1

Measurement Invariance for Social Phobia Example

This document reports measurement invariance tests for the social phobia example in Chapter 11. I assume you have read the document on measurement invariance tests on the resources tab for Chapter 3. Here, I perform two tests. First, I test for longitudinal measurement non-invariance of the indicators of social phobia from baseline to posttreatment. Second, I test for measurement non-invariance of the indicators at posttest as a function of the treatment vs. control conditions. The latter test evaluates if the program changed the way people interpret the scales used to measure social phobia. I do not perform the latter tests for the baseline indicators of social phobia because assignment to treatment condition is random and I would not expect group differences as a function of treatment condition.

LONGITUDINAL MEASUREMENT NON-INVARIANCE

A goal of longitudinal tests of measurement non-invariance is to evaluate if factor loadings and measurement intercepts for latent variable indicators differ across time. My first task is to choose a reference indicator for latent social phobia at the two time points under the assumption that the reference indicator has an invariant loading and invariant measurement intercept across time. Knowing that this is the case, I can then empirically evaluate the invariance of the other indicators.

There is no assumption free method for choosing the reference indicator for purposes of measurement non-invariance testing. The forward analysis strategy I described in the Chapter 3 measurement invariance primer makes the assumption that the latent variable variances are equal across time. This was a reasonable assumption for the data I analyzed in Chapter 3, but it is not reasonable for the social phobia example. For instance, in the primary RET analysis I reported in Chapter 11, the estimated latent variance of social phobia at baseline was 0.45 and at posttreatment, it was 1.59, with the latter being about 3.5 times larger than the former.[1] This result rules out the forward analysis strategy for choosing the reference indicator. A second strategy is to use a longitudinal implementation of the backward analysis method by Raykov et al. (2013) described in the primer. Table 1 presents the relevant Mplus syntax.

[1] I obtained this information from the TECH4 output section of the analysis output.

Table 1: Mplus Syntax to Find Reference Indicator for Across Time Analysis

```
1. TITLE: MEASUREMENT INVARIANCE BASE MODEL ;
2. DATA: FILE IS c:\mplus\ret\chap11M.txt ;
3. VARIABLE:
4.   NAMES ARE ID CR1 SPAI1 SPIN1 CR3 SPAI3 SPIN3
5.      NEGAPP2 PSKILLS2 EXTERN2 NEGAPP1 PSKILLS1 EXTERN1
6.      HYPER SEX TREAT ;
7.   USEVARIABLES ARE CR1 SPAI1 SPIN1 CR3 SPAI3 SPIN3   ;
8.   MISSING ARE ALL (-9999) ;
9. ANALYSIS:
10.  ESTIMATOR = MLR;
11. MODEL:
12.  LSP1 BY CR1* SPAI1* SPIN1* (L1a L2a L3a) ;
13.  LSP1@1 (varLSP1) ;
14.  [CR1] (i1a) ;  [SPAI1] (i2a) ; [SPIN1] (i3a) ;
15.  [LSP1@0] (mean1) ;
16.  LSP3 BY CR3* SPAI3* SPIN3* (L1a L2a L3a) ;
17.  LSP3* (varLSP3) ;
18.  [CR3] (i1a) ;  [SPAI3] (i2a) ; [SPIN3] (i3a) ;
19.  [LSP3] (mean2) ;
20.  LSP1 WITH LSP3;
21. OUTPUT: SAMP RESIDUAL CINTERVAL TECH4 STAND(STDYX) ;
```

I number lines for purposes of exposition, but the numbers are not part of standard Mplus syntax. Lines 1 through 11 should be familiar to you so I do not comment on them. Lines 12 and 13 use the fixed factor variance method to identify the measurement model for the three indicators of latent social phobia at baseline. Line 14 tells Mplus to estimate the measurement intercepts for the baseline indicators and line 14 fixes the mean of the baseline latent social phobia variable to zero. The entries contained within the parentheses for these lines are labels for the parameters.

Lines 16 to 19 repeat the above but for the posttreatment latent social phobia variable, with some exceptions. First, rather than fix the variance of latent social phobia at 1.0, the variance is estimated (line 17). This obviates the need to assume equal latent variances across time. Second, the mean of the latent social phobia is estimated rather than fixed at 0 (Line 19). This permits the latent means to vary across time.

An important feature of the code is the use of equality constraints via common labels for parameters. Note that the three factor loadings are constrained to be equal across time as are the three measurement intercepts. The resulting chi square test of model fit is used as a "base model" standard against which significance tests of loading and measurement intercept differences across time are evaluated. I illustrate this process shortly. The chi square statistic for the model was 7.03, df = 12, $p < 0.86$, suggesting a reasonable model fit. This is important because the Raykov et al. (2013) test can fail if the chi square fit is poor.

With three loading differences and three measurement intercept differences across time, I perform a total of six contrasts. Consider the contrast to compare the factor loading for CR1 with the factor loading for CR3. To test the difference between these two loadings to address across-time loading non-invariance for the clinician rating, I make the labels for the two loadings different. I accomplish this by changing Line 16 in Table 1 from

```
LSP3 BY CR3* SPAI3* SPIN3* (L1a L2a L3a) ;
```

to

```
LSP3 BY CR3* SPAI3* SPIN3* (L1b L2a L3a) ;
```

I then re-estimate the model with this one equality constraint relaxed and obtain a model chi square of 6.74 with df = 11. I perform a chi square difference test of the relaxed model relative to the base model but I use a scaled chi square difference test because of my use of robust MLR estimation. A program called *scaled chi sqr difference test* is provided on my webpage to execute the analysis with the correction factors provided by Mplus. The chi square difference was 0.267, df = 1 p < 0.605. The statistically non-significant result means the fit of the model is negligibly affected by introducing the equality constraint. The null hypothesis of equal loadings across time cannot be rejected, which is consistent with (but does not prove) loading invariance, as I discuss in the measurement invariance primer in Chapter 3. I repeat this process for each of the five other contrasts. The results are summarized in Table 2. Raykov et al. (2013) recommend applying a False Discovery Rate (FDR) correction to adjust for contrast multiplicity. I did so using the program called *FDR p values* on my website and these results also appear in Table 2. None of the contrasts were statistically significant, which is consistent with loading and intercept invariance.

Table 2: Scaled chi squared difference test

Contrast	Scaled chi sqr diff	p value	FDR p value
FL: CR3	0.282	0.595	0.752
FL: SPAI	0.570	0.450	0.752
FL: SPIN3	0.018	0.893	0.893
MI: CR3	0.538	0.463	0.752
MI: SPAI	1.659	0.198	0.752
MI: SPIN3	0.236	0.627	0.752

Based on the above, I can reasonably use any of the three indicators as the reference indicator and I can also conclude that there is not empirical support for the presence of measurement non-invariance.

I supplemented the above analyses using the full FISEM model in the main text to evaluate loading non-invariance but taking into account *all* variables in the model rather than just the subset of variables used in the Raykov et al. (2013) procedure. I also used the supplemental analysis to bring to bear the effect size perspectives of Oberski as discussed in the Chapter 3 primer. Table 3 reproduces the Mplus syntax from the full model per Chapter 11. This syntax in its current form ignores matters of measurement non-invariance.

Table 3: Mplus Syntax for Social Phobia Example in Chapter 11

```
TITLE: EXAMPLE CHAPTER 11 ;
DATA: FILE IS c:\mplus\ret\chap11M.txt ;
VARIABLE:
NAMES ARE ID CR1 SPAI1 SPIN1 CR3 SPAI3 SPIN3
NEGAPP2 PSKILLS2 EXTERN2 NEGAPP1 PSKILLS1 EXTERN1
HYPER SEX TREAT ;
USEVARIABLES ARE CR1 SPAI1 SPIN1 CR3 SPAI3 SPIN3
NEGAPP2 PSKILLS2 EXTERN2 NEGAPP1 PSKILLS1 EXTERN1
HYPER SEX TREAT ;
MISSING ARE ALL (-9999) ;
ANALYSIS:
ESTIMATOR = MLR ;  !Robust maximum likelihood
MODEL:
!Specify latent variables
   LSP1 BY CR1 SPAI1 SPIN1 ;
   LSP3 BY CR3 SPAI3 SPIN3 ;
[CR1@0] ; [CR3@0] ; [LSP1] (mean1) ; [LSP3] (int1) ;
!Specify equations
LSP3 ON LSP1 NEGAPP2 PSKILLS2 EXTERN2 TREAT SEX (b10 p4-p7 b11) ;
LSP3 ON HYPER (b12) ;
NEGAPP2 ON TREAT HYPER SEX NEGAPP1 PSKILLS2 (p1 b1-b3 p8) ;
PSKILLS2 ON TREAT HYPER SEX PSKILLS1 (p2 b4-b6) ;
EXTERN2 ON TREAT HYPER SEX EXTERN1 PSKILLS2 (p3 b7-b9 p9) ;
!Specify correlations of latent variable with exogenous variables
LSP1 WITH NEGAPP1 PSKILLS1 EXTERN1 TREAT SEX HYPER ;
MODEL INDIRECT:
LSP3 IND TREAT ;
LSP3 IND PSKILLS2 ;
NEGAPP2 IND TREAT ;
EXTERN2 IND TREAT ;
OUTPUT:
SAMP STANDARDIZED(STDYX) MOD(ALL 4) RESIDUAL CINTERVAL TECH4 ;
```

The model uses CR1 and CR3 as reference indicators for latent social phobia, which the backward analysis suggests is loading (and measurement intercept) invariant across time. The chi square for the model fit was 50.34, df = 57, $p < 0.57$. I make the following changes to the code. First, I add labels to the indicators on Lines 15 and 16:

```
LSP1 BY CR1 SPAI1 SPIN1 (L11 L21 L31) ;
LSP3 BY CR3 SPAI3 SPIN3 (L13 L23 L33) ;
```

Second, I remove the `MOD(All 4)` option in the output line (Line 32). Then I add the following syntax before line 31:

```
MODEL CONSTRAINT:
NEW(diff1 diff2) ;
diff1=L21-L23 ;
diff2=L31-L33 ;
```

This syntax calculates and tests the significance of the loading difference between SPAI1 and SPAI3 across time and between SPIN1 and SPIN3 across time (see the primer on measurement invariance for Chapter 3 for further explication of the syntax). Here is the relevant output from the section MODEL RESULTS:

```
                                                      Two-Tailed
                    Estimate       S.E.  Est./S.E.     P-Value

New/Additional Parameters
    DIFF1             -0.009      0.049     -0.179       0.858
    DIFF2             -0.003      0.057     -0.055       0.956
```

The across time unstandardized loading differences are in the `Estimate` column, with the estimated standard errors of the difference in the `S.E.` column, the critical ratio for the difference in the `EST./S.E.` column, and the p value for the difference in the last column. The results are consistent with the backward analysis of loading invariance but are now grounded in the full substantive model.

I can alter this new syntax further by removing all of the `MODEL CONSTRAINT` commands and introducing across time equality constraints for the non-reference indictors by creating common labels between them in Lines 15 and 16 as follows:

```
LSP1 BY CR1 SPAI1 SPIN1 (L11 L21 L31) ;
LSP3 BY CR3 SPAI3 SPIN3 (L13 L21 L31) ;
```

This imposes a model with across time loading invariance and the question is how such an imposition affects model fit. The chi square fit for this revised model was 50.45, df = 59 and $p < 0.78$, which compares favorably with the chi square for the original model of 50.34,

$df = 57$, $p < 0.57$. The scaled chi square difference test between the two models was decidedly non-significant. However, my primary interest in conducting this analysis is to apply the logic of Oberski (2014) to determine how substantive parameters of interest in the model are affected by imposing versus ignoring loading invariance. For example, in the original model tested in Chapter 11 that ignored loading invariance, the path coefficient linking negative cognitive appraisals and posttreatment latent social phobia (`NEGAPP2` and `LSP3`) was 0.390 ± 0.19 (critical ratio = 4.10, $p < 0.05$). In the model that imposed loading invariance constraints, the same parameter was 0.391 ± 0.19 (critical ratio = 4.09, $p < 0.05$). The difference in the parameter values was trivial and this was true of other coefficients of substantive interest as well. This result provides additional support of modeling the data without imposing loading invariance constraints.

TREATMENT-CONTROL MEASUREMENT INVARIANCE

To test for measurement invariance for the posttreatment social phobia indicators as a function of treatment condition, I used the alignment strategy of Asparouhov and Muthén (2014). Table 4 present the relevant Mplus syntax.

Table 4: Alignment Approach

```
TITLE: ALIGNMENT TEST ;
DATA: FILE IS invariance.dat ;
VARIABLE:
NAMES ARE id d1 d2 d3 d4 d5 d6 adhere adhere2 dfemale income ethnic ;
USEVARIABLES ARE CR3 SPAI3 SPIN3 ;
MISSING ARE ALL (-9999) ;
CLASSES = c(2) ; !number of classes
KNOWNCLASS = c(TREAT = 0 1) ; !variable  values for groups
ANALYSIS:
TYPE=MIXTURE ;
ESTIMATOR=MLR ;
! ALIGNMENT=FREE;
ALIGNMENT=FIXED(0);
MODEL:
%OVERALL%
LSP3 BY CR3* SPAI3* SPIN3* ;
[CR3] ; [SPAI3] ; [SPIN3] ;
%c#1%
LSP3 BY CR3* SPAI3* SPIN3* (L1_1 L1_2 L1_3) ;
[CR3] (i1_1) ; [SPAI3] (i1_2) ; [SPIN3] (i1_3);
%c#2%
LSP3 BY CR3* SPAI3* SPIN3* (L2_1 L2_2 L2_3) ;
[CR3] (i2_1) ; [SPAI3] (i2_2) ; [SPIN3] (i2_3);
OUTPUT: ALIGN CINTERVAL SAMP RESIDUAL TECH4 TECH8 ;
```

The syntax format is fully explained in the primer on measurement invariance on the resources tab for Chapter 3, so I do not comment on it here. The control group is group 0 and the treatment group is group 1. The output of primary interest is the alignment output that appears in the section called `ALIGNMENT OUTPUT`. Here is the relevant (edited) output:

```
Intercepts/Thresholds
Intercept for CR3
Group     Group      Value      Value     Difference  SE       P-value
    1         0      3.069      3.118     -0.049      0.088      0.577
Approximate Measurement Invariance Holds For Groups: 0 1

Intercept for SPAI3
Group     Group      Value      Value     Difference  SE       P-value
    1         0      3.157      3.033      0.124      0.136      0.362
Approximate Measurement Invariance Holds For Groups: 0 1

Intercept for SPIN3
Group     Group      Value      Value     Difference  SE       P-value
    1         0      3.089      3.110     -0.021      0.052      0.686
Approximate Measurement Invariance Holds For Groups: 0 1

Loadings
Loadings for CR3
Group     Group      Value      Value     Difference  SE       P-value
    1         0      0.926      0.993     -0.067      0.062      0.281
Approximate Measurement Invariance Holds For Groups: 0 1

Loadings for SPAI3
Group     Group      Value      Value     Difference  SE       P-value
    1         0      0.986      0.918      0.068      0.067      0.312
Approximate Measurement Invariance Holds For Groups: 0 1

Loadings for SPIN3
Group     Group      Value      Value     Difference  SE       P-value
    1         0      0.993      0.994     -0.001      0.032      0.984
Approximate Measurement Invariance Holds For Groups: 0 1
```

The results support both measurement intercept and loading invariance across the treatment and control groups (see the primer on measurement invariance for Chapter 3 for details of interpretation). I can apply the logic of Oberski (2014) as well but given that none of the results were statistically significant, I leave that as an exercise for you to do based on my discussion of such methods in the Chapter 3 primer.

REFERENCES

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling*, 21, 495-508.

Oberski, D. L. (2014). Evaluating sensitivity of parameters of interest to measurement invariance in latent variable models. *Political Analysis*, 22, 45-60.

Raykov, T., Marcoulides, G. A., & Millsap, R. E. (2013). Factorial invariance in multiple populations: A multiple testing procedure. *Educational and Psychological Measurement*, 73, 713-727.