

# Utrecht University Summer School S20

## Exercises

Jeroen D. Mulder<sup>1</sup>, Ellen Hamaker<sup>1</sup>, & Caspar J. van Lissa<sup>1</sup>

<sup>1</sup>Utrecht University, Methodology & Statistics



# Contents

<b>Introduction</b>	<b>5</b>
<b>Exercises - Day 4</b>	<b>7</b>
The mean-structure . . . . .	8
The CLPM . . . . .	9
The RI-CLPM . . . . .	12



# Introduction

These course materials are part of the course Introduction to Structural Equation Modeling using Mplus (S20). It is a five-day summer school course hosted by Utrecht University's department of Methodology and Statistics. The main objective is to learn how to analyze several models with Mplus (e.g. path models, multiple group models, mediation and moderation, confirmatory factor analysis, and longitudinal models). More in-depth lectures on the fundamentals of Mplus and advanced longitudinal models are discussed in the advanced course (S23).

Using the menu on the left, you can navigate to exercises of day 4 of S20.



# Exercises - Day 4

These exercises belong to day 4 of S20, "Traits, Measurement Errors, and Time-Lags". We are going to analyze the same data that were used in the Hamaker et al. (2015) paper and were presented in the lecture. The goals of these exercises are to get some experience with fitting (random intercept) cross-lagged panel models, interpreting their results, and comparing them. Here, we focus on the relationship between *Parental Responsiveness* and *Adolescents' Depressive Symptomatology* (note, in the lecture the focus was on a different concept, namely *Parental Psychological Control*).

The data (means, standard deviations, and the correlation matrix) are included in the file `Soenens.dat`. The number of observations is 396. The data command for these data should be:

```
DATA:
  TYPE = MEANS STDEVIATIONS CORRELATION;
  FILE = Soenens.dat;
  NOBSERVATIONS = 396;
```

The variable command should be:

```
VARIABLE:
  NAMES = PsCon1-PsCon3 Res1-Res3 BeCon1-BeCon3 Dep1-Dep3;
  USEVARIABLES = Res1 Dep1 Res2 Dep2 Res3 Dep3;
```

Do the exercises. You can navigate through them using the left-hand menu or the arrows on left- and right-hand side of this screen. Answers are included below each exercise. The Mplus input files can be found in the Summer School Dropbox-folder (`./Dropbox/Lab material - Day 4/S20 - Day 4 - 4MplusInputFiles`).

## The mean-structure

### Exercise 1

Specify a model in which you investigate whether the group *means* can be constrained across time. Does this model fit? Tip: allow all variables to covary with each other (i.e. putting no constraints on the covariance structure) such that we are estimating means, and not intercepts.

[Click to show answers](#)

The model specification is given below. The model does not fit well, according to the  $\chi^2$  and the RMSEA; CFI and TLI indicate a good and barely acceptable model fit, respectively. See the slides of day 1 of this summer school if you need a refresher on the recommended cutoff values for these fit indices (although such cutoff recommendations remain a hotly debated topic among statisticians).

- $\chi^2(4) = 19.620, p < .001$
- RMSEA = 0.099
- CFI = 0.986
- TLI = 0.946

### Exercise 2

What can you do to improve the model fit?

[Click to show answers](#)

We can look at the modification indices in the output. Here, we specified `MOD(4)` in the `OUTPUT` command, which indicates all parameters that can be added/freed that will lead to a decrease in chi-square of at least 4 points (which is a significant improvement combined with 1 *df*).

Based on the Mplus output, it appears that freeing the mean of `Dep1` leads to the largest change in chi-square (M.I. = 10.53).

### Exercise 3

Run the adjusted model and discuss the results.

[Click to show answers](#)

Syntax for the adjusted model is in `Means2.inp`. Looking at the fit indices of the adjusted model, the chi-square for this model is still significant, but all other measures indicate the model fit is okay to excellent. Moreover, the model is a significant improvement in comparison to the previous model:  $\Delta\chi^2 = 19.62 - 8.71 = 10.91$ , with  $\Delta df = 1$  and,  $p = .001$ .



You can calculate the  $p$ -value of the  $\Delta\chi^2$  using the `pchisq()`-function in R (with the `lower.tail` argument set to `FALSE`), or an online tool

## The CLPM

### Exercise 4

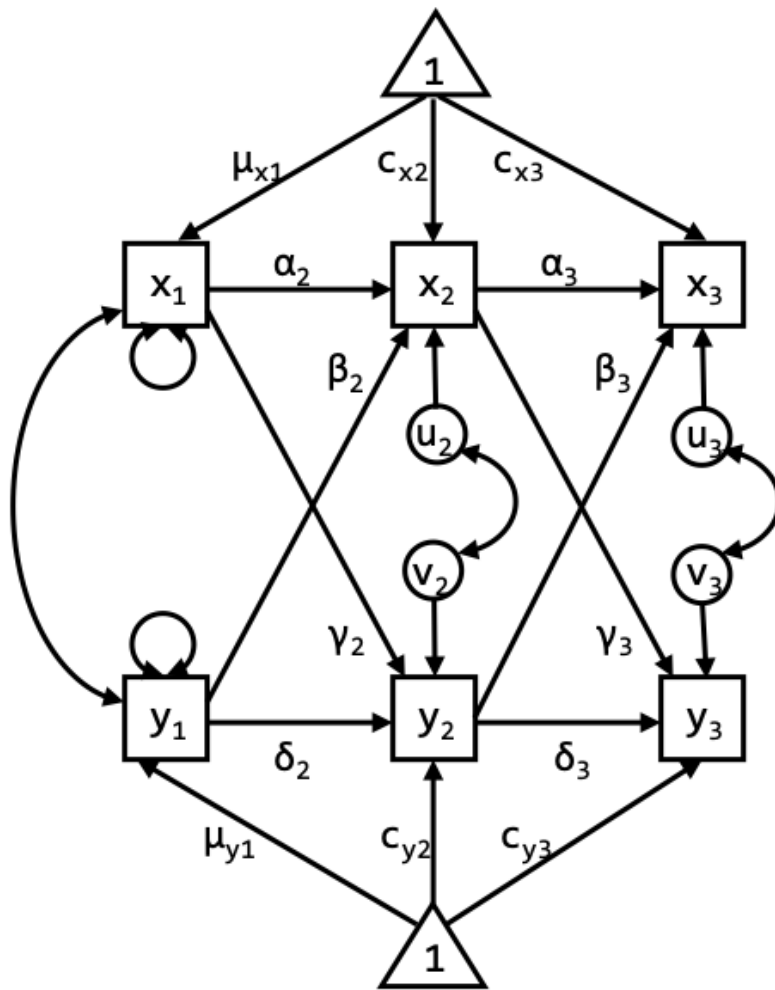


Figure 1: The cross-lagged panel model.

Specify the CLPM in Mplus. Regress the observed variables directly on each other over time, as in the graph above. Try to think of where each parameter

you estimate should go in the graph.

Click to show answers

The input for this models can be found in `CLPM.inp`.

### Exercise 5

Specify the model as in the graph below (the CLPM with centered latent variables), and again include the (significant) parameter estimates in this graph. Tip: do not forget to constrain the measurement error variances of the observed variables to 0.

Click to show answers

The input for this models can be found in `CLPMalt.inp`.

### Exercise 6

The two models should lead to the exact same model fit. Estimate the model and discuss the model fit. When comparing the parameter estimates, what is the difference between the model in question 4 and the model in question 5?

Click to show answers

Model 5 is, just like model 4, a traditional CLPM, but specified in such a way that the mean structure is first separated from the regression part of the model. This implies we model the means, rather than intercepts, which makes it easier to impose the constraint of identical means over time. So, the fit indices and most parameter estimates are equivalent, except for... (see next question)

### Exercise 7

Which parameter estimates differ? How can they be related?

Click to show answers

The difference in parameter estimates is in the means/intercepts. For the first model we estimate the means at wave 1, and intercepts thereafter. For the second model we obtain mean estimates at each wave (although they are referred to as intercepts in the Mplus output).

Again, the reason for this is that in the first model, the means at wave 2 and 3 are partly predicted by the means from previous waves (through the lagged relationships). Therefore, the constants that are estimated should be interpreted as intercepts. In the second model, the mean part (triangles) and the regression part (lagged parameters) are separated.

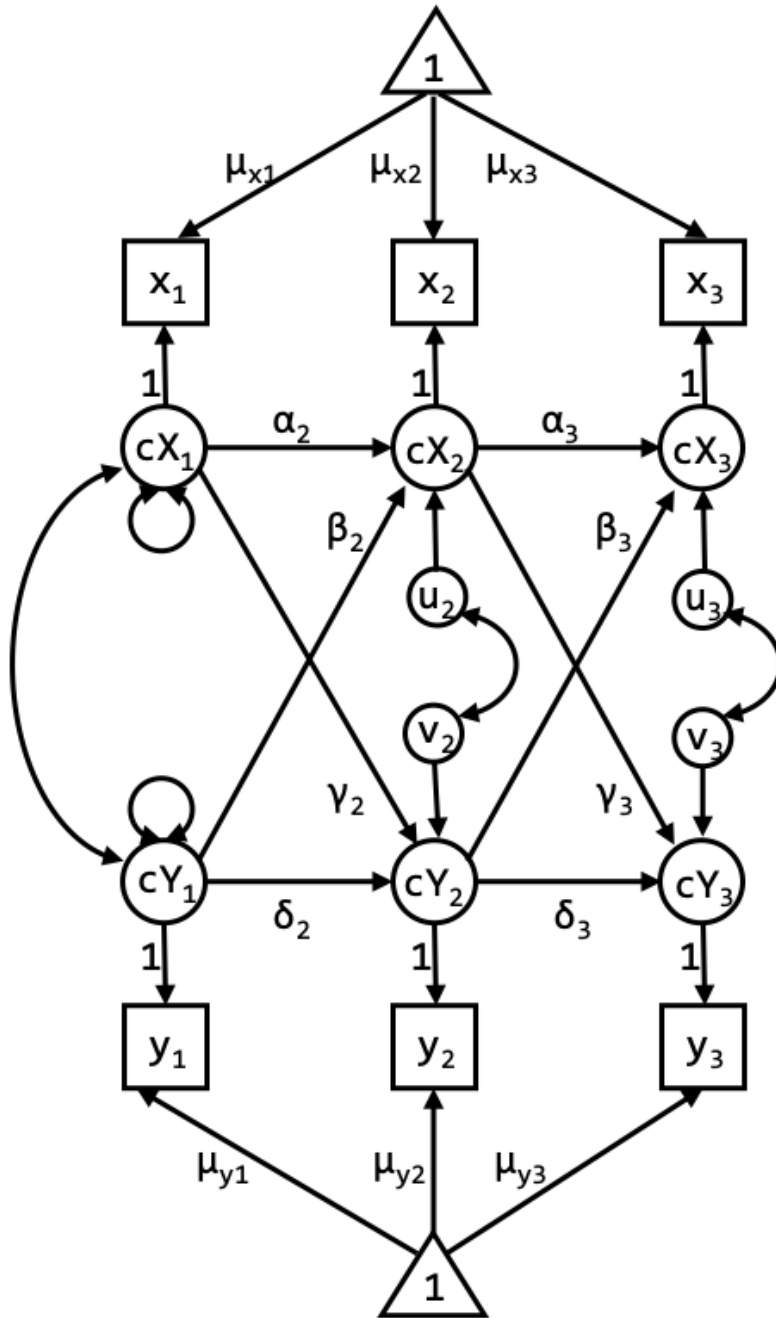


Figure 2: The cross-lagged panel model with centered latent variables.

**Exercise 8**

Constrain the means in the second version of the CLPM. Please do not constrain the mean of `Dep1`, however, as this constraint proved untenable in exercise 3. Discuss the model fit.

[Click to show answers](#)

The input for these models are in `CLPMalt2.inp`. The chi-square has increased in comparison with the previous model: The reason is that this is a restricted model (a special case of the previous model). We can test whether the increase in misfit is significant using the chi-square difference test:  $\Delta\chi^2 = 69.59 - 62.98 = 6.61$  with  $\Delta df = 7 - 4 = 3$ ,  $p = .0854$ . Hence, the constraints are tenable. We can also look at information criteria such as the AIC or BIC. The model with the lowest AIC or BIC is preferred.

**Exercise 9**

Continue with the model from the previous question and constrain all the lagged parameters between wave 1 and 2 to be identical to the lagged parameters between wave 2 and 3. Discuss the model fit, and compare this model to the previous one to see whether the constraint holds.

[Click to show answers](#)

The input for these models are in `CLPMalt3.inp`. The chi-square has again increased in comparison with the previous model: The reason is that this is a restricted model (a special case of the previous model). We can test whether the increase in misfit is significant:  $\Delta\chi^2 81.52 - 69.59 = 11.93$ ,  $\Delta(df) = 11 - 7 = 4$ ,  $p = .0178$ .

If we stick to an alpha of .05, the constraints must be rejected (as the increase in misfit is significant). Note that we are doing several model comparisons here, so one could also argue that we should use a smaller alpha to guard against an inflated overall type I error. Again, an alternative would be to use information criteria for model comparisons.

**The RI-CLPM****Exercise 10**

Specify the RI-CLPM for these data (without any constraints on the lagged parameters over time). Discuss the model fit.

[Click to show answers](#)

The input for these models are in `RICLPM.inp`. This model fits very well; however, it only has 1  $df$ , so it is not doing much for data reduction.

**Exercise 11**

Specify the RI-CLPM with the constraints on the grand means at each occasions. Again, please do not constrain the mean of **Dep1**, however, as this constraint proved untenable in exercise 3.

[Click to show answers](#)

The input for these models are in **RICLPM2.inp**. This model has an acceptable (RMSEA,  $\chi^2$ ) to good (CFI and TLI) fit. Comparing it to the previous model, we get a chi-square difference test of  $\Delta\chi^2 = 9.590 - 0.84 = 8.75$ ,  $\Delta df = 4 - 1 = 3$ ,  $p = .0328$ . This implies that the constraints we impose on the means are untenable (they lead to a significant worsening of model fit). You might therefore decide to ultimately not impose these constraints. However, for didactical reasons, we continue with these constraints in place in the next exercise.

**Exercise 12**

Add the constraints on the lagged parameters to the previous model.

[Click to show answers](#)

The input for these models are in **RICLPM3.inp**. This model fits okay (depending on the measure). Comparing it to the previous model, we get  $\Delta\chi^2 = 19.74 - 9.59 = 8.15$ ,  $\Delta df = 8 - 4 = 4$ ,  $p = .0379$ . Again, this implies that constraining the lagged effects to be time-invariant is untenable (it leads to a significant worsening of model fit).

**Exercise 13**

Compare all the models so far simultaneously, using the AIC and BIC. Which model is selected?

[Click to show answers](#)

Based on the AIC, the RI-CLPM without any constraints provides the closest fit to the data, followed by the two other RI-CLPMs (and the adjusted means model). Based on the BIC, the RI-CLPM with constraints on the means and the lagged parameters is superior to the other models.

Exercise	Input file	AIC	BIC
1	Means.inp	9451	9543
3	Means2.inp	9442	9538
6	CLPM.inp and CLPMalt.inp	9494	9586
7	CLPMalt2.inp	9495	9575
8	CLPMalt3.inp	9499	9563

Exercise	Input file	AIC	BIC
9	RICLPM.inp	<b>9438</b>	9541
10	RICLPM2.inp	9441	9533
11	RICLPM3.inp	9443	<b>9518</b>