Producing Partial or Fully Standardized Solutions in Mplus with Constrained Estimation

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Producing Partial or Fully Standardized Solutions in Mplus with Constrained Estimation

1. Introduction
2. Standardizing Endogenous Latent Variables with Constrained Estimation
    - Deriving the Constraints
    - Mplus Setup for Constrained Estimations
3. A Fully Standardized Solution from a Correlation Matrix

Introduction

- In the module on *Selected Models*, we discovered that some models are not invariant under change of scale of their latent variables.
- Such models might be considered meaningful under a fixed metric. Two types of models are of particular interest:
  1. A partially standardized metric where all latent variables are standardized (but not the manifest variables) are standardized to have unit variances;
  2. A fully standardized solution in which all variables, manifest and latent, are standardized to have unit variances.

Introduction

- Unfortunately, in the preceding discussion, we discovered that the default approach to a fully standardized model in Mplus need not generate a correct standardized solution when equality constraints interact with ULI constraints.
- This is because Mplus employs a "compute and transform" approach to standardizaion that involves first obtaining an unstandardized solution, then converting it to a standardized metric.
- However, as we demonstrated, the initial model in this approach is actually a different model than the standardized model, and so not only is the model discrepancy incorrect, but the transformed solution yields parameters that do not satisfy the equality constraints.

Introduction

- Fortunately, there is a solution to this problem available within Mplus.
- One uses the MODEL CONSTRAINT capability in Mplus to constrain all manifest and latent variable variances to unity during estimation, thereby avoiding the need to transform the solution to standardized form.
- This approach requires some careful use of the algebra of variances and covariances.
- We'll begin with the relatively simple partially standardized model, then move on to discuss how to fit models on correlation structures and fully standardized models.

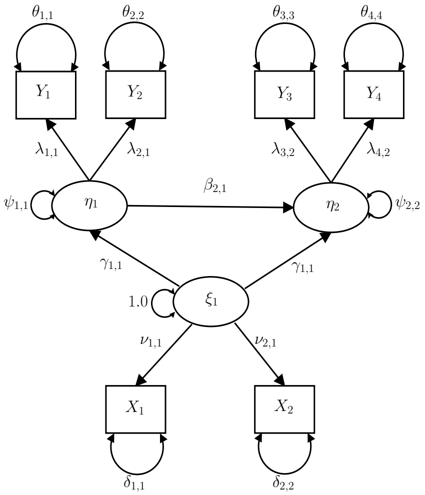
Standardizing Endogenous Latent Variables with Constrained Estimation

- In a structural equation model, we can control the variance of exogenous latent variables directly by parameterizing the model so that the latent variable has a fixed variance of 1.
- However, endogenous latent variables have a variance that is determined by the paths leading into them, and so their variances cannot be directly parameterized.
- Mplus allows one to name parameters in a model, and establish model constraints on those parameters.
- We can use those constraints to control the variance of an endogenous latent variable during the estimation process.
- Consider the general structural equation model presented on the next slide.
- In this model, we wish to constrain two structural coefficients to be equal, so that ξ_1 has the same regression coefficient for both η_1 and η_2. Hence both coefficients are labeled γ_{11} in the diagram.
- We have used the parameterization for the exogenous factor ξ_1 that fixes its variance at 1.0. Hence no ULI constraint is employed.

Standardizing Endogenous Latent Variables with Constrained Estimation

- Note also that this reduced form diagram employs does not show residual and disturbance terms of endogenous variables directly. Rather, it shows them as variances attached to the variables.
- So, for example, $\delta_{1,1}$ is not the variance of X_1, it is its error variance.
- The variances of η_1 and η_2 need to be constrained at 1.0.

Standardizing Endogenous Latent Variables with Constrained Estimation



Standardizing Endogenous Latent Variables with Constrained Estimation

Deriving the Constraints

- We have two latent variables that are endogeneous and need their variances constrained at 1.0.
- To do this, we first need to derive an expression for the variance of each variable.
- This involves simply examining the diagram, converting it to a pair of linear equations, and employing the algebra of variances and covariances.

Standardizing Endogenous Latent Variables with Constrained Estimation

Deriving the Constraints

- Consider η_1 first. From the diagram, we can write the formula

$$\eta_1 = \gamma_{1,1}\xi_1 + \zeta_1 \tag{1}$$

  where ζ_1 is the "invisible" residual with a variance of $\psi_{1,1}$.
- Note that ζ_1 is not represented explicitly in the diagram.
- From the algebra of variances and covariances, keeping in mind that ζ_1 and η_1 are uncorrelated, we may immediately write

$$\begin{aligned}\mathrm{Var}(\eta_1) &= \gamma_{1,1}^2\,\mathrm{Var}(\xi_1) + \mathrm{Var}(\zeta_1) + 2\,\mathrm{Cov}(\xi_1, \zeta_1) \\ &= \gamma_{1,1}^2(1) + \psi_{1,1} + (2)(0) \\ &= \gamma_{1,1}^2 + \psi_{1,1} \qquad (2)\end{aligned}$$

Standardizing Endogenous Latent Variables with Constrained Estimation

Deriving the Constraints

- Mplus requires that the constraint be expressed in the form $0 = f(\boldsymbol{\theta})$, that is, that some function of model parameters is zero.
- Substituting the required value of 1 for $\text{Var}(\eta_1)$, we can re-express Equation 2 easily as

$$0 = \gamma_{1,1}^2 + \psi_{1,1} - 1 \tag{3}$$

Standardizing Endogenous Latent Variables with Constrained Estimation

Deriving the Constraints

- The second constraint takes a bit more work.
- Examining η_2 in the diagram, we see that

$$\eta_2 = (\beta_{2,1}\eta_1 + \gamma_{1,1}\xi_1) + \zeta_2 \tag{4}$$

- Note the strategically placed parentheses in the above equation. They are to remind us that ζ_2 is uncorrelated with the remainder of the terms in the equation, so we may write

$$\begin{aligned} \text{Var}(\eta_2) &= \text{Var}(\beta_{2,1}\eta_1 + \gamma_{1,1}\xi_1) + \text{Var}(\zeta_2) \\ &= (\beta_{2,1}^2 \text{Var}(\eta_1) + \gamma_{1,1}^2 \text{Var}(\xi_1) + 2\beta_{2,1}\gamma_{1,1} \text{Cov}(\eta_1, \xi_1)) + \psi_{2,2} \end{aligned} \tag{5}$$

- Since ξ_1 has its variance fixed at 1.0, and η_1 has a variance constrained to be 1.0 by the first constraint, and we want to constrain the variance of η_2 at 1, we can simplify this a bit more as

$$0 = \beta_{2,1}^2 + \gamma_{1,1}^2 + 2\beta_{2,1}\gamma_{1,1} \text{Cov}(\eta_1, \xi_1) + \psi_{2,2} - 1 \tag{6}$$

Standardizing Endogenous Latent Variables with Constrained Estimation

Deriving the Constraints

- Since $\eta_1 = \gamma_{1,1}\xi_1 + \zeta_1$, it is easily established that $\text{Cov}(\eta_1, \xi_1) = \gamma_{1,1}$, whence our constraint becomes

$$0 = \beta_{2,1}^2 + (1 + 2\beta_{2,1})\gamma_{1,1}^2 + \psi_{2,2} - 1 \qquad (7)$$

Standardizing Endogenous Latent Variables with Constrained Estimation

Mplus Setup for Constrained Estimations

- Setting up our model for estimation in Mplus is straightforward. We need to
  1. Name all parameters that are included in the constraint equations, and
  2. Provide Mplus with the constraint equations in Mplus syntax.
- In the Mplus input on the next slide, we set up our model to fit the raw data analyzed previously near the end of the *Selected Models* lecture slides.

Standardizing Endogenous Latent Variables with Constrained Estimation

Mplus Setup for Constrained Estimations

```
TITLE: TEST DATA FROM STEIGER(2002) -- CONSTRAINED ESTIMATION
DATA: FILE IS MonteCarloFullCovNonPerfect.txt;
TYPE IS FULLCOV;
NOBSERVATIONS=931;
VARIABLE: NAMES ARE Y1 Y2 Y3 Y4 X1 X2;
MODEL: XI1 BY X1*(NU11);
XI1 BY X2*(NU21);
XI1@1;
ETA1 BY Y1*(LAMBDA11);
ETA1 BY Y2*(LAMBDA21);
ETA2 BY Y3*(LAMBDA32);
ETA2 BY Y4*(LAMBDA42);
ETA1 ON XI1*(GAMMA11);
ETA2 ON XI1*(GAMMA11);
ETA2 ON ETA1*(BETA21);
ETA1*(PSI11);
ETA2*(PSI22);

MODEL CONSTRAINT:
0 = GAMMA11**2 + PSI11 - 1;
0 = BETA21**2 + (1 + 2*BETA21)*GAMMA11**2 + PSI22 - 1;
```

Standardizing Endogenous Latent Variables with Constrained Estimation

Mplus Setup for Constrained Estimations

```
Chi-Square Test of Model Fit

          Value                             10.331
          Degrees of Freedom                     7
          P-Value                           0.1706

MODEL RESULTS

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

 XI1      BY
    X1                 0.365      0.032     11.438      0.000
    X2                 0.259      0.028      9.260      0.000

 ETA1     BY
    Y1                 0.257      0.031      8.333      0.000
    Y2                 0.342      0.036      9.468      0.000

 ETA2     BY
    Y3                 0.306      0.028     10.863      0.000
    Y4                 0.390      0.032     12.198      0.000

 ETA1     ON
    XI1                0.703      0.094      7.463      0.000

 ETA2     ON
    XI1                0.703      0.094      7.463      0.000
    ETA1               0.082      0.125      0.660      0.509

 Variances
    XI1                1.000      0.000    999.000    999.000

 Residual Variances
    Y1                 0.326      0.019     16.793      0.000
    Y2                 0.287      0.025     11.601      0.000
    Y3                 0.295      0.019     15.864      0.000
    Y4                 0.248      0.023     10.634      0.000
    X1                 0.274      0.023     12.141      0.000
    X2                 0.335      0.019     18.036      0.000
    ETA1               0.506      0.132      3.818      0.000
    ETA2               0.418      0.078      5.332      0.000
```

Standardizing Endogenous Latent Variables with Constrained Estimation

Mplus Setup for Constrained Estimations

- The results agree almost precisely with those of SEPATH's constrained estimation approach.
- It is not clear to what extent the differences are attributable to differences in the way Mplus calculates the covariance matrix, a minor difference in convergence, or a difference in the way standard errors are calculated when estimates are obtained using model constraints.
- However, in this case it does not appear that the differences have any substantive impact.

A Fully Standardized Solution from a Correlation Matrix

- Mels(1989) described how to analyze a correlation matrix correctly using constrained estimation.
- His approach eliminates the problems described by Cudeck(1989), and yields model estimates that match those of Lawley and Maxwell for their classic confirmatory factor analysis problem.
- Moreover, this approach, when employed with standardized latent variables, allows computation of a "fully standardized" solution whether the covariance matrix *or the correlation matrix* is input.
- This approach eliminates the occasional anomalies that occur with the more "nuanced" examples described by Steiger (2002).

A Fully Standardized Solution from a Correlation Matrix

- The general approach of Mels is as follows:
- Start with the path diagram as it would be if covariances were analyzed.
- Replace each manifest variables with a "dummy" latent variable, with a single arrow (having a free "scaling parameter") pointing to the original manifest variable.
- If the original manifest variable is exogenous, its dummy latent variable will be exogenous.
- If the original manifest variable is endogenous, its dummy latent variable will be endogenous.
- Fix the variance of each exogenous dummy latent variable at 1.
- Constrain the variance of each endogenous dummy latent variable at 1, using model constraints.

A Fully Standardized Solution from a Correlation Matrix

- Let's examine how to revise our constrained estimation of the General Model to yield a fully standardized solution.
- We begin by "dummifying" each manifest variable name by replacing it with a its previous name prefixed with a "D".
- We also add a path from each dummy to its corresponding manifest variable.
- We must also remove all residuals from the manifest variables by setting their variances equal to zero. This is necessary because the manifest variable residuals are now attached to their dummy latent counterparts — if we don't forcibly remove them, Mplus will leave them on, resulting in an identification problem.

A Fully Standardized Solution from a Correlation Matrix

```
TITLE: TEST DATA FROM STEIGER(2002) -- CONSTRAINED ESTIMATION
DATA: FILE IS FullCorNonPerfect.txt;
TYPE IS FULLCOV;
NOBSERVATIONS=931;
VARIABLE: NAMES ARE Y1 Y2 Y3 Y4 X1 X2;
MODEL: DX1 BY X1*;
DX2 BY X2*;
DY1 BY Y1*;
DY2 BY Y2*;
DY3 BY Y3*;
DY4 BY Y4*;
X1-X2@0;
Y1-Y4@0;
XI1 BY DX1*(NU11);
XI1 BY DX2*(NU21);
XI1@1;
ETA1 BY DY1*(LAMBDA11);
ETA1 BY DY2*(LAMBDA21);
ETA2 BY DY3*(LAMBDA32);
ETA2 BY DY4*(LAMBDA42);
ETA1 ON XI1*(GAMMA11);
ETA2 ON XI1*(GAMMA11);
ETA2 ON ETA1*(BETA21);
ETA1*(PSI11);
ETA2*(PSI22);

MODEL CONSTRAINT:
0 = GAMMA11**2 + PSI11 - 1;
0 = BETA21**2 + (1 + 2*BETA21)*GAMMA11**2 + PSI22 - 1;
```

A Fully Standardized Solution from a Correlation Matrix

- Finally, we need to add constraints on the variances of the 6 endogenous dummy latents, so that each has a variance of 1.
- The exogenous dummy latents can have their variances fixed directly at 1.0.
- However, we need to use constrained estimation to fix the variances of the endogenous dummy latents.
- Fortunately, the fact that we have already constrained the variances of the endogenous and exogenous factors to be 1 makes the calculation straightforward.
- For example, $DX_1 = \nu_{11}\xi_1 + \delta_1$, whence $\mathrm{Var}(DX_1) = 1 = \nu_{11}^2 \,\mathrm{Var}\, \xi_1 + \mathrm{Var}(\delta_1) = \nu_{11}^2 + \delta_{11}$.
- Consequently, the model constraint is written

  ```
  0 = NU11^2 + DELTA11 - 1
  ```
- We note also that, while we didn't need to name residual variance parameters in our previous model commands, now we need to name them explicitly as DELTA11,DELTA22,THETA11,THETA22,THETA33,THETA44.
- Adding statements for the residuals, and similar restrictions for the remaining 5 dummy latents, we end up with the model instructions shown on the next slide.

A Fully Standardized Solution from a Correlation Matrix

```
TITLE: TEST DATA FROM STEIGER(2002) -- CONSTRAINED ESTIMATION
DATA: FILE IS FullCorNonPerfect.txt;
TYPE IS FULLCOV;
NOBSERVATIONS=931;
VARIABLE: NAMES ARE Y1 Y2 Y3 Y4 X1 X2;
MODEL:
DX1 BY X1*;
DX2 BY X2*;
DY1 BY Y1*;
DY2 BY Y2*;
DY3 BY Y3*;
DY4 BY Y4*;
X1-X2@0;
Y1-Y4@0;
XI1 BY DX1*(NU11);
XI1 BY DX2*(NU21);
XI1@1;
ETA1 BY DY1*(LAMBDA11);
ETA1 BY DY2*(LAMBDA21);
ETA2 BY DY3*(LAMBDA32);
ETA2 BY DY4*(LAMBDA42);
ETA1 ON XI1*(GAMMA11);
ETA2 ON XI1*(GAMMA11);
ETA2 ON ETA1*(BETA21);
ETA1*(PSI11);
ETA2*(PSI22);
DX1*(DELTA11);
DX2*(DELTA22);
DY1*(THETA11);
DY2*(THETA22);
DY3*(THETA33);
DY4*(THETA44);

MODEL CONSTRAINT:
0 = GAMMA11**2 + PSI11 - 1;
0 = BETA21**2 + (1 + 2*BETA21)*GAMMA11**2 + PSI22 - 1;
0 = NU11**2 + DELTA11 - 1;
0 = NU21**2 + DELTA22 - 1;
0 = LAMBDA11**2 + THETA11 - 1;
0 = LAMBDA21**2 + THETA22 - 1;
0 = LAMBDA32**2 + THETA33 - 1;
0 = LAMBDA42**2 + THETA44 - 1;
```

A Fully Standardized Solution from a Correlation Matrix

- The output is shown on the next slides. The loadings of the dummy latents on the manifest variables are nuisance parameters, of no interest.
- The key output follows these loadings.
- Note also that the residual variances for the manifest variables are fixed at zero, and are of no interest.

A Fully Standardized Solution from a Correlation Matrix

```
Chi-Square Test of Model Fit

        Value                              10.331
        Degrees of Freedom                      7
        P-Value                            0.1706

MODEL RESULTS

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

 DX1      BY
    X1                 0.999      0.023     43.150      0.000

 DX2      BY
    X2                 0.999      0.023     43.151      0.000

 DY1      BY
    Y1                 0.999      0.023     43.151      0.000

 DY2      BY
    Y2                 0.999      0.023     43.152      0.000

 DY3      BY
    Y3                 1.000      0.023     43.151      0.000

 DY4      BY
    Y4                 1.000      0.023     43.152      0.000

 XI1      BY
    DX1                0.572      0.046     12.509      0.000
    DX2                0.408      0.041      9.885      0.000

 ETA1     BY
    DY1                0.410      0.047      8.779      0.000
    DY2                0.538      0.053     10.078      0.000

 ETA2     BY
    DY3                0.490      0.041     11.848      0.000
    DY4                0.617      0.046     13.463      0.000

 ETA1     ON
    XI1                0.703      0.094      7.464      0.000

 ETA2     ON
    XI1                0.703      0.094      7.464      0.000
    ETA1               0.082      0.125      0.659      0.510

 Variances
    XI1                1.000      0.000    999.000    999.000

 Residual Variances
    Y1                 0.000      0.000    999.000    999.000
    Y2                 0.000      0.000    999.000    999.000
    Y3                 0.000      0.000    999.000    999.000
    Y4                 0.000      0.000    999.000    999.000
    X1                 0.000      0.000    999.000    999.000
    X2                 0.000      0.000    999.000    999.000
    DX1                0.673      0.052     12.878      0.000
    DX2                0.833      0.034     24.684      0.000
    DY1                0.832      0.038     21.719      0.000
    DY2                0.710      0.057     12.361      0.000
    DY3                0.759      0.041     18.701      0.000
    DY4                0.619      0.057     10.951      0.000
    ETA1               0.506      0.132      3.818      0.000
    ETA2               0.418      0.078      5.335      0.000
```

A Fully Standardized Solution from a Correlation Matrix

- The next slide shows the model coefficients with the nuisance parameters and zero variances removed for easier readability.

A Fully Standardized Solution from a Correlation Matrix

```
Chi-Square Test of Model Fit

          Value                              10.331
          Degrees of Freedom                      7
          P-Value                            0.1706

MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
 XI1      BY
    DX1                0.572      0.046     12.509      0.000
    DX2                0.408      0.041      9.885      0.000

 ETA1     BY
    DY1                0.410      0.047      8.779      0.000
    DY2                0.538      0.053     10.078      0.000

 ETA2     BY
    DY3                0.490      0.041     11.848      0.000
    DY4                0.617      0.046     13.463      0.000

 ETA1     ON
    XI1                0.703      0.094      7.464      0.000

 ETA2     ON
    XI1                0.703      0.094      7.464      0.000
    ETA1               0.082      0.125      0.659      0.510

 Variances
    XI1                1.000      0.000    999.000    999.000

 Residual Variances
    DX1                0.673      0.052     12.878      0.000
    DX2                0.833      0.034     24.684      0.000
    DY1                0.832      0.038     21.719      0.000
    DY2                0.710      0.057     12.361      0.000
    DY3                0.759      0.041     18.701      0.000
    DY4                0.619      0.057     10.951      0.000
    ETA1               0.506      0.132      3.818      0.000
    ETA2               0.418      0.078      5.335      0.000
```

A Fully Standardized Solution from a Correlation Matrix

- On the next slide, we show the corresponding estimates from SEPATH, which are virtually identical. Standard errors show some minor discrepancies in the third decimal place.
- It is not clear at this point whether these discrepancies result from minor differences in convergence, Mplus's "correction" of the covariance matrix, or some difference in the way standard errors are computed with constrained estimation.
- In a subsequent module, we shall investigate the use of Monte Carlo methods to investigate the performance of the estimation procedure for standard errors.

A Fully Standardized Solution from a Correlation Matrix

	Model Estimates (Test Data NonPerfect Fit)			
	Parameter Estimate	Standard Error	T Statistic	Prob. Level
(XI1)-1->[X1]	0.572	0.045	12.614	0.000
(XI1)-2->[X2]	0.408	0.041	9.890	0.000
(XI1)-{1}-(XI1)				
(DELTA1)-->[X1]				
(DELTA2)-->[X2]				
(DELTA1)-3-(DELTA1)	0.673	0.052	12.986	0.000
(DELTA2)-4-(DELTA2)	0.833	0.034	24.695	0.000
(ETA1)-98->[Y1]	0.410	0.047	8.781	0.000
(ETA1)-5->[Y2]	0.538	0.053	10.142	0.000
(ETA2)-99->[Y3]	0.490	0.041	11.848	0.000
(ETA2)-6->[Y4]	0.617	0.046	13.442	0.000
(EPSILON1)-->[Y1]				
(EPSILON2)-->[Y2]				
(EPSILON3)-->[Y3]				
(EPSILON4)-->[Y4]				
(EPSILON1)-7-(EPSILON1)	0.832	0.038	21.721	0.000
(EPSILON2)-8-(EPSILON2)	0.710	0.057	12.435	0.000
(EPSILON3)-9-(EPSILON3)	0.759	0.041	18.704	0.000
(EPSILON4)-10-(EPSILON4)	0.619	0.057	10.931	0.000
(ZETA1)-->(ETA1)				
(ZETA2)-->(ETA2)				
(ZETA1)-11-(ZETA1)	0.506	0.126	4.020	0.000
(ZETA2)-12-(ZETA2)	0.418	0.079	5.290	0.000
(XI1)-13->(ETA1)	0.703	0.090	7.845	0.000
(XI1)-13->(ETA2)	0.703	0.090	7.845	0.000
(ETA1)-15->(ETA2)	0.083	0.120	0.688	0.491