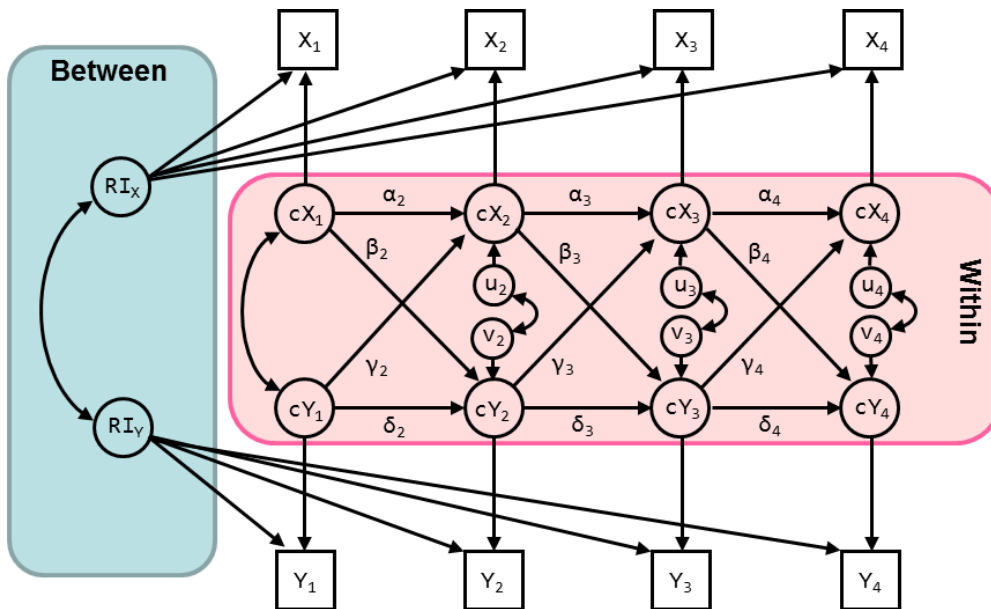# How to run a multiple indicator RI-CLPM with Mplus

By Ellen Hamaker

November 10, 2018

The random intercept cross-lagged panel model (RI-CLPM) as proposed by Hamaker, Kuiper and Grasman (2015, Psychological Methods) is a model that decomposes each observed score into a between-person part and a within-person part. For two variables measured at four waves, it can be represented like this:


I have described how to estimate this model and which steps to go through elsewhere (see https://www.researchgate.net/publication/323935149_How_to_run_the_RI-CLPM_with_Mplus).

I have been asked repeatedly whether this model can be extended to allow for **multiple indicators** for each of the constructs. This is certainly possible, but it requires first to establish whether there is **measurement invariance over time**. Specifically, one should constrain the factor loadings over time to be identical; if this is not the case, the latent variable at one occasion cannot be compared in a meaningful way with the latent variable at another occasion, because they will represent qualitative different constructs (rather than mere quantitative differences on the same underlying construct).

Whether one needs to establish **weak or strong factorial invariance** may be considered a matter of taste: Strong factorial invariance is important if one want to compare the means over time. If one is not interested in the way the (grand) means change over time (or whether they stay invariant), one may settle on weak factorial invariance.

The **multiple indicator RI-CLPM** that we need to work towards (for two constructs, each measured with three indicators at four waves) can be represented like this:
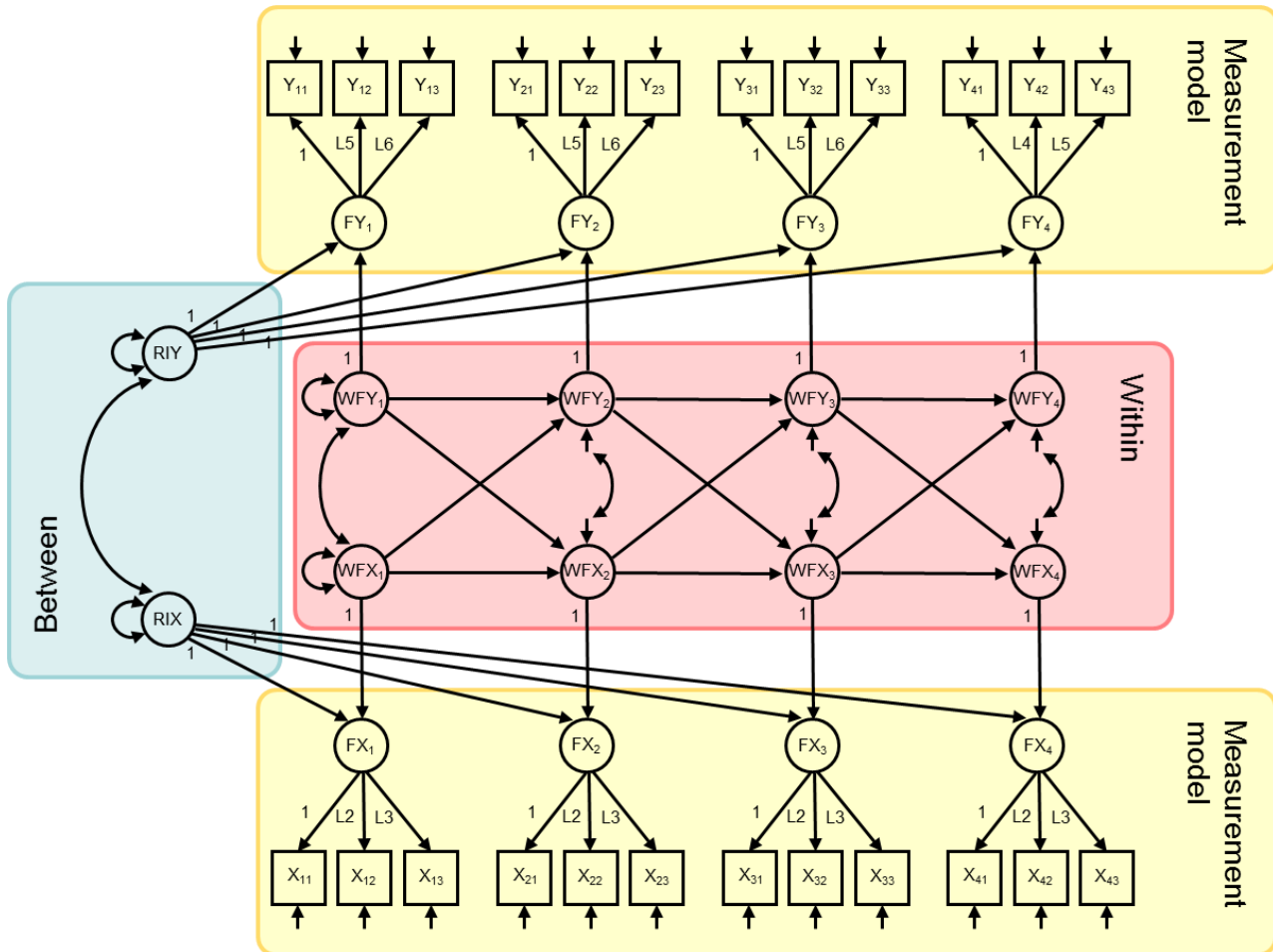
It consists of the following **three parts**:

- **Measurement model part** (in yellow): Here the indicators are linked to the underlying constructs (i.e., the latent variables, or factors)
- **Between part** (blue): These capture the parts of the latent variables that are person-specific and that do not vary over time; hence, we could refer to these as the traits
- **Within part** (red): These represent the temporary fluctuations of the latent variables around the person-specific scores; here we model the dynamics over time using the lagged relationships

But before we can run this model, we should first make sure we have at least weak factorial invariance, and possibly strong factorial invariance over time. The constraints for this are **very similar to the constraints for measurement invariance across groups**. I discuss the sequence of needed models below.

In **Model 1**, we impose the factor structure without any constraints on the factor loadings, means or covariance structure at the latent level; it can be referred to as **configural invariance**. It can be specified as:

```
DATA:   FILE = MIRICLPM.dat;

VARIABLE: NAMES ARE x11 x12 x13 x21 x22 x23 x31 x32 x33 x41 x42 x43
             y11 y12 y13 y21 y22 y23 y31 y32 y33 y41 y42 y43;

MODEL:        ! Factor models for x at 4 waves
              FX1 BY x11-x13;
              FX2 BY x21-x23;
              FX3 BY x31-x33;
              FX4 BY x41-x43;

              ! Factor models for y at 4 waves
              FY1 BY y11-y13;
              FY2 BY y21-y23;
              FY3 BY y31-y33;
              FY4 BY y41-y43;

OUTPUT:       TECH1;
```

In this model, the eight latent variables (FX1 to FY4) are allowed to be freely correlated (default for exogenous variables in Mplus).

In **Model 2**, we constrain the factor loadings to be invariant by giving them the same labels; this model is known as **weak factorial invariance**:

```
MODEL:        ! Factor models for x at 4 waves
              FX1 BY x11-x13 (L1-L3);
              FX2 BY x21-x23 (L1-L3);
              FX3 BY x31-x33 (L1-L3);
              FX4 BY x41-x43 (L1-L3);

              ! Factor models for y at 4 waves
              FY1 BY y11-y13 (L4-L6);
              FY2 BY y21-y23 (L4-L6);
              FY3 BY y31-y33 (L4-L6);
              FY4 BY y41-y43 (L4-L6);
```

These two model are **nested**, such that we can do a **chi-square difference test**. This test will have 12df, because that is the number of constraints we impose to go from model 1 to model 2.

When the chi-square test is **nonsignificant**, this implies the factor loadings are not significantly different from each other over time. In other words, we can assume **weak factorial invariance holds**.

If however the test is **significant**, this implies that the factor loadings cannot be constrained over time. This makes further comparisons between the latent variables **very problematic** or even impossible.

Assuming weak factorial invariance holds, we can proceed with testing whether strong factorial invariance holds. To this end, we specify a model in which we constrain the intercepts over time to be invariant. In addition, we free up the latent means from t=2 onward; if we would not do that, we would actually not only specify strong factorial invariance, but also specify a model in which there cannot be mean changes over time. Such a model may be of interest, but that should be tested separately.

Model 3 can be specified as Model 2 with the following addition to ensure **strong factorial invariance**:

```
! Constrained intercepts over time
[x11 x12 x13] (i1-i3);
[x21 x22 x23] (i1-i3);
[x31 x32 x33] (i1-i3);
[x41 x42 x43] (i1-i3);
[y11 y12 y13] (i4-i6);
[y21 y22 y23] (i4-i6);
[y31 y32 y33] (i4-i6);
[y41 y42 y43] (i4-i6);

! Free latent means from t=2 onward
[FX2* FX3* FX4*];
[FY2* FY3* FY4*];
```

Again, this model is **nested** under the previous model, such that a **chi-square difference test** can be performed to see whether the constraints hold. Again, this test will have 12 df, because that is the number of constraints that were added (18-6=12).

If this chi-square difference test is **nonsignicant**, this means we can assume that **strong factorial invariance** holds over time. In that case we could consider investigating whether the means change over time, by omitting the last two lines in the code above. This is just optional.

If the chi-square test is **significant**, this means strong factorial invariance does not hold. It implies that the **scores cannot be compared over time**, but individual difference in scores may still be meaningfully compared. The latter is probably of most interest when considering cross-lagged panel models, but we should realize this is a less ideal situation.

Once we have established **strong factorial invariance**, we can continue with the actual model of interest. For each construct we specify a random intercept and a within-person centered part, and for the latter we specify the cross-lagged model.

I present the **code in four parts** below, that is, the three parts as presented in the figure, and a fourth part to overrule particular defaults. First, we need to specify the **measurement model**, based on strong factorial invariance as discussed above.

```
! Factor models for x at 4 waves
! constrained factor loadings over time
FX1 BY x11-x13 (L1-L3);
FX2 BY x21-x23 (L1-L3);
FX3 BY x31-x33 (L1-L3);
FX4 BY x41-x43 (L1-L3);

! Factor models for y at 4 waves
! constrained factor loadings over time
FY1 BY y11-y13 (L4-L6);
FY2 BY y21-y23 (L4-L6);
FY3 BY y31-y33 (L4-L6);
FY4 BY y41-y43 (L4-L6);

! Constrained intercepts over time
! (this is necessary for strong factorial
! invariance; without these constraints
! we have weak factorial invariance)
[x11 x12 x13] (i1-i3);  [x21 x22 x23] (i1-i3);
[x31 x32 x33] (i1-i3);  [x41 x42 x43] (i1-i3);
[y11 y12 y13] (i4-i6);  [y21 y22 y23] (i4-i6);
[y31 y32 y33] (i4-i6);  [y41 y42 y43] (i4-i6);

! Free latent means from t=2 onward
! (only do this in combination with the
! constraints on the intercepts; without
! these, this would not be identified)
[FX2* FX3* FX4*];
[FY2* FY3* FY4*];
```

Note that the latter grey part in the code is for **strong factorial invariance**; without it, the model would be based on **weak factorial invariance**.

The second part of the model code contains the **between part** of the model. It looks like this:

```
! Create between factors (random intercepts)
RIX BY FX1@1 FX2@1 FX3@1 FX4@1;
RIY BY FY1@1 FY2@1 FY3@1 FY4@1;

! Set the residual variances of all FX and FY variables to zero
FX1-FY4@0;
```

The third part of the model code contains the **within part**, that is:

```
! Create the within part
WFX1 BY FX1@1;
WFX2 BY FX2@1;
WFX3 BY FX3@1;
WFX4 BY FX4@1;
WFY1 BY FY1@1;
WFY2 BY FY2@1;
WFY3 BY FY3@1;
WFY4 BY FY4@1;

! Specify the lagged effects between the
! within-person centered latent variables
WFX2 ON WFY1 WFX1;
WFX3 ON WFY2 WFX2;
WFX4 ON WFY3 WFX3;
WFY2 ON WFY1 WFX1;
WFY3 ON WFY2 WFX2;
WFY4 ON WFY3 WFX3;

! Estimate the correlations within the same wave
WFX1 WITH WFY1;
WFX2 WITH WFY2;
WFX3 WITH WFY3;
WFX4 WITH WFY4;
```

Finally, we have to **overrule the Mplus default that all exogenous latent variables are correlated**. Specifically, we do not want the between components to be correlated with the within components at the first wave (all of these are exogenous). Hence we add to the code:

```
RIX WITH WFX1@0 WFY1@0;
RIY WITH WFX1@0 WFY1@0;
```

The model presented above can be considered the **basic extension of the RI-CLPM with multiple indicators**. If the model fits well, one can impose further constraints if these are of interest, such as:

- equal lagged effects over time (only when intervals are of equal length!)
- time-invariant variances and covariances in the dynamic part (only from t=2 forward, and only when intervals between measurements are of equal length!)
- time-invariant measurement error variances (for strict factorial invariance)