@0;  
f2 by y2@1; y2@0;  
f3 by y3@1; y3@0;  
f4 by y4@1; y4@0;  
f5 by y5@1; y5@0;  
f6 by y6@1; y6@0;  
f7 by y7@1; y7@0;  
f1 with i@0;  
f2-f7 pon f1-f6;  
f3-f7 pon f2^-f6^;
```

3.3.4 RI-MEAR model

Next, we introduce the RI-MEAR (random intercept measurement error auto-regressive) model which is equivalent to the RI-ARMA model given in equations (23-25). The random intercept part of the model is the same as in the RI-ARMA model and is given by the following equation

$$Y_{it} = \alpha_t + \eta_i + \hat{Y}_{it}. \quad (26)$$

The MEAR model is then defined for the residual variables \hat{Y}_{it} . The measurement error part of the model is given by

$$\hat{Y}_{it} = f_{it} + e_{it} \quad (27)$$

where e_{it} represents the measurement error. The auto-regressive part of the model is given for $t > 1$

$$f_{it} = \rho_t f_{i,t-1} + \xi_{it}. \quad (28)$$

For $t = 1$ we augment the model with $f_{i1} = \xi_{i1}$ for convenience. This model has $4T$ parameters: T intercept parameters α_t , T measurement error variance parameters $\sigma_t = Var(e_{it})$, T residual variance parameters $v_t = Var(\xi_t)$, $T-1$ auto-regressive parameters ρ_t , and 1 random intercept variance parameter $\phi = Var(\eta_i)$. This means the RI-MEAR model has two extra parameters as compared to the RI-ARMA model. Note however that v_T and σ_T are indistinguishable parameters. Both play the same role in the model implied variance covariance matrix $Var(Y)$. Thus one of the two parameters must be eliminated from the model, otherwise the model will not be identified. For convenience, we set $\sigma_T = 0$, i.e., the measurement error at the last measurement occasion cannot be identified separately. The same thing in fact applies to the first measurement occasion, although the non-identification there is slightly more complex and it would involve also the first auto-regressive parameter ρ_2 . The three parameters v_1 , σ_1 and ρ_2 are involved in the following indeterminacy. A set of other parameters v'_1 , σ'_1 and ρ'_2 would imply exactly the same variance covariance matrix $Var(Y)$ if these two equations are satisfied

$$v'_1 + \sigma'_1 = v_1 + \sigma_1 \quad (29)$$

$$(\rho'_2)^2 v'_1 = (\rho_2)^2 v_1. \quad (30)$$

Since these two quantities can not be used to determine 3 parameters, one of the three parameters is unidentified. We conclude that the measurement

error at the first occasion also can not be identified and we set $\sigma_1 = 0$. With the two necessary identification constraints $\sigma_1 = \sigma_T = 0$, the RI-MEAR model has the same number of parameters as the RI-ARMA model. The identification issues for this model are exactly the same as the identification issues of the quasi simplex model discussed in Joreskog (1978). The RI-MEAR model reduces to the quasi simplex model if we remove the random intercept η_i . The identification constraints typically used with the quasi simplex model are to hold the unidentified variance components to be equal to the ones in the neighbouring period, that is $\sigma_1 = \sigma_2$ and $\sigma_T = \sigma_{T-1}$. Such alternate identification constraints can be used with the RI-MEAR model as well. The choice of the identification constraints does not affect the model fit as the models are equivalent and there is a simple reparameterization between the alternate versions.

The equivalence between the RI-MEAR and RI-ARMA models is established in the Appendix. The RI-MEAR model gives an alternative interpretation of the RI-ARMA model. The main interpretation is that the variable Y_{it} has a measurement error e_{it} . If that measurement error is removed, the model will be reduced to the simpler RI-AR model. Note, however, that in practice e_{it} doesn't necessarily need to be interpreted as measurement error. It can simply be interpreted as a instantaneous effect that has no carry-over effect to the next period, i.e., instantaneous input to the process that leave no trace past the current period. The RI-ARMA model, however, appears to be more robust in small and medium sample sizes and smaller values of T . The RI-ARMA model estimation is more likely to converge, it is less likely to have exploding parameter values, it is less likely to require multiple random starting values in the estimation procedure, it is less likely to have multiple solutions, and it is less likely to yield inadmissible solutions, such as negative variances. The reparameterization formulas given in the Appendix show that the RI-MEAR model is expected to perform well when all moving average parameters are negative and all auto-regressive parameters are positive. In simulation studies, where parameters vary in a certain range, these restrictions are likely to be violated, particularly so when T and N are small, when the range in the parameter estimates is bigger.

There are two different ways to estimate the RI-MEAR model in Mplus. The first approach is given in Figure 9, where the factors f_t correspond to f_{it} in equation (27) and the residual variance of Y_t correspond to the measurement error e_{it} . The constraints $\sigma_1 = \sigma_T = 0$ are implemented by fixing the first and the last residual variance to zero for the observed variables.

The second approach is given in Figure 10, where the factors f_t correspond to the measurement errors e_{it} and the hat variables \hat{Y}_t correspond to factors f_{it} in equation (27). The constraints $\sigma_1 = \sigma_T = 0$ are implemented by not having factors f_t for the first and the last variable. The second approach appears to have two advantages and is therefore the recommended approach. Limited experimentation suggests that the second approach is more likely to converge. Also, the second approach is more efficient with the Bayesian estimation as it avoids fixing residual variances to 0.

The RI-MEAR model is very similar to the TSE (trait–state–error) model of Kenny and Zautra (1995) and discussed further by Cole et al. (2005). The difference between these models is only in the identification of the model parameters. In the RI-MEAR model, all parameters are time-specific, while in the TSE model the parameters are time-invariant. The TSE model has just 5 parameters: $\rho_t = \rho$, $\alpha_t = \alpha$, $\sigma_t = \sigma$, $v_t = v$, for $t > 1$, and v_1 . An important difference between the models is in the identification of the variance of the measurement error. In the RI-MEAR model, the measurement errors in the first and last time points cannot be identified and are thus removed from the model. On the other hand, the TSE model identifies the variance of the measurement error by holding it equal across time. Overall, the RI-MEAR model can be viewed as a more flexible version of the TSE model. The TSE model is nested within the RI-MEAR model. If time invariance of the model parameters is not supported by the data, the TSE model is prone to convergence failures and inadmissible parameter solutions, such as negative variance parameters.

Figure 9: Mplus RI-MEAR model statement

```
Analysis: estimator = ml; model=nocov;
```

```
Model:
```

```
i by y1-y8@1;  
f1 by y1;  
f2 by y2;  
f3 by y3;  
f4 by y4;  
f5 by y5;  
f6 by y6;  
f7 by y7;  
f8 by y8;  
f2-f8 pon f1-f7;  
y1@0 y8@0;
```


Figure 10: Alternate (preferred) Mplus RI-MEAR model statement

```
Analysis: estimator = ml; model=nocov;
```

```
Model:
```

```
i by y1-y8@1;
```

```
y2^-y8^ pon y1^-y7^;
```

```
f2 by y2@1;
```

```
f3 by y3@1;
```

```
f4 by y4@1;
```

```
f5 by y5@1;
```

```
f6 by y6@1;
```

```
f7 by y7@1;
```

3.3.5 RI-ARMA and RI-MEAR models for categorical data

Next, we consider the RI-ARMA and the two different ways of writing the RI-MEAR model for categorical data. Here, the equivalence between the models is somewhat more complicated. The issue that complicates the comparison is how the metric is set for the variables. In the RI-ARMA model the metric is set by fixing the residual variance of the factors f_i in Figure 8 to 1, that is $\theta_t = 1$. In the RI-MEAR model given in Figure 10, the metric will be set by fixing the residual variances of \hat{Y}_t to 1, that is $v_t = 1$. In the RI-MEAR model given in Figure 9, the metric will be set by fixing the residual variances of the factors f_i to 1, that is also $v_t = 1$, but one must free the variances of Y_t for all but the first and the last time points (those will be fixed by default). The two RI-MEAR methods will yield the same model, however, that model will be different from the RI-ARMA model. The reparameterization formulas given in the Appendix show that $\theta_t = 1$ doesn't translate into $v_t = 1$. To make the conversion from the RI-ARMA to the RI-MEAR model, the constraint $\theta_t = 1$ should be translated into a constraint for v_t . This is illustrated in Figure 11. In this figure, we generate data according to the RI-ARMA model and we estimate the RI-MEAR model using the approach of Figure 10. Instead of using $v_t = 1$ as the scale setter in the estimation, we use $\theta_t = 1$ as it is in the data generating model. That amounts to freeing v_t for $t > 1$ and restricting these parameters in the model constraints command through the conversion formulas so that $\theta_t = 1$. Doing so will result in equivalence of the RI-MEAR and RI-ARMA models. When we estimate this model over 100 replications, we obtain an average chi-square value of 26, which with 27 degrees of freedom results in 3% rejection rate. If we instead use $v_t = 1$ as the scale setter in this simulation, we obtain an average chi-square value of 44 and a rejection rate of 54%. We conclude that the RI-ARMA and the RI-MEAR models are equivalent for categorical data as well, however, in their native scale setter, the two models are different. In practical settings, both models should be estimated in their default state as we expect to see different levels of fit.

Figure 11: RI-MEAR to RI-ARMA conversion for categorical data

```
montecarlo: names = y1-y10;
nobs = 1000; nreps = 1;
categorical=all; generate=y1-y10(4);

analysis: estimator=wlsmv; param=theta; model=nocov;

model population:
f1 by y1@1; f2 by y2@1; f3 by y3@1; f4 by y4@1;
f5 by y5@1; f6 by y6@1; f7 by y7@1; f8 by y8@1;
f9 by y9@1; f10 by y10@1;
f1-f10@1; y1-y10@0;
f2-f10 pon f1-f9*0.5;
f3-f10 pon f2^-f9^-0.3;
f by y1-y10@1; f*1;
[y1$1-y10$1*-1.5];
[y1$2-y10$2*0];
[y1$3-y10$3*1];
[y1$4-y10$4*2];

model:
f2 by y2@1; f3 by y3@1; f4 by y4@1;
f5 by y5@1; f6 by y6@1; f7 by y7@1; f8 by y8@1;
f9 by y9@1; f2-f9 (s2-s9);
y2^-y10^ pon y1^-y9^*0.5 (r2-r10);
y2-y10 (v2-v10);
f by y1-y10@1; f*1;

model constraints:
v2=1-s2;
v3=1-s3-r3*r3*s2*(1-s2);
v4=1-s4-r4*r4*s3*(1-s3);
v5=1-s5-r5*r5*s4*(1-s4);
v6=1-s6-r6*r6*s5*(1-s5);
v7=1-s7-r7*r7*s6*(1-s6);
v8=1-s8-r8*r8*s7*(1-s7);
v9=1-s9-r9*r9*s8*(1-s8);
v10=1-r10*r10*s9*(1-s9);35
```

3.3.6 Simulation studies with continuous variables

To illustrate the D-RI-ARMA model for continuous variables, we conduct a simulation study using $T = 10$ time points for $N = 1000$ observations. The Mplus input file for this simulation is given in Figure 12 and the results for some of the parameters are given in Figure 13. The results indicate that the estimation performs well. In fact, the standard errors for all parameters are sufficiently small and all the parameters were significant in all 100 replications. The chi-square test of fit has an average value of 26 in this case which matches the degrees of freedom and yields a rejection rate of 6%. If we estimate the model without the moving-average parameters β_t on the same data, i.e. using the D-RI-AR model, the autoregressive coefficients are underestimated, the random intercept variance is overestimated and the average chi-square value is 233. With 35 degrees of freedom the model is rejected for all 100 replications. If we estimate the RI-AR model on the same data, the AR coefficients are also underestimated and the random intercept variance overestimated. In that case the chi-square average value is 690, which with 35 DF also results in 100% rejection rate. If we estimate the RI-ARMA model on the same data, the chi-square average value is 56, which with 27 DF also results in 85% rejection rate. Importantly, when the RI-ARMA model was estimated on the D-RI-ARMA generated data, multiple local solutions were found. Using 50 random starting values, between 0 and 3 local solutions are found in each replication. In 14% of the replications, the model estimation did not converge. Therefore, in practical situations, the RI-ARMA and the D-RI-ARMA models should be estimated in combination with the **starts** option in Mplus. In conclusion, the D-RI-ARMA model is well identified and can be used to improve model fit for those situations when the RI-AR/D-RI-AR/RI-ARMA models are rejected.

Figure 12: D-RI-ARMA simulation study

```
montecarlo: names = y1-y10;  
nobs = 1000; nreps = 100;  
  
analysis: estimator=ml; proc=2;  
  
model population:  
y1-y10*1;  
y2-y10 pon y1-y9*0.5;  
y2-y10 pon y1^-y9^*-0.3;  
f by y1-y10@1;  
f*1;  
  
model:  
y1-y10*1;  
y2-y10 pon y1-y9*0.5;  
y2-y10 pon y1^-y9^*-0.3;  
f by y1@1 y2-y10@1;  
f*1;
```

Figure 13: D-RI-ARMA simulation study results

MODEL RESULTS

		Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Y2	ON							
Y1^		-0.300	-0.3016	0.0621	0.0638	0.0038	0.960	1.000
Y3	ON							
Y2^		-0.300	-0.3012	0.0520	0.0507	0.0027	0.940	1.000
Y4	ON							
Y3^		-0.300	-0.2982	0.0453	0.0483	0.0020	0.950	1.000
Y2	ON							
Y1		0.500	0.4999	0.0537	0.0536	0.0028	0.950	1.000
Y3	ON							
Y2		0.500	0.4998	0.0270	0.0311	0.0007	0.990	1.000
Y4	ON							
Y3		0.500	0.5010	0.0239	0.0264	0.0006	0.960	1.000
Variances								
F		1.000	0.9938	0.0727	0.0781	0.0053	0.980	1.000

3.3.7 Simulation studies with categorical variables

A simulation study for the D-RI-ARMA model with categorical data using the WLSMV estimator is shown in Figure 14. The results of the simulation study are given in Figure 15. Similarly, a simulation study for the RI-ARMA model with categorical data is shown in Figure 16 and the results are given in Figure 17. The results indicate that for both models the estimation performs well with categorical data. The average chi-square value for the RI-ARMA model in Figure 16 is 26 and with 27 degrees of freedom that yields a rejection rate of 4%. For comparison, we analyze the RI-ARMA generated data with the D-RI-ARMA, D-RI-AR and RI-AR models. The D-RI-ARMA model appears to yield unbiased estimates for the random intercept variance, but the auto-regressive parameters are underestimated. The average chi-square in that case is 78 and with 26 degrees of freedom that yields a rejection rate of 98%. Multiple solutions are found with 10 random starting values in some of the replications. The average chi-square value for the D-RI-AR model is 150 and with 35 degrees of freedom that yields 100% rejection rate. In that case, both the random intercept variance and the autoregressive coefficients are underestimated. The average chi-square value for the RI-AR model is 51 and with 35 degrees of freedom that yields 55% rejection rate. In that case, the random intercept variance is slightly overestimated while the auto-regressive coefficients are underestimated. Overall, the RI-ARMA model with categorical variables is clearly distinguishable from the alternative models and can be pursued in those situations when the simpler D-RI-AR and RI-AR models are rejected.

Figure 14: D-RI-ARMA with categorical data simulation study

```
montecarlo: names = y1-y10;
  nobs = 1000; nreps = 100;
  categorical=all;
  generate=y1-y10(4);

analysis: estimator=wlsmv; proc=2;
param=theta;

model population:
y1-y10*1;
y2-y10 pon y1-y9*0.5;
y2-y10 pon y1^-y9^*-0.3;
f by y1-y10@1; f*1;
[y1$1-y10$1*-1.5];
[y1$2-y10$2*0];
[y1$3-y10$3*1];
[y1$4-y10$4*2];

model:
y2-y10 pon y1-y9*0.5;
y2-y10 pon y1^-y9^*-0.3;
f by y1-y10@1; f*1;
[y1$1-y10$1*-1.5];
[y1$2-y10$2*0];
[y1$3-y10$3*1];
[y1$4-y10$4*2];
```


Figure 15: D-RI-ARMA with categorical data simulation study results

MODEL RESULTS		Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Y2	ON	-0.300	-0.3061	0.1047	0.0999	0.0109	0.970	0.890
Y1^								
Y3	ON	-0.300	-0.2990	0.0833	0.0819	0.0069	0.950	0.970
Y2^								
Y4	ON	-0.300	-0.3059	0.0698	0.0759	0.0049	0.970	0.990
Y3^								
Y2	ON	0.500	0.5004	0.1023	0.0958	0.0104	0.940	1.000
Y1								
Y3	ON	0.500	0.4974	0.0621	0.0634	0.0038	0.970	1.000
Y2								
Y4	ON	0.500	0.5053	0.0521	0.0563	0.0027	0.980	1.000
Y3								
Variances								
F		1.000	0.9982	0.1028	0.1098	0.0105	0.970	1.000

Figure 16: RI-ARMA with categorical data simulation study

```
montecarlo: names = y1-y10;
nobs = 1000; nreps = 100;
categorical=all; generate=y1-y10(4);

analysis: estimator=wlsmv; param=theta; model=nocov;

model population:
f1 by y1@1; f2 by y2@1; f3 by y3@1; f4 by y4@1;
f5 by y5@1; f6 by y6@1; f7 by y7@1; f8 by y8@1;
f9 by y9@1; f10 by y10@1;
f1-f10@1; y1-y10@0;
f2-f10 pon f1-f9*0.5;
f3-f10 pon f2^-f9^-0.3;
f by y1-y10@1; f*1;
[y1$1-y10$1*-1.5];
[y1$2-y10$2*0];
[y1$3-y10$3*1];
[y1$4-y10$4*2];

model:
f1 by y1@1; f2 by y2@1; f3 by y3@1; f4 by y4@1;
f5 by y5@1; f6 by y6@1; f7 by y7@1; f8 by y8@1;
f9 by y9@1; f10 by y10@1;
f1-f10@1; y1-y10@0;
f2-f10 pon f1-f9*0.5;
f3-f10 pon f2^-f9^-0.3;
f by y1-y10@1; f*1;
[y1$1-y10$1*-1.5];
[y1$2-y10$2*0];
[y1$3-y10$3*1];
[y1$4-y10$4*2];
```

Figure 17: RI-ARMA with categorical data simulation study results

MODEL RESULTS		Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
F2	ON							
F1		0.500	0.5075	0.0621	0.0529	0.0039	0.960	1.000
F3	ON							
F2		0.500	0.5049	0.1019	0.0988	0.0103	0.950	1.000
F2^		-0.300	-0.3076	0.1162	0.1067	0.0134	0.920	0.830
F4	ON							
F3		0.500	0.5464	0.1633	0.1614	0.0286	0.940	0.950
F3^		-0.300	-0.3461	0.1632	0.1608	0.0285	0.970	0.650
Thresholds								
Y1\$1		-1.500	-1.4863	0.0720	0.0703	0.0053	0.940	1.000
Y1\$2		0.000	0.0093	0.0541	0.0557	0.0030	0.970	0.030
Y1\$3		1.000	1.0027	0.0607	0.0620	0.0037	0.970	1.000
Y1\$4		2.000	2.0012	0.0811	0.0838	0.0065	0.980	1.000
Y2\$1		-1.500	-1.4963	0.0627	0.0704	0.0039	0.960	1.000
Y2\$2		0.000	0.0050	0.0523	0.0593	0.0027	0.980	0.020
Y2\$3		1.000	1.0023	0.0662	0.0639	0.0043	0.950	1.000
Y2\$4		2.000	2.0042	0.0762	0.0814	0.0058	0.970	1.000
Variances								
F		1.000	0.9743	0.0944	0.0792	0.0095	0.970	0.990

3.3.8 Bivariate models

Next we illustrate the bivariate version of the RI-ARMA model where all cross-lagged relations are present. This model can be viewed as the matrix version of the univariate RI-ARMA model given in equations (23-25). In this bivariate version, Y_{it} , α_t , η_i , ε_{it} , $\hat{\varepsilon}_{it}$ are all vectors of size 2, while the regression parameters ρ_t and β_t are square matrices of size 2. This model can also be viewed as the moving average extension of the RI-CLPM described in equations (16-19). As in the univariate case, this model is recommended only when $T > 7$, unless time-invariance in some of the parameters is assumed. Figure 18 shows the input file for a bivariate RI-ARMA model simulation study with $T = 10$. The results of the simulation, given in Figure 19, indicate the ML estimation performs well. The average chi-square value in this simulation study is 107, which with 110 degrees of freedom results in 3% rejection rate. For comparison, using the same data generation, we estimate the bivariate RI-ARMA model without the cross-lagged moving average components, i.e., restricting the β_t regression matrix to be a diagonal square matrix. In that case, the average chi-square value is 147, which with 126 degrees of freedom results in 34% rejection rate. In 66% of the replications, the bivariate RI-ARMA model without the cross-lagged moving average components provides sufficient fit for the data even though the data was generated with the full bivariate RI-ARMA model. Thus, excluding the cross-lagged moving average components provides a more parsimonious alternative to the full bivariate RI-ARMA model, which can be useful in practical applications and may provide sufficient model fit. For comparison, we also estimate the RI-CLPM for these data. The average chi-square test of fit for the RI-CLPM model is 255, which with 142 degrees of freedom results in 100% rejection rate. To summarize, the bivariate RI-ARMA model can be used in those situations where the simpler RI-CLPM model is rejected. We also note here that the bivariate RI-ARMA model is equivalent to a bivariate RI-MEAR model, just as this is so in the univariate case. The bivariate RI-MEAR model, however, must include residual cross-correlation between the measurement error and the observed variables.

The bivariate version of the D-RI-ARMA model has also been discussed in Zyphur et al. (2020a) and Zyphur et al. (2020b). The GCLM model defined in these articles, however, has different parameter restrictions. The loading parameters for the random intercept are all free parameters, except for one time point, while all auto-regressive and moving-average parameters

are time-invariant. The new hats language in Mplus can be used to simplify the specification for the GCLM model as well. Using data from Van Iddekinge et al. (2009), the GCLM is illustrated in the online materials provided in Zyphur et al. (2020a). The original Mplus model statement used in Zyphur et al. (2020a) is given in Figure 20, while the new simplified model statement is given in Figure 21.

Figure 18: Bivariate RI-ARMA simulation study

```
montecarlo: names = y1-y10 z1-z10;
nobs = 1000; nreps = 100;

analysis: estimator=ml; model=nocov;

model population:
f1 by y1@1; f2 by y2@1; f3 by y3@1; f4 by y4@1;
f5 by y5@1; f6 by y6@1; f7 by y7@1; f8 by y8@1;
f9 by y9@1; f10 by y10@1;
g1 by z1@1; g2 by z2@1; g3 by z3@1; g4 by z4@1;
g5 by z5@1; g6 by z6@1; g7 by z7@1; g8 by z8@1;
g9 by z9@1; g10 by z10@1;
f1-f10*1; y1-y10@0; g1-g10*1; z1-z10@0;
f by y1-y10@1; f*1; g by z1-z10@1; g*1;
f2-f10 g2-g10 pon f1-f9*0.5 g1-g9*0.5;
g2-g10 f2-f10 pon f1-f9*0.2 g1-g9*0.2;
f3-f10 g3-g10 pon f2^-f9^*-0.3 g2^-g9^*-0.3;
g3-g10 f3-f10 pon f2^-f9^*-0.2 g2^-g9^*-0.2;
f1-f10 pwith g1-g10*0.3; f with g*0.3;

model:
f1 by y1@1; f2 by y2@1; f3 by y3@1; f4 by y4@1;
f5 by y5@1; f6 by y6@1; f7 by y7@1; f8 by y8@1;
f9 by y9@1; f10 by y10@1;
g1 by z1@1; g2 by z2@1; g3 by z3@1; g4 by z4@1;
g5 by z5@1; g6 by z6@1; g7 by z7@1; g8 by z8@1;
g9 by z9@1; g10 by z10@1;
f1-f10*1; y1-y10@0; g1-g10*1; z1-z10@0;
f by y1-y10@1; f*1; g by z1-z10@1; g*1;
f2-f10 g2-g10 pon f1-f9*0.5 g1-g9*0.5;
g2-g10 f2-f10 pon f1-f9*0.2 g1-g9*0.2;
f3-f10 g3-g10 pon f2^-f9^*-0.3 g2^-g9^*-0.3;
g3-g10 f3-f10 pon f2^-f9^*-0.2 g2^-g9^*-0.2;
f1-f10 pwith g1-g10*0.3; f with g*0.3;
```

Figure 19: Bivariate RI-ARMA simulation study results

MODEL RESULTS		Population	ESTIMATES		S. E.	M. S. E.	95% Cover	% Sig
			Average	Std. Dev.	Average			Coeff
F2	ON							
F1		0.500	0.4950	0.0423	0.0427	0.0018	0.946	1.000
G1		0.200	0.2033	0.0443	0.0412	0.0020	0.946	1.000
F3	ON							
F2		0.500	0.4984	0.1415	0.1248	0.0198	0.913	0.978
G2		0.200	0.2048	0.1194	0.1223	0.0141	0.978	0.380
F2^		-0.300	-0.3027	0.1369	0.1229	0.0185	0.935	0.707
G2^		-0.200	-0.2044	0.1088	0.1208	0.0117	0.967	0.380
G2	ON							
G1		0.500	0.4959	0.0447	0.0435	0.0020	0.946	1.000
F1		0.200	0.1960	0.0424	0.0413	0.0018	0.924	1.000
G3	ON							
G2		0.500	0.4907	0.1239	0.1263	0.0153	0.946	0.946
F2		0.200	0.2108	0.1290	0.1212	0.0166	0.946	0.402
G2^		-0.300	-0.2923	0.1238	0.1239	0.0152	0.967	0.674
F2^		-0.200	-0.2070	0.1292	0.1194	0.0166	0.946	0.391
F	WITH							
G		0.300	0.2800	0.1897	0.1266	0.0360	0.902	0.891
F1	WITH							
G1		0.300	0.3254	0.1891	0.1236	0.0360	0.924	0.946
Variances								
F		1.000	0.9766	0.1927	0.1389	0.0373	0.924	0.978
G		1.000	0.9820	0.1904	0.1406	0.0362	0.924	0.967

Figure 20: Zyphur et al. (2020a) GCLM model specification

```
ANALYSIS: Estimator = ML;

MODEL:
! Unit Effects
Eta1_X by X6-X1;
Eta1_Y by Y6-Y1;

! Impulses
u_X6 by X6; X6@0; u_X5 by X5; X5@0; u_X4 by X4; X4@0; u_X3 by X3;
X3@0; u_X2 by X2; X2@0; u_X1 by X1; X1@0;
u_Y6 by Y6; Y6@0; u_Y5 by Y5; Y5@0; u_Y4 by Y4; Y4@0; u_Y3 by Y3;
Y3@0; u_Y2 by Y2; Y2@0; u_Y1 by Y1; Y1@0;

! AR Terms
X6-X2 PON X5-X1 (AR1_X);
Y6-Y2 PON Y5-Y1 (AR1_Y);

! MA Terms
X6-X2 PON u_X5-u_X1 (MA1_X);
Y6-Y2 PON u_Y5-u_Y1 (MA1_Y);

! CL Terms
X6-X2 PON Y5-Y1 (CL1_XY);
Y6-Y2 PON X5-X1 (CL1_YX);

! CLMA Terms
X6-X2 PON u_Y5-u_Y1 (CLMA1_XY);
Y6-Y2 PON u_X5-u_X1 (CLMA1_YX);

! Restrictions
u_X6-u_X1 WITH Eta1_X@0 Eta1_Y@0;
u_Y6-u_Y1 WITH Eta1_X@0 Eta1_Y@0;
u_X6-u_X1 WITH u_X6-u_X1@0;
u_Y6-u_Y1 WITH u_Y6-u_Y1@0; u_X6-u_X1 WITH u_Y6-u_Y1@0;

! Co-Movements
u_X6 WITH u_Y6;
u_X5 WITH u_Y5;
u_X4 WITH u_Y4;
u_X3 WITH u_Y3;
u_X2 WITH u_Y2;
u_X1 WITH u_Y1;
```


Figure 21: Zyp hur et al. (2020a) GCLM simplified model specification using RSEM hats language

```
ANALYSIS: Estimator = ML; model=nocov;

MODEL:
! Unit Effects
Eta1_X by X6-X1;
Eta1_Y by Y6-Y1;

! AR Terms
X6-X2 PON X5-X1 (AR1_X);
Y6-Y2 PON Y5-Y1 (AR1_Y);

! MA Terms
X6-X2 PON X5^-X1^ (MA1_X);
Y6-Y2 PON Y5^-Y1^ (MA1_Y);

! CL Terms
X6-X2 PON Y5-Y1 (CL1_XY);
Y6-Y2 PON X5-X1 (CL1_YX);

! CLMA Terms
X6-X2 PON Y5^-Y1^ (CLMA1_XY);
Y6-Y2 PON X5^-X1^ (CLMA1_YX);

X6-X1 PWITH Y6-Y1;
ETA1_Y WITH ETA1_X;
```

3.4 Residual variables as predictors

In certain situations it is possible to use the residual variables as predictors for other variables. As an example we consider a factor analysis model where the factor and some of the residuals can be used to predict a distal outcome variable. The factor analysis model is given by the following equation. For $p = 1, \dots, P$

$$Y_p = \nu_p + \lambda_p \eta + \varepsilon_p \quad (31)$$

$$\varepsilon_p \sim N(0, \theta_p) \quad (32)$$

$$\eta \sim N(0, 1). \quad (33)$$

We can now use the latent factor η as well as the residuals ε_p to predict a distal outcome variable Z as follows

$$Z = \alpha + \beta_0 \eta + \sum_{p=1}^P \beta_p \varepsilon_p + \zeta, \quad (34)$$

where $\zeta \sim N(0, \psi)$. Not all of the β_i parameters can be identified. A maximum of P regression parameters can be identified in the above equation and therefore at least one of these must be fixed to zero. Within the BSEM framework of Muthén and Asparouhov (2012), it is possible to estimate all of the regression parameters in an exploratory sense where tiny priors are specified for β_1, \dots, β_P .

The above model allows us explore predictive relations between the indicator variables Y_p and the distal outcome variable that go beyond the predictive effect of the measured factor. In principle, in the above model, one can use Y_p directly instead of ε_p . In certain situations, however, it is preferable to use the independent predictors ε_p and η because additional predictors would not affect the existing regression coefficients.

Note also that equation (34) is equivalent to the following model

$$Z = \alpha + \beta_0 \eta + \hat{Z} \quad (35)$$

$$\hat{Z} = \sum_{p=1}^P \beta_p \varepsilon_p + \zeta. \quad (36)$$

Even though these models are equivalent, they are coded differently in Mplus. Model (34) would be specified as **Z on Y1 ^ -YP ^** while model (35-36) would

use the \mathbf{Z}^{\wedge} on $\mathbf{Y1}^{\wedge} - \mathbf{YP}^{\wedge}$ specification. There are two important differences between these two versions. First, the two models will not be equivalent if the residual of Z is used as a predictor in an another equation. That is because in model (34) the residual of Z is ζ , while in model (35-36) the residual of Z is \hat{Z} . The second difference is specific to the estimators. With the current capabilities in Mplus, the Bayesian estimation allows residual variables to be regressed only among each other. That means that only model (35-36) can be estimated with the Bayesian estimator. The ML estimator can be used to estimate both models and residual variables can be used as predictors for all variables in the model.

Figure 22 shows an example of a simulation input file for the above model where two of the residuals of the factor model are used as predictors for a distal outcome. Figure 23 shows the results from this simulation for a selection of the parameters. The results indicate that the model estimation performs well.

The models discussed in this section extend to categorical variables as well. The Bayesian estimator as well as the WLSMV estimator can be used to estimate models where the residuals of categorical variables are used to predict distal outcomes.

Figure 22: Input file for a simulation study where residual variables are used as predictors

```
montecarlo:
    names = y1-y7 z;
    nobs = 1000;
    nreps = 100;

analysis: estimator=bayes; proc=2;

model population:
f by y1-y7*1; f@1;
y1-y7*1 z*1;
z^ on y3^*.1;
z^ on y4^*.3;
z on f*.5;

model:
f by y1-y7*1; f@1;
y1-y7*1 z*1;
z^ on y3^*.1;
z^ on y4^*.3;
z on f*.5;
```

Figure 23: Output results for a simulation study where residual variables are used as predictors

MODEL RESULTS		Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
F	BY							
Y1		1.000	1.0098	0.0392	0.0406	0.0016	0.950	1.000
Y2		1.000	1.0075	0.0378	0.0406	0.0015	0.970	1.000
Y3		1.000	1.0040	0.0383	0.0409	0.0015	0.950	1.000
Z	ON							
F		0.500	0.5101	0.0444	0.0385	0.0021	0.910	1.000
Z^	ON							
Y3^		0.100	0.0975	0.0359	0.0343	0.0013	0.920	0.780
Y4^		0.300	0.3020	0.0358	0.0348	0.0013	0.940	1.000

3.5 Incomplete variance covariance blocks with the Bayesian estimation

The Bayesian estimation is most efficient when conjugate priors are used for all model parameters. For the variance covariance parameters, the conjugate prior is the Inverse-Wishart prior. However, that prior is only available when the variance covariance matrices are block diagonal, see Asparouhov and Muthén (2010). If a variance covariance structure is not block diagonal, conjugate priors are not available. In such situations, the Mplus Bayesian estimation uses the much less efficient algorithm of random walk based Metropolis-Hastings (MH). This algorithm is specified in Mplus with the option **algo = gibbs(rw)**. In simpler models, the MH algorithm performs well, although it usually takes a very large number of MCMC iterations to converge. In more complex models, however, where other estimation challenges may exist, the MH algorithm will simply fail to converge even with a very large number of MCMC iterations.

The block-diagonal model restriction can be formulated in practical terms as follows. If a variable Z is correlated with variables Y_1 and Y_2 , then Y_1 and Y_2 must also be correlated. In certain modeling situations, however, the covariance between Y_1 and Y_2 can not be estimated. If for example Y_1 is regressed on Y_2 , the covariance between Y_1 and Y_2 can not be estimated because the model would become unidentified. In such situations, the only alternative would be to use the inefficient MH algorithm. The RSEM model, however, appears to provide a solution to this problem. Regressing the residual of Z on the residuals of Y_1 and Y_2 produces a variance covariance matrix of the desired non-block diagonal pattern. That is, by using **Z^ on Y1^ Y2^** model specification, we obtain a variance covariance matrix for the residuals of Z , Y_1 and Y_2 , where Z is correlated with Y_1 and Y_2 , but Y_1 is not correlated with Y_2 .

We consider two more practical examples. The first example is related to auxiliary variables that are used for improving the missing data handling, see Asparouhov and Muthén (2008). In particular, formula (1) in that article shows how adding auxiliary variable to a model, for the purpose of improving the missing data handling, leads to incomplete / non-block diagonal variance covariance matrices. Consider again the model given in equations (31-36). For the purposes of this illustration, however, we interpret the variable Z as an auxiliary variable that may contain information related to the missing values of the factor indicators Y_p . We want Z to be correlated with all

residual variables ε_p . If we are to form a complete block diagonal matrix in such an example, we would have to model all correlations among all the residual variables ε_p as well. This of course will make the factor model unidentified. Thus, it is important to correlate Z with ε_p without introducing any correlations among the residuals. The RSEM model in equation (36) provides exactly that. In fact, the example provided in Figure 22 implies precisely that kind of incomplete variance covariance matrix between Z , ε_3 and ε_4 . The model implied variance covariance shows that $Cov(\varepsilon_3, \varepsilon_4) = 0$ while both $Cov(Z, \varepsilon_3)$ and $Cov(Z, \varepsilon_4)$ are not zero.

The second practical example is related to growth models, similar to the model given in equation (7). It is a fairly common practice to include in that model all correlation parameters between consecutive observations, i.e., the parameters $Cov(\varepsilon_{i1}, \varepsilon_{i2})$, $Cov(\varepsilon_{i2}, \varepsilon_{i3})$, etc. Such a model would technically not be an autoregressive model but because the correlations in the neighboring observations are the strongest, a vast portion of the autocorrelations would be modeled with the inclusion of these parameters. The resulting variance covariance matrix is not block diagonal as the only parameters that are not zero in that variance covariance matrix are on the diagonal next to the main diagonal. Estimating such a model would require the use of the MH algorithm and would frequently become impossible to estimate due to the inefficiency of the algorithm. We clearly cannot estimate the full variance covariance matrix for the residuals as the growth model will become unidentified. The solution to this problem is given precisely in the RSEM model given in equation (11) and is illustrated in Figure 1. We can replace the neighboring correlation model with a full auto-regressive model. While the two models are clearly not the same model as the autoregressive model implies non-zero (but diminishing) correlations even in non-neighboring residuals, for all practical purposes the RSEM model (11) resolves the problem completely.

3.6 BSEM estimation with unconstrained variance covariance for the residuals

The BSEM methodology described in Muthén and Asparouhov (2012) and Asparouhov et al. (2015) can be used to discover residual correlation in a general SEM model. Such a methodology is the Bayesian equivalent to the modification indices methodology used with the ML estimator, see Sörbom (1989). The method is based on estimating all residual covariance parameters

within the SEM model. Because the SEM model becomes unidentified when all residual correlations are included, a very restrictive prior is specified for the residual variance covariance matrix. The prior would generally hold the residual correlation near zero, unless, within the estimation, substantial information is found in the data to indicate that a covariance parameter is not zero. This is an iterative process that requires multiple model estimations with varying degree of prior restrictiveness. The process is described in detail in Asparouhov et al. (2015). Specifying a restrictive prior for a variance covariance matrix, however, is somewhat difficult. Typically, the Inverse Wishart prior is used for these parameters as this is the conjugate prior needed for the most efficient Bayesian estimation. The Inverse Wishart prior is a multivariate prior which makes the process difficult for two reasons. First, the prior is specified not just for the covariances but also for the variances in the variance/covariance matrix. Second, the level of prior restrictiveness is the same for both the covariances and the variances. Thus, if we want the covariances to be nearly fixed to 0, then we will need the variances to also be nearly fixed. The values to which the variances must be nearly fixed should be carefully picked as to not obstruct the SEM estimation. Within the BSEM iterative process, this becomes cumbersome as the priors must be carefully calibrated with every level of prior restrictiveness.

These complexities can be avoided with the RSEM model where conjugate priors can be specified separately for all the variance covariance parameters because they are essentially converted to regression among the residuals. Restrictive conjugate priors for regression parameters are simply the normal priors with mean zero and small variances. In the RSEM model, restrictive conjugate priors are not given for the variances.

We illustrate the BSEM methodology utilization of the RSEM model with a simple factor analysis model with P indicators Y_p , $p = 1, \dots, P$ and one factor η . The factor model is given by the following equation

$$Y_p = \nu_p + \lambda_p \eta + \hat{Y}_p, \quad (37)$$

where as usual \hat{Y}_p are the residual variables and $\eta \sim N(0, 1)$. To explore possible non-independence between the residuals we introduce the following RSEM model

$$\hat{Y}_p = \sum_{i=1}^{p-1} \beta_{pi} \hat{Y}_i + \zeta_p. \quad (38)$$

In this RSEM model, every residual is regressed on all the previous residuals.

The model is equivalent to estimating the full variance covariance matrix for the residuals. Next we impose the BSEM style prior on all the regression parameters β_{pi}

$$\beta_{pi} \sim N(0, \sigma), \quad (39)$$

where σ is the intended to be a small value which controls the level of prior restrictiveness.

We conduct a simulation study to illustrate the above process. We use $P = 7$ factor indicators in the simulation study and sample size of $N = 1000$. In the data generation we include 3 residual variances between indicators: Y_1 and Y_3 , Y_3 and Y_5 , and Y_6 and Y_7 , as illustrated in Figure 24. This figure also illustrates the BSEM utilization of RSEM and the tiny priors for the residual regression parameters with $\sigma = 0.01$. As usual, we have to vary the value of σ in this estimation process. With $\sigma = 0$ (i.e. the pure factor analysis model without the RSEM extension), $\sigma = 0.0001$ and $\sigma = 0.001$, the PPP rejects all 100 of the replications in this study. With $\sigma = 0.01$, the PPP rejects 0 of the replications. Therefore, we select $\sigma = 0.01$ as the level of prior restrictiveness that would enable us to select the most promising residual covariances. The results of this simulation study are shown in Figure 25, which contains the top 6 in order of magnitude residual regression parameters. All other residual regression parameters are smaller than 0.1 by absolute value. The top 3 of these are precisely the ones used in the data generation model, i.e., the most promising residual correlations reported by this BSEM+RSEM analysis are exactly those that are needed. The final step in such an analysis would be to include only those top few residual correlations (or residual regressions) in the factor analysis model, without any restrictive priors.

Figure 24: BSEM utilization of RSEM simulation study to determine the residual variances needed for acceptable model fit

```
montecarlo:
    names = y1-y7;
    nobs = 1000;
    nreps = 100;

analysis: estimator=bayes;

model population:
f by y1-y7*1; f@1;
y1-y7*1;
y1 with y3*0.3;
y3 with y5*0.3;
y6 with y7*0.6;

model:
f by y1-y7*1; f@1;
y1-y7*1;
y7^ on y1^-y6^*0 (p1-p6);
y6^ on y1^-y5^*0 (p7-p11);
y5^ on y1^-y4^*0 (p12-p15);
y4^ on y1^-y3^*0 (p16-p18);
y3^ on y1^-y2^*0 (p19-p20);
y2^ on y1^*0 (p21);

model prior: p1-p21~N(0,0.01);
```

Figure 25: BSEM utilization of RSEM output results

MODEL RESULTS

		Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Y5^	ON							
Y3^		0.000	0.2654	0.0271	0.0584	0.0711	0.000	1.000
Y7^	ON							
Y6^		0.000	0.2463	0.0494	0.1161	0.0631	0.360	0.640
Y3^	ON							
Y1^		0.000	0.2416	0.0277	0.0574	0.0591	0.000	1.000
Y6^	ON							
Y3^		0.000	-0.1420	0.0206	0.0657	0.0206	0.210	0.790
Y7^	ON							
Y3^		0.000	-0.1383	0.0218	0.0603	0.0196	0.140	0.860
Y5^		0.000	-0.1067	0.0215	0.0594	0.0118	0.460	0.540

3.7 LCA analysis with conditional dependence

LCA analysis can be combined with the RSEM methodology to explore conditional dependence between the class indicators. Suppose that Y_1, \dots, Y_p are categorical variables that measure a latent class variable C with K categories. The standard LCA model is given by the following equations

$$P(Y_p = l | C = k) = \Phi(\tau_{pk l}) - \Phi(\tau_{pk, l-1}) \quad (40)$$

$$P(C = k) = p_k \quad (41)$$

where $l = 1, \dots, L_p$ represents the observed values for Y_p and the threshold parameters $\tau_{pk l}$ determine the prevalence of every category for every class. As usual, we assume that $\tau_{pk 0} = -\infty$ and $\tau_{pk L_p} = \infty$. The standard LCA model assumes that all observed variables are conditionally independent, i.e., within each class or conditional on the class variables, the Y_p variables are independent. This assumption is also sometimes referred to as the local independence assumption. The latent class variables C is used to model all correlations among the indicator variables. In practical applications, however, the assumption is often violated. As an unfortunate consequence, spurious classes that are difficult to interpret are introduced in the model to achieve good model fit. One way to introduce local dependence, without additional spurious classes, is via continuous latent variables that influence pairs of variables to create correlations between them. This approach, however, is limited in the number of correlations that can be modeled. The ML estimation will use numerical integration for each continuous latent variable and therefore the modeling will be limited to a maximum of 3 or 4 correlations. With the Bayes estimation, it is possible to add more latent variables but the model estimation efficiency will decrease as more latent components are added to the model. With the ML estimation, it is possible to use the association model, Asparouhov and Muthén (2015), using the Mplus option **param=rescov**, however, such a model does not accommodate any additional predictors for the observed variables at this time and will become computationally intensive when the size of the variance/covariance matrix is larger. In addition, the parameter estimates of the association model are on a somewhat of a different metric than the standard LCA model which makes it a bit difficult to use from a practical perspective. The most straight forward way to address the local dependence is to simply add the correlation model for Y^* , see Asparouhov and Muthén (2011), as follows. We replace equation (40) by

$$Y_p = l \iff \tau_{pk, l-1} < Y_p^* \leq \tau_{pk l} \quad (42)$$

$$[Y^*|C = k] = [Y_1^*, \dots, Y_p^*|C = k] \sim N(0, \Sigma_k) \quad (43)$$

where Σ_k is a correlation matrix with 1 on the main diagonal. Currently this model can be estimated only with the Bayesian estimator. For binary variables, the model is discussed in detail in Asparouhov and Muthén (2011) and in Mplus 8.7 it can also be estimated with ordered categorical variables. The model can also include continuous latent variables which would essentially amount to estimating the model with Σ_k being a full variance covariance matrix rather than a correlation matrix.

The algorithmic advance that allows us to now estimate the model with ordered categorical variables can be described as follows. When the Y^* variance covariance matrix Σ_k is diagonal, the most efficient MCMC estimation method is based on estimating the model by grouping Y^* and C in one updating block, using Method 3 in Asparouhov and Muthén (2010) for updating the latent class variable. This way, the latent class variable is updated directly from Y . When Σ_k is not diagonal, C and Y^* are not grouped together because their joint conditional distribution is not explicit. They are two separate blocks that will be updated conditional on each other. Method 3 is not available and instead Method 2 is used which is less efficient and has the potential to create convergence problems due to high correlations between C and Y^* . This can happen not just on the population level, where the model is estimated, but also on the individual level. This would be difficult to detect as one would have to evaluate the mixing quality of the estimation for each individual. Furthermore, an individual may be clearly categorized in one of the classes where the posterior probability is 100%. This leads to no mixing, i.e., the latent variable remains constant during the MCMC estimation for that one individual. It would be difficult to distinguish if such an individual is "clearly categorized" or the MCMC estimation for that individual involving Y^* and C is "poorly mixing", due to for example poor starting values for C and Y^* (neither of which are provided as starting values of the estimation but are randomly generated; starting values can be provided in Mplus only for the model parameters). Because of these complexities when using Bayesian Mixture estimation based on Method 2 (which will be used with residual covariances), it is important to use long MCMC sequences, multiple chains, and repeating the estimation with a different random seed generation. The estimation quality can also be evaluated with Montecarlo simulation.

If the variables Y are binary, we can re-parameterize the model by replacing the threshold parameter in the model with the mean of Y^* (with

opposite sign), i.e., the threshold parameter can be assumed to be zero and then $[Y^*|C] \sim N(\mu_k, \Sigma_k)$, where $\mu_k = -(\tau_{1k1}, \dots, \tau_{pk1})$. The conditional distribution $[C|Y, Y^*, \mu_k, \Sigma_k]$ is the same as $[C|Y^*, \mu_k, \Sigma_k]$ because Y is determined by Y^* . Finally, the conditional distribution $[C|Y^*, \mu_k, \Sigma_k]$ is easy to compute because Y^* is multivariate normal. With ordered categorical variables, this construction does not carry over as there are multiple threshold parameters. In Mplus 8.7 we have implemented an approximation method which is designed to mimic the binary variables implementation. As a mean for Y^* we use the negative of the average threshold value, i.e., we use a reparameterization where the mean of Y^* is estimated but the threshold parameters are constrained to add up to zero. If $\tau_{pk.} = (\tau_{pk1} + \dots + \tau_{pkL_p})/L_p$ is the average threshold, and $\tau'_{pkl} = \tau_{pkl} - \tau_{pk.}$ are the centered thresholds then $[Y^*|C] \sim N(\mu_k, \Sigma_k)$ where $\mu_k = (-\tau_{1k.}, \dots, -\tau_{pk.})$. With this reparameterization, we approximate $[C|Y, Y^*, \mu_k, \Sigma_k, \tau'_{pkl}]$ with $[C|Y^*, \mu_k, \Sigma_k]$ and avoid using the computationally intensive multivariate probit function.

Note also that the non-diagonal Σ_k can be the result of multiple modeling features: direct modeling of covariances, continuous latent variables measured by the categorical variables, residual regressions as in RSEM, as well as combinations of any of the these. While these are different models, conceptually they result in the same framework when it comes to Bayesian Mixture estimation.

Next we illustrate the methodology with a simulation study where a 2 class categorical latent variable is measured by 7 categorical variables with 3 categories each. In addition, 3 residual correlations are introduced via residual regressions as in Figure 26. The results for some of the parameters are reported in Figure 27. Some small biases are visible in the estimates and these will not disappear asymptotically. The biases are due to the approximate nature of the computation, although with binary and continuous variables the computation is exact. The PPP value based on the chi-square test of fit within each class, which can detect residual correlations, has 0% rejection rate. The entropy in this example is 0.9. The Bayesian estimator appears to work well only when the entropy is on the high end, i.e., the classes are somewhat well separated. This is particularly so when Method 2 for the latent class updating is used. When the entropy is low, mixture models tend to have multiple solutions which are often blurred together by the Bayesian estimator. Label switching is also a possible estimation problem when the entropy is low and the sample size is small. Nevertheless, this simulation example illustrates that the Bayesian Mixture estimation is often feasible

and is especially valuable for those situations where the maximum-likelihood estimation is not available.

3.8 Pearson Posterior Predictive P-value

The standard chi-square PPP value can be used to evaluate the fit of RSEM and RSEM mixture models. With categorical data, however, this test of fit is limited to testing the model fit to the underlying latent variables Y^* and not to the observed categorical variables Y . In the Bayesian estimation framework, we can construct multiple PPP tests based on different fit functions. In Mplus 8.7 a new PPP is introduced based on the fit function formed by summing the Pearson statistics for all univariate and bivariate contingency tables. This fit function tests the model directly against the observed categorical data Y . The fit function is given by the following equation

$$F = N \sum_{p=1}^P \sum_{i=1}^{L_p} \frac{(o_{pi} - e_{pi})^2}{e_{pi}} + N \sum_{p_1=1}^P \sum_{p_2=1}^{p_1-1} \sum_{i_1=1}^{L_{p_1}} \sum_{i_2=1}^{L_{p_2}} \frac{(o_{p_1 p_2 i_1 i_2} - e_{p_1 p_2 i_1 i_2})^2}{e_{p_1 p_2 i_1 i_2}}. \quad (44)$$

Here N is the sample size, o_{pi} is the observed proportion of $Y_p = i$, e_{pi} is the model estimated probability $P(Y_p = i)$, $o_{p_1 p_2 i_1 i_2}$ is the observed proportion of $Y_{p_1} = i_1, Y_{p_2} = i_2$ in the bivariate observed joint distribution of Y_{p_1} and Y_{p_2} , and $e_{p_1 p_2 i_1 i_2}$ is the model estimated probability $P(Y_{p_1} = i_1, Y_{p_2} = i_2)$. The Pearson PPP can be obtained in Mplus with the specification **output:tech10**. The Pearson PPP obtained using the Figure 26 simulation study has 0% rejection rate.

To illustrate the Pearson PPP, using the same data generation as in Figure 26, we estimate the two-class model excluding the residual correlations. In that case, the chi-square PPP rejects the model in 100% of the replications while the Pearson PPP rejects the model in 97% of the replications. Using the data generation in Figure 26, we also estimate the single class model where all residual correlations are included in the model. This is essentially the unconstrained multivariate probit model for the observed variables. This model represents the most flexible model we can estimate without Mixture modeling. In that case, the chi-square PPP rejects the model in 0% of the replications while the Pearson PPP rejects the model in 100% of the replications. Note here, that such an outcome is not a contradiction. A Bayes test of fit can be constructed using any meaningful fit function. A model is acceptable fit for the data if all such tests yield acceptable PPP values. The

Figure 26: LCA analysis with conditional dependence simulation study

```
montecarlo:
  names = Y1-Y7; nobs = 2000; nreps = 100;
  generate=Y1-Y7(2); categorical=Y1-Y7;
  genclasses=c(2); classes=c(2);

analysis: estimator=bayes; proc=2; type=mixture;

model population:
  %overall%
  Y1^-Y3^ pon Y5^-Y7^*0.3;
  [c#1*0.5];
  %C#1%
  [Y1$1-Y4$1*-1]; [Y1$2-Y4$2*1];
  [Y5$1-Y7$1*0.5]; [Y5$2-Y7$2*1.5];
  %C#2%
  [Y1$1-Y4$1*0.5]; [Y1$2-Y4$2*1.5];
  [Y5$1-Y7$1*-1]; [Y5$2-Y7$2*0];

model:
  %overall%
  Y1^-Y3^ pon Y5^-Y7^*0.3;
  [c#1*0.5];
  %C#1%
  [Y1$1-Y4$1*-1]; [Y1$2-Y4$2*1];
  [Y5$1-Y7$1*0.5]; [Y5$2-Y7$2*1.5];
  %C#2%
  [Y1$1-Y4$1*0.5]; [Y1$2-Y4$2*1.5];
  [Y5$1-Y7$1*-1]; [Y5$2-Y7$2*0];

output:tech10;
```


Figure 27: LCA analysis with conditional dependence simulation study results

MODEL RESULTS		Population	ESTIMATES		S. E.	M. S. E.	95% Cover	% Sig	Coeff
			Average	Std. Dev.	Average				
Y1^	ON								
Y5^		0.300	0.2824	0.0414	0.0377	0.0020	0.900	1.000	
Y2^	ON								
Y6^		0.300	0.2873	0.0346	0.0376	0.0013	0.990	1.000	
Y3^	ON								
Y7^		0.300	0.2750	0.0386	0.0376	0.0021	0.900	1.000	
Latent Class 1									
Thresholds									
Y1\$1		-1.000	-0.9679	0.0446	0.0486	0.0030	0.920	1.000	
Y1\$2		1.000	0.9624	0.0444	0.0444	0.0034	0.830	1.000	
Y2\$1		-1.000	-0.9729	0.0430	0.0484	0.0026	0.930	1.000	
Y2\$2		1.000	0.9704	0.0437	0.0449	0.0028	0.920	1.000	
Latent Class 2									
Thresholds									
Y1\$1		0.500	0.5039	0.0566	0.0569	0.0032	0.950	1.000	
Y1\$2		1.500	1.6354	0.0880	0.0866	0.0260	0.670	1.000	
Y2\$1		0.500	0.4857	0.0570	0.0561	0.0034	0.950	1.000	
Y2\$2		1.500	1.6091	0.0837	0.0837	0.0188	0.740	1.000	
Categorical Latent Variables									
Means									
C#1		0.500	0.5256	0.0562	0.0536	0.0038	0.910	1.000	

unrestricted multivariate probit model is the saturated single class model and therefore we can expect that the model will never be rejected by the chi-square PPP value. The model, however, has very few parameters when compared to full contingency table for the categorical variables. When the unrestricted multivariate probit model is rejected by the Pearson PPP, the only option for improving the model fit is to introduce latent classes in the model. This further emphasizes the need to efficiently estimate LCA models with conditional dependence.

The Pearson statistic based on the univariate and bivariate contingency tables can be used to compare models across different estimators. In Mplus, all estimators (ML, WLSMV and Bayes) compute the Pearson statistic in the tech10 output. The Pearson PPP in the Bayes estimator is the only case, however, that provides a direct test of fit procedure. With the ML and WLSMV estimators, the Pearson statistic is not a chi-square statistic and it will not provide a formal p-value. Those statistics can be used, however, for comparative purposes, among different estimators and models, similarly to how BIC criterion is used. At this time, however, we have no way of determining what constitutes a significant improvement in the statistic. The full multivariate Pearson statistic does provide a formal chi-square test but that procedure is useful only when the multivariate contingency table is smaller. For larger models, the multivariate contingency table has very large number of parameters, compared to the estimated model, and the Pearson test will have very large DF. In that case, the asymptotic argument that supports the Pearson testing will require extremely large sample size, beyond what is practical. The tech10 output contains also very detailed information for all quantities used in (44). Standardized residuals are computed for every cell in the univariate, bivariate and multivariate contingency tables. These can be used to pinpoint where a model is inadequate representation of the data.

If data is missing and it is not missing completely at random (MCAR), the utilization of the Pearson PPP and the standardized residuals in the contingency tables is somewhat complex. Two things can occur when data is missing at random (MAR) and not MCAR. First, the observed proportions need not agree with the estimated proportions even when the model is correct. Observed proportions in the univariate and bivariate tables are essentially obtained with listwise deletion, i.e., using an inferior estimation method. We can expect that the multivariate model estimated quantities, which yields unbiased estimated under MAR, will be more accurate than the observed proportions. Therefore, discrepancies between the observed and es-

estimated quantities can be due to strong MAR effects rather than to incorrect model. The second thing that occurs is that the Pearson PPP may not perform well. The Pearson PPP is based on comparing the Pearson statistic for the sample data and for hypothetical (replicated) data generated from the estimated model. We have no way of producing similar MAR missing data for the replicated data, since the missing data mechanism is not estimated with likelihood based approaches. Therefore, the replicated data would have MCAR missing data and the two data sets would not be entirely comparable. Thus, if the missing data is MAR, the discrepancy between the observed and the estimated contingency tables can artificially cause the Pearson PPP to reject the model, even when the model is correct. In practical applications, however, this is a relatively unlikely scenario. If the amount of missing data is relatively small or the MAR effects are not very strong, the Pearson PPP is expected to perform well. In the rare situation where the amount of missing data is large and strong MAR effects are suspected, the Pearson PPP should not be used. Further discussion on Bayesian PPP methodology in the presence of MAR missing data is available in Asparouhov and Muthén (2021).

The Pearson PPP and the chi-square PPP may often disagree in terms of model fit. Consider the case when the Pearson PPP does not reject the model but the chi-square PPP rejects the model. In this case, the unrestricted correlation model will fit both the Pearson PPP and the chi-square PPP. That means that some modifications in the structural model can be implemented to resolve the chi-square PPP rejection. When the unrestricted correlation model is not rejected by both PPP procedures we can assume that this model fits the data well. In that case, i.e., when the multivariate probit model is true, the chi-square PPP has bigger power to detect misspecifications as it is more directly connected to the concept of polychoric correlation., which need to be fitted. The Pearson PPP will consider the fit of many more quantities, than the chi-square PPP and as a results of that some model misfits will be washed off as insignificant. In the case when the Pearson PPP does not reject the model but the chi-square PPP rejects the model, the results can be traced to the lower power of the Pearson PPP when the multivariate probit model is correct.

Note also that, models which include categorical and continuous dependent variables may have different PPP outcomes for a different reason. The Pearson PPP tests the contingency tables for the categorical variables only, while the chi-square PPP tests the entire model. If the Pearson PPP does not

reject the model but the chi-square PPP rejects the model, the reason may be that the continuous part of the model is misfitted or that the part of the model that correlates the categorical variables with the continuous variables is misfitted. In that case, the difference between the two PPP procedures may not be due to the difference in power at all.

Next, consider the case when the Pearson PPP rejects the model but the chi-square PPP does not. To decide how to proceed in that case, one needs to estimate the unrestricted correlation model and obtain the Pearson PPP for that model. If the Pearson PPP does not reject that model, modifications in the structural model can be implemented to resolve the Pearson PPP rejection. If, however, the unrestricted correlation model is rejected by the Pearson PPP, no model modification will be available to resolve the problem. The cause of the Pearson PPP rejection can then be traced in the failure of the multivariate probit model. There are two different interpretations in that case, which are actually mathematically equivalent. The first interpretation is that the Pearson PPP rejects the model because the underlying continuous variables Y^* does not have a multivariate normal distribution. The second interpretation is that a single-class structural model does not provide a good fit for the data and a mixture model with more than one class must be estimated. The reason, the two interpretations are equivalent is because any non-normal multivariate distribution (for Y^*) can be approximated by a mixture of multivariate normal distributions. Regardless of which interpretation is used, however, the only option in this case is to use mixture modeling, i.e., increase the number of classes until the Pearson PPP does not reject the model. Further discussion on testing categorical variable models is available in Muthén (1993).

3.9 Mixture RI-AR model with categorical variables

In this section we illustrate the Bayesian Mixture methodology with a more advanced mixture model: a two-class mixture RI-AR model where all the dependent variables are ordered categorical variables. In this model, there are three completely different and competing modeling components attempting to model the dependencies among the categorical variables. These three modeling components are: the latent class variable, the random intercept and the auto-regressive residuals. Separating these three sources of correlational dependencies generally requires a fairly rich data set and in many practical situations this might not be available. This kind of analysis, should always be

preceded by analyzing the data first with one and then two of the modeling components. In addition, we recommend that any mixture RI-AR analysis with ordered categorical data be accompanied by a Montecarlo simulation study that verifies the quality of the estimation.

As in the previous example, the model can be estimated only with the Bayesian estimator. Figure 28 shows the Mplus input file for a mixture RI-AR simulation study and Figure 29 gives the results for some of the model parameters. As in the previous case, small biases are visible in some of the parameter estimates but the overall performance is satisfactory and all three of the correlation components were allocated properly. In this simulation study, we held the auto-regressive parameter equal across time and across classes. This helps with the model identification, although it is not necessary in general. Also, in this simulation study, the effect of the latent class variable and the random intercept on the observed variables are orthogonal to each other. The random intercept always weighs equally on all indicators. If the latent class variable does so as well, it will be difficult to distinguish between the effect of C and the effect of I . In the simulation study given in Figure 28, the orthogonality is achieved by giving completely different sets of threshold parameters for indicators $Y_1 - Y_4$ and $Y_5 - Y_8$. Because of that, the latent class variable affects the class indicators differently and it would be distinguishable from the effect of the random intercept variable.

Figure 28: Mixture RI-AR with ordered categorical simulation study

```
montecarlo:
  names = y1-y8; nobs = 2000; nreps = 100;
  generate=y1-y8(2); categorical=y1-y8;
  genclasses=c(2); classes=c(2);

analysis: estimator=bayes; proc=2; type=mixture;

model population:
  %overall%
  Y2^-Y8^ pon Y1^-Y7^*0.225 (ar);
  i by y1-y8@1; [i@0]; i*0.4;
  [c#1*-0.4 ];
  %C#1%
  [Y1$1-Y4$1*-1]; [Y1$2-Y4$2*1];
  [Y5$1-Y8$1*1]; [Y5$2-Y8$2*2];
  %C#2%
  [Y1$1-Y4$1*1]; [Y1$2-Y4$2*2];
  [Y5$1-Y8$1*-1]; [Y5$2-Y8$2*0];

model:
  %overall%
  Y2^-Y8^ pon Y1^-Y7^*0.225 (ar);
  i by y1-y8@1; [i@0]; i*0.4;
  [c#1*-0.4 ];
  %C#1%
  [Y1$1-Y4$1*-1]; [Y1$2-Y4$2*1];
  [Y5$1-Y8$1*1]; [Y5$2-Y8$2*2];
  %C#2%
  [Y1$1-Y4$1*1]; 70[Y1$2-Y4$2*2];
  [Y5$1-Y8$1*-1]; [Y5$2-Y8$2*0];
```

Figure 29: Mixture RI-AR with ordered categorical simulation study results

MODEL RESULTS			ESTIMATES		S. E.	M. S. E.	95% % Sig
		Population	Average	Std. Dev.	Average		Cover Coeff
Y2^	ON						
Y1^		0.225	0.2287	0.0174	0.0191	0.0003	0.960 1.000
Variances							
I		0.400	0.3841	0.0269	0.0273	0.0010	0.900 1.000
Latent Class 1							
Thresholds							
Y1\$1		-1.000	-0.9250	0.0655	0.0628	0.0099	0.740 1.000
Y1\$2		1.000	0.9864	0.0643	0.0600	0.0043	0.910 1.000
Y2\$1		-1.000	-0.9349	0.0588	0.0636	0.0077	0.820 1.000
Y2\$2		1.000	0.9846	0.0586	0.0604	0.0036	0.920 1.000
Latent Class 2							
Thresholds							
Y1\$1		1.000	0.9782	0.0484	0.0520	0.0028	0.940 1.000
Y1\$2		2.000	2.0302	0.0785	0.0801	0.0070	0.920 1.000
Y2\$1		1.000	0.9768	0.0522	0.0522	0.0032	0.940 1.000
Y2\$2		2.000	2.0321	0.0763	0.0784	0.0068	0.950 1.000
Categorical Latent Variables							
Means							
C#1		-0.400	-0.3704	0.0454	0.0493	0.0029	0.930 1.000

3.10 Structural residual modeling in RDSEM

The residual dynamic structural equation model (RDSEM), as defined in Asparouhov et al. (2018), allows for lagged modeling of the residuals. The residuals from the current time period in RDSEM can be predicted by residuals from the previous periods. The model, however, did not allow for the residuals to be structurally modeled within the same period (contemporaneous relations). Such modeling however is sometimes necessary, see Hamaker et al. (2021). In Mplus 8.7 we have extended the RDSEM model to accommodate such structural models among the residuals. This extension applies to single and two-level RDSEM models. Since the two-level RDSEM model can be viewed as a generalization of the standard two-level SEM model (where the lag is 0), this extension can be viewed as the two-level version of RSEM. Both categorical and continuous variables can be used in this RDSEM extension.

The RDSEM model allows for contemporaneous relations not just between the residuals but also between the variables. In principle, this creates a methodological challenge. When a contemporaneous relationship between two variables must be modeled, it may not be clear whether the relationship should be between the variables or whether it should be between the residuals. In some practical applications, a compelling substantive argument may be available to make that choice. When a substantive argument is not available, one can use statistical techniques, such as the DIC criterion, for guidance in that decision.

We illustrate the contemporaneous residual modeling in RDSEM with a bivariate example that mimics the modeling in Hamaker et al. (2021). Suppose that Y_{pit} is the p -th variable, $p = 1, 2$, for individual i at time t . Suppose that X_{it} is a covariate for individual i observed at time t . The RDSEM model is given by the following equations

$$Y_{pit} = \alpha_{pi} + \beta_{pi}X_{it} + \hat{Y}_{pit} \quad (45)$$

$$\hat{Y}_{1it} = \gamma_i \hat{Y}_{2it} + r_{1i} \hat{Y}_{1i,t-1} + \varepsilon_{1it} \quad (46)$$

$$\hat{Y}_{2it} = r_{2i} \hat{Y}_{2i,t-1} + r_{3i} \hat{Y}_{1i,t-1} + \varepsilon_{2it}. \quad (47)$$

There are 8 random effects in this model: α_{pi} and β_{pi} are the random intercept and slope in the regression of Y_{pit} on X_{it} , γ_i represents the contemporaneous relationship between the residuals \hat{Y}_{1it} and \hat{Y}_{2it} , and r_{ji} represent the lagged relationships in the residuals. Figure 30 shows a simulation study setup for

this model. Figure 31 shows the results of the simulation study. The Bayesian estimation performs well in this example.

4 Conclusion

In this article, we illustrate how the residual variables in a structural equation model can be used for constructing a secondary structural model. The residual variables can also be used as additional variables and predictors in the original model. Using this residual structural equation modeling framework, we enhance the utility of the primary structural model. The new Mplus hats language notation, previously used only for RDSEM, is expanded in Mplus 8.7 to standard SEM models. The hats notation allows us to construct RSEM models in an efficient and compact way. In addition, the hats notation will guarantee that the most efficient estimation method is used, particularly so with the Bayesian estimator. This compact presentation will also help with proper model interpretation and conceptualization.

Figure 30: Contemporaneous residual modeling in RDSEM simulation study

```
MONTECARLO:
NAMES ARE y1-y2 x;
NOBS = 5000; NREP = 100;
NCSIZES = 1; CSIZES = 100(50);
lagged=y1(1) y2(1);
within=x;

ANALYSIS:    TYPE IS TWOLEVEL RANDOM;
              estimator=bayes; proc=2;

MODEL MONTECARLO:

%WITHIN%
y1-y2*1 x*1;
b1 | y1 on x;
b2 | y2 on x;
g | y1^ on y2^;
r1 | y1^ on y1^1;
r2 | y2^ on y2^1;
r3 | y2^ on y1^1;

%BETWEEN%
[y1*1 y2*2 b1*0.8 b2*-0.5 g*0.2 r1*0.3 r2*0.4 r3*0.2 ];
y1*0.5; y2*0.7 b1*0.2 b2*0.2 g*0.02 r1*0.02 r2*0.02 r3*0.02;

MODEL:

%WITHIN%
y1-y2*1;
b1 | y1 on x;
b2 | y2 on x;
g | y1^ on y2^;
r1 | y1^ on y1^1;
r2 | y2^ on y2^1;
r3 | y2^ on y1^1;

%BETWEEN%
74
[y1*1 y2*2 b1*0.8 b2*-0.5 g*0.2 r1*0.3 r2*0.4 r3*0.2 ];
y1*0.5; y2*0.7 b1*0.2 b2*0.2 g*0.02 r1*0.02 r2*0.02 r3*0.02;
```

Figure 31: Contemporaneous residual modeling in RDSEM simulation study results

MODEL RESULTS							
	Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Within Level							
Residual Variances							
Y1	1.000	1.0035	0.0258	0.0211	0.0007	0.920	1.000
Y2	1.000	1.0203	0.1378	0.0215	0.0192	0.920	1.000
Between Level							
Means							
Y1	1.000	1.0027	0.0771	0.0764	0.0059	0.930	1.000
Y2	2.000	1.9938	0.0904	0.0915	0.0081	0.950	1.000
B1	0.800	0.8036	0.0504	0.0476	0.0025	0.920	1.000
B2	-0.500	-0.5032	0.0474	0.0487	0.0022	0.950	1.000
G	0.200	0.2042	0.0185	0.0202	0.0004	0.960	1.000
R1	0.300	0.2975	0.0164	0.0205	0.0003	0.990	1.000
R2	0.400	0.4009	0.0196	0.0203	0.0004	0.950	1.000
R3	0.200	0.2019	0.0203	0.0210	0.0004	0.940	1.000
Variances							
Y1	0.500	0.5023	0.0899	0.0851	0.0080	0.900	1.000
Y2	0.700	0.7211	0.1156	0.1224	0.0137	0.940	1.000
B1	0.200	0.2086	0.0290	0.0338	0.0009	0.960	1.000
B2	0.200	0.2189	0.0779	0.0352	0.0064	0.920	1.000
G	0.020	0.0222	0.0060	0.0058	0.0000	0.870	1.000
R1	0.020	0.0217	0.0052	0.0059	0.0000	0.960	1.000
R2	0.020	0.0210	0.0052	0.0055	0.0000	0.930	1.000
R3	0.020	0.0243	0.0147	0.0063	0.0002	0.890	1.000

5 Appendix

In this section we discuss the equivalence of the RI-MEAR and RI-ARMA models. We determine the parameter space for which this equivalence holds and show how to obtain the parameters for one model from the parameters of the other model. Out of the total $4T - 2$ model parameters, $2T$ parameters: α_t , ρ_t and ψ are identical between the two models. That leaves us with the $2T - 2$ parameters (σ_t and v_t) of the RI-MEAR model and the $2T - 2$ parameters (θ_t and β_t) of the RI-ARMA model. We want to show that there is a reparameterization of these $2T - 2$ parameters for which the two models are equivalent.

The models will be equivalent if the variance covariance matrix for \hat{Y}_{it} in the RI-MEAR model is the same as the variance covariance matrix for ε_{it} in the RI-ARMA model. This is because the random intercept part of the model is the same for the two models. For convenience, in this discussion we drop the index i which refers to the individual i data and we will refer to \hat{Y}_{it} as simply \hat{Y}_t , etc.

First, we algebraically rewrite the RI-MEAR model as follows. For $t = 1$

$$\hat{Y}_1 = \xi_1. \quad (48)$$

For $t = 2$

$$\hat{Y}_2 = \rho_2 \hat{Y}_1 + e_2 + \xi_2. \quad (49)$$

For $t > 2$

$$\hat{Y}_t = \rho_t \hat{Y}_{t-1} + e_t + \xi_t - \rho_t e_{t-1}. \quad (50)$$

The RI-MEAR model can be given in matrix form as follows. Define A to be the following matrix

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ -\rho_2 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -\rho_3 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -\rho_4 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -\rho_T & 1 \end{pmatrix}.$$

Let \hat{Y} be the vector of all hat variables \hat{Y}_t . The RI-MEAR model becomes

$$A\hat{Y} = \begin{pmatrix} \xi_1 \\ \xi_2 + e_2 \\ \xi_3 + e_3 - \rho_3 e_2 \\ \xi_4 + e_4 - \rho_4 e_3 \\ \vdots \\ \xi_T + e_T - \rho_T e_{T-1} \end{pmatrix}.$$

Similarly, the RI-ARMA model can be written as

$$A\varepsilon = \begin{pmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \hat{\varepsilon}_3 + \beta_3 \hat{\varepsilon}_2 \\ \hat{\varepsilon}_4 + \beta_4 \hat{\varepsilon}_3 \\ \vdots \\ \hat{\varepsilon}_T + \beta_T \hat{\varepsilon}_{T-1} \end{pmatrix}.$$

The variance covariance of \hat{Y} and ε will be the same when the variance covariance matrices of the RHS vectors in the above two equations are the same. Both of these vectors have fairly simple variance covariance matrices and both take the form of MA (moving-average) processes. Inspecting both vectors reveals that there are precisely $2T - 2$ non-zero entries in these MA variance covariance matrices. If we denote that variance covariance matrix by Ω , the non-zero entries are the T diagonal entries $\omega(t, t)$, and the $T - 2$ off diagonal entries $\omega(t, t - 1)$ for $t > 2$. All other entries are zero for both vectors. Setting these $2T - 2$ variance covariance elements to be equal between the two models provides precisely the $2T - 2$ equations needed for the reparameterization between the RI-ARMA and the RI-MEAR models. These equations are given below. For $t = 1$

$$\omega(1, 1) = \theta_1 = v_1. \quad (51)$$

For $t = 2$

$$\omega(2, 2) = \theta_2 = v_2 + \sigma_2. \quad (52)$$

For $t > 2$

$$\omega(t, t) = \theta_t + \beta_t^2 \theta_{t-1} = v_t + \sigma_t + \rho_t^2 \sigma_{t-1} \quad (53)$$

$$\omega(t, t - 1) = \beta_t \theta_{t-1} = -\rho_t \sigma_{t-1} \quad (54)$$

The RI-MEAR parameters can be obtained from the RI-ARMA parameters as follows

$$v_1 = \theta_1 \quad (55)$$

$$v_2 = \theta_2(\rho_3 + \beta_3)/\rho_3. \quad (56)$$

For $t = 3, \dots, T - 1$

$$v_t = \theta_t(\rho_{t+1} + \beta_{t+1})/\rho_{t+1} + (\rho_t + \beta_t)\beta_t\theta_{t-1}. \quad (57)$$

For $t = T$

$$v_t = \theta_t + (\rho_t + \beta_t)\beta_t\theta_{t-1}. \quad (58)$$

For $t = 2, \dots, T - 1$

$$\sigma_t = -\beta_{t+1}\theta_t/\rho_{t+1}. \quad (59)$$

Conversely, the RI-ARMA parameters can be obtained from the RI-MEAR parameters in the following recursive sequence

$$\theta_1 = v_1 \quad (60)$$

$$\theta_2 = v_2 + \sigma_2. \quad (61)$$

For $t > 2$

$$\beta_t = -\rho_t\sigma_{t-1}/\theta_{t-1} \quad (62)$$

$$\theta_t = v_t + \sigma_t + \rho_t\sigma_{t-1}(\beta_t + \rho_t). \quad (63)$$

One key question about this reparameterization is whether all variance component parameters will remain positive. The ML estimator, unless restricted, can estimate negative residual variances, even though such a thing would not be interpretable. Thus, the RI-ARMA to RI-MEAR conversion can always be done as long as we allow residual variances to be negative. In general, however, if one of the two models has negative residual variances but the other does not, one should clearly prefer the model with all positive residual variances.

Here we analyze this issue through the above conversion formulas. We restrict ourselves to the most common scenario $\rho_t \geq 0$. First, we consider the situation when the RI-ARMA model is estimated and all variance components are positive. We want to know under what circumstance the RI-MEAR variance components will also be positive. Equation (59) implies that all β_t (moving average) parameters in the RI-ARMA model must be negative to produce positive σ_t . The situation with v_t , however is much more complex.

In DSEM-ARMA to DSEM-MEAR conversion, one of the parameter requirements is that $\beta + \rho > 0$. In this situation, however, assuming $\beta_t + \rho_t > 0$, $\beta_t < 0$ for every t , would not be enough to guarantee positive v_t . That is because in equation (57) the second term is negative. The additional equality that must be satisfied by the RI-ARMA parameters to guarantee positive v_t is as follows

$$\frac{\theta_t}{\theta_{t-1}} > \frac{\rho_t + \beta_t}{\rho_{t+1} + \beta_{t+1}}(-\beta_t)\rho_{t+1}. \quad (64)$$

If all parameters are time-invariant and the process is stationary and invertible, i.e. $\rho < 1$ and $\beta < 1$, the above inequality is satisfied.

The opposite situation is simpler. If we estimate the RI-MEAR model and all variance components are positive, the variance components in the RI-ARMA model are always guaranteed to be positive. To see this, going through equations (60-63) we can inductively establish the following inequalities: $\theta_t > \sigma_t$, $\beta_t < 0$, $\beta_t + \rho_t > 0$. Since σ_t is positive, θ_t must be positive as well.

With the Bayesian estimator negative variances can not be estimated. This means that the RI-MEAR and RI-ARMA equivalence in the Bayesian framework is subject to the parameter space restriction where the RI-ARMA to RI-MEAR conversion yields positive variance components. Furthermore, the inequality constraints must be satisfied for the entire posteriors distribution and not just for the point estimates. If for a portion of the posterior distribution of the RI-ARMA parameters, the implied RI-MEAR parameters have negative variances, the equivalence between the RI-ARMA and RI-MEAR will not hold. This is particularly consequential when the sample size is small or moderate and the posterior distributions are wide enough to cross over in the parameter space that doesn't support the RI-ARMA to RI-MEAR conversion.

The RI-ARMA and RI-MEAR equivalence is very specific to the model we discussed above. It applies precisely when all parameters are non-invariant and there are exactly $4T - 2$ parameters. In practical settings, some simplifications of these models may be desirable, such as holding parameters equal across time or fixing some insignificant parameters to zero. Such simplifications do not translate from one of the models to the other. A simplified RI-ARMA model may not result in a simplified RI-MEAR model and vice versa. The reparametrization formulas given above must be used to determine how parameter restrictions in one model translate into parameter restriction for the other model. An example of such a situation is given in

Figure 11. In that figure, the models are applied to categorical data, which require residual variances to be fixed to 1. Residual variances fixed to 1 in one of the models then results in a completely different restriction for the other model.

References

- [1] Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 25 , 359-388.
- [2] Asparouhov, T & Muthén, B. (2008) Auxiliary Variables Predicting Missing Data. Mplus Technical Appendices. <http://www.statmodel.com/download/AuxM2.pdf>
- [3] Asparouhov, T. & Muthén, B. (2010). Bayesian analysis using Mplus: Technical implementation. Technical appendix. Los Angeles: Muthén & Muthén <http://www.statmodel.com/download/Bayes3.pdf>
- [4] Asparouhov, T., & Muthén, B. O. (2011). Using Bayesian priors for more flexible latent class analysis. In *Joint Statistical Proceedings, Government Statistics Section* (pp. 4979–4993). Alexandria, VA: American Statistical Association.
- [5] Asparouhov, T., & Muthén, B. (2015). Residual associations in latent class and latent transition analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(2), 169-177.
- [6] Asparouhov, T., & Muthén, B. (2016). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. In J. R. Harring, L. M. Stapleton, & S. N. Beretvas (Eds.), *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications* (pp. 163–192). Charlotte, NC: Information Age.
- [7] Asparouhov, T., & Muthén, B. (2019). Latent variable centering of predictors and mediators in multilevel and time-series models. *Structural Equation Modeling: A Multidisciplinary Journal*, 26, 119–142.
- [8] Asparouhov, T., & Muthén, B. (2020). Comparison of models for the analysis of intensive longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 27 , 275-297.
- [9] Asparouhov, T., & Muthén, B. (2021). Advances in Bayesian model fit evaluation for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 28, 1–14.

- [10] Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeier et al. *Journal of Management*, 41, 1561-1577.
- [11] Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2013). The separation of between-person and within-person components of individual change over time: A latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology*, 82, 879–894.
- [12] Bollen, K. A., & Brand, J. E. (2010). A General Panel Model with Random and Fixed Effects: A Structural Equations Approach. *Social Forces*, 89(1), 1-34.
- [13] Box, G.E.P. & J.M. Jenkins (1976) *Time series analysis: Forecasting and control*. Holden-Day, San Francisco, CA.
- [14] Cole, D. A., Martin, N. C., & Steiger, J. H. (2005). Empirical and conceptual problems with longitudinal trait-state models: Introducing a trait-state-occasion model. *Psychological Methods*, 10, 3-20.
- [15] Du Toit, S. H. C., & Browne, M. W. (2001). The covariance structure of a vector ARMA time series. In R. Cudeck, S. Du Toit, & D. Sörbom (Eds), *Structural equation modeling: Present and future* (pp. 279–314). Lincolnwood, IL: Scientific Software International.
- [16] Granger, C. W. J., & Morris, M. J. (1976). Time series modelling and interpretation. *Journal of the Royal Statistical Society, Series A*, 139, 246–257.
- [17] Hamaker, E.L. (2018) How to run the RI-CLPM with Mplus.
<https://www.statmodel.com/download/RI-CLPM%20Hamaker%20input.pdf>
- [18] Hamaker, E.L., Asparouhov, T, & Muthén, B. (2021). Dynamic structural equation modeling as a combination of time series modeling, multilevel modeling, and structural equation modeling. To be published as Chapter 31 in: *The Handbook of Structural Equation Modeling* (2nd edition); Rick H. Hoyle (Ed.); Publisher: Guilford Press.

- [19] Hamaker, E.L., Kuiper R.M., Grasman, R.P. (2015) A critique of the cross-lagged panel model. *Psychol Methods* 20:102–116.
- [20] Joreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, 43, 443-477.
- [21] Kenny, D. A., & Zautra, A. (1995). The trait–state– error model for multiwave data. *Journal of Consulting and Clinical Psychology*, 63, 52–59.
- [22] Mulder, J. D., & Hamaker, E. L. (2021). Three extensions of the random intercept cross-lagged panel model. *Structural Equation Modeling*, 28 (4), 638-648. Supplementary materials at <https://jeroendmulder.github.io/RI-CLPM/>
- [23] Muthén, B. O. (1993). Goodness of fit test with categorical and other nonnormal variables. In K. A. Bollen and J. S. Long (Eds.), *Testing structural equation models* (pp. 205-234). Newbury Park: Sage.
- [24] Muthén, B. O., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus Web Note No.4*.
<http://www.statmodel.com/download/webnotes/CatMGLong.pdf>
- [25] Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17, 313-335.
- [26] Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus User’s Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén.
- [27] Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (2021). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology*, 120(4),1013-1034.
- [28] Sörbom, D (1989). Model modification. *Psychometrika* 54, 371–384.
- [29] Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, 24(5), 637-657.

- [30] Usami, S. (2021). On the differences between general cross-lagged panel model and random-intercept cross-lagged panel model: Interpretation of cross-lagged parameters and model choice. *Structural Equation Modeling*, 28 (3), 331–344.
- [31] Van Iddekinge, C. H., Ferris, G. R., Perrewe, P. L., Perryman, A. A., Blass, F. R., & Heetderks, T. D. (2009). Effects of selection and training on unit-level performance over time: A latent growth modeling approach. *Journal of Applied Psychology*, 94(4), 829-843.
- [32] Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., . . . Diener, E. (2020a). From data to causes I: Building A general cross-lagged panel model (GCLM). *Organizational Research Methods*, 23(4), 651-687.
- [33] Zyphur, M. J., Voelkle, M. C., Tay, L., Allison, P. D., Preacher, K. J., Zhang, Z., . . . Diener, E. (2020b). From data to causes II: Comparing approaches to panel data analysis. *Organizational Research Methods*. 23(4), 688-716.