Multigroup SEM Example[1]

All Parameters Free Across Groups

```
title:  Multigroup structural model example--All parameters free ;

  data:  file=C:\Jason\mplus\semclass\stack1.dat;  format=11f1.0;
      listwise=on;

  variable:  names = widow panas1 panas2 panas3 panas4 panas5
             panas6 panas7 panas8 panas9 panas10 ;
             grouping is widow (0=notwidow,1=widow);

             missing = blank;

  analysis:  type=general; iterations = 200;
    model=nomeanstructure; information=expected;

  model:
          posaff by panas1-panas5*;
          posaff@1;
          negaff by panas6-panas10* ;
          negaff@1;

  !  Note:  by default in Mplus, measurement errors and factor correlations are not
  !   constrained to be equal across groups;

  Model notwidow:
          posaff by panas1-panas5*;
          posaff@1;
          negaff by panas6-panas10* ;
          negaff@1;

  Model widow:
          posaff by panas1-panas5*;
          posaff@1;
          negaff by panas6-panas10* ;
          negaff@1;

    ! I've freed all factor loadings but set variances to 1, alternatively, some authors
    !    have suggested using the highest loading item as a reference loading and freeing
    !    factor variances. Choice is arbitrary if all are tested together, but the reference
    !    variable approach must constrain the factor variances to be equal across groups;

  output:  stdyx ;
```

```
Multigroup structural model example--All parameters free ;

SUMMARY OF ANALYSIS

Number of groups                                                 2
Number of observations
   Group NOTWIDOW                                              159
   Group WIDOW                                                  40

Number of dependent variables                                  10
Number of independent variables                                 0
Number of continuous latent variables                           2



  Grouping variable     WIDOW

Estimator                                                      ML
Information matrix                                       EXPECTED
Maximum number of iterations                                  200
Convergence criterion                                  0.500D-04
Maximum number of steepest descent iterations                 20
```

[1] Note: I only illustrate some of the tests that may be needed (e.g., all parameters equal vs. all parameters free, factor variance comparisons, or other specific paths). For brevity sake, I do not illustrate comparison of variances, covariances, or mean and intercept comparisons. See the handout "Multigroup SEM" for an overview. By default in Mplus Version 6 and later, analyses with mean structures set the intercepts to zero in the first group and allow them to be freely estimated in the second group. This should be done when loadings are also constrained (strong factoral invariance).

```
Input data file(s)
  C:\Jason\mplus\semclass\stack1.dat

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

          Value                            112.947
          Degrees of Freedom                    68
          P-Value                           0.0005

Chi-Square Contributions From Each Group

          NOTWIDOW                          72.686
          WIDOW                             40.260

Chi-Square Test of Model Fit for the Baseline Model

          Value                           1643.772
          Degrees of Freedom                    90
          P-Value                           0.0000

CFI/TLI

          CFI                                0.971
          TLI                                0.962

Loglikelihood

          H0 Value                       -1958.440
          H1 Value                       -1901.967

Information Criteria

          Number of Free Parameters             42
          Akaike (AIC)                    4000.880
          Bayesian (BIC)                  4139.198
          Sample-Size Adjusted BIC        4006.140
            (n* = (n + 2) / 24)

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.082
          90 Percent C.I.                    0.054  0.108

SRMR (Standardized Root Mean Square Residual)

          Value                              0.046

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

Group NOTWIDOW

 POSAFF   BY
    PANAS1             0.745      0.054     13.734      0.000
    PANAS2             0.865      0.071     12.145      0.000
    PANAS3             0.791      0.055     14.401      0.000
    PANAS4             0.855      0.057     14.881      0.000
    PANAS5             0.830      0.062     13.367      0.000

 NEGAFF   BY
    PANAS6             0.798      0.054     14.910      0.000
    PANAS7             0.597      0.065      9.234      0.000
    PANAS8             0.748      0.058     12.804      0.000
    PANAS9             0.380      0.054      7.049      0.000
    PANAS10            0.813      0.083      9.777      0.000

 NEGAFF   WITH
    POSAFF            -0.674      0.050    -13.557      0.000
 Variances
```

```
    POSAFF             1.000      0.000    999.000    999.000
    NEGAFF             1.000      0.000    999.000    999.000

Group WIDOW

 POSAFF   BY
    PANAS1             1.092      0.127      8.599      0.000
    PANAS2             1.252      0.160      7.841      0.000
    PANAS3             1.037      0.132      7.825      0.000
    PANAS4             1.054      0.135      7.788      0.000
    PANAS5             0.914      0.136      6.719      0.000

 NEGAFF   BY
    PANAS6             1.099      0.136      8.075      0.000
    PANAS7             0.952      0.142      6.710      0.000
    PANAS8             0.843      0.139      6.047      0.000
    PANAS9             0.405      0.127      3.188      0.001
    PANAS10            0.676      0.188      3.601      0.000

 NEGAFF   WITH
    POSAFF            -0.472      0.129     -3.661      0.000

 Variances
    POSAFF             1.000      0.000    999.000    999.000
    NEGAFF             1.000      0.000    999.000    999.000
```

Loadings Only Constrained Equal Across Groups

```
  title:  Multigroup structural model example--Loadings Only Equal ;

  data:  file=C:\Jason\mplus\semclass\stack1.dat;  format=11f1.0;
      listwise=on;

  variable:  names = widow panas1 panas2 panas3 panas4 panas5
             panas6 panas7 panas8 panas9 panas10 ;
             grouping is widow (0=notwidow,1=widow);

             missing = blank;

  analysis:  type=general; iterations = 200;
    model=nomeanstructure; information=expected;

  model:  posaff by panas1* panas2-panas5;
          negaff by panas6* panas7-panas10*;

  !  Note:  by default in Mplus, measurement errors and factor correlations are not
  !   constrained to be equal across groups;

  Model notwidow:
    posaff by panas1* (1)
              panas2* (2)
              panas3 (3)
              panas4 (4)
              panas5 (5);
    posaff@1;

    negaff by panas6* (6)
              panas7* (7)
              panas8* (8)
              panas9* (9)
              panas10* (10);
    negaff@1;

  Model widow:
    posaff by panas1* (1)
              panas2* (2)
              panas3 (3)
              panas4 (4)
              panas5 (5);
    posaff@1;

    negaff by panas6* (6)
              panas7* (7)
              panas8* (8)
              panas9* (9)
```

```
              panas10* (10);
    negaff@1;

  output:  stdyx ;
```

```
INPUT READING TERMINATED NORMALLY

Multigroup structural model example--Loadings Only Equal ;

SUMMARY OF ANALYSIS

Number of groups                                                 2
Number of observations
   Group NOTWIDOW                                              159
   Group WIDOW                                                  40

Number of dependent variables                                   10
Number of independent variables                                  0
Number of continuous latent variables                            2


Variables with special functions

  Grouping variable     WIDOW

Estimator                                                       ML
Information matrix                                        EXPECTED
Maximum number of iterations                                   200
Convergence criterion                                  0.500D-04
Maximum number of steepest descent iterations                  20



THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT
Chi-Square Test of Model Fit
          Value                             136.569
          Degrees of Freedom                    78
          P-Value                           0.0000

Chi-Square Contributions From Each Group
          NOTWIDOW                          80.162
          WIDOW                             56.407

Chi-Square Test of Model Fit for the Baseline Model
          Value                           1643.772
          Degrees of Freedom                    90
          P-Value                           0.0000

CFI/TLI
          CFI                                0.962
          TLI                                0.957

Loglikelihood
          H0 Value                       -1970.251
          H1 Value                       -1901.967

Information Criteria
          Number of Free Parameters             32
          Akaike (AIC)                     4004.502
          Bayesian (BIC)                   4109.888
          Sample-Size Adjusted BIC         4008.510
            (n* = (n + 2) / 24)

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.087
          90 Percent C.I.                    0.062  0.111

SRMR (Standardized Root Mean Square Residual)
          Value                              0.143

MODEL RESULTS
```

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Group NOTWIDOW				
POSAFF BY				
PANAS1	0.858	0.050	17.035	0.000
PANAS2	0.979	0.065	15.071	0.000
PANAS3	0.854	0.051	16.632	0.000
PANAS4	0.906	0.054	16.915	0.000
PANAS5	0.856	0.057	15.016	0.000
NEGAFF BY				
PANAS6	0.867	0.051	17.098	0.000
PANAS7	0.694	0.059	11.671	0.000
PANAS8	0.786	0.054	14.484	0.000
PANAS9	0.396	0.050	7.973	0.000
PANAS10	0.816	0.076	10.688	0.000
NEGAFF WITH				
POSAFF	-0.711	0.043	-16.386	0.000
Variances				
POSAFF	1.000	0.000	999.000	999.000
NEGAFF	1.000	0.000	999.000	999.000
Group WIDOW				
POSAFF BY				
PANAS1	0.858	0.050	17.035	0.000
PANAS2	0.979	0.065	15.071	0.000
PANAS3	0.854	0.051	16.632	0.000
PANAS4	0.906	0.054	16.915	0.000
PANAS5	0.856	0.057	15.016	0.000
NEGAFF BY				
PANAS6	0.867	0.051	17.098	0.000
PANAS7	0.694	0.059	11.671	0.000
PANAS8	0.786	0.054	14.484	0.000
PANAS9	0.396	0.050	7.973	0.000
PANAS10	0.816	0.076	10.688	0.000
NEGAFF WITH				
POSAFF	-0.405	0.139	-2.922	0.003
Variances				
POSAFF	1.000	0.000	999.000	999.000
NEGAFF	1.000	0.000	999.000	999.000

Chi-square comparisons

			Comparison to all-free model	
	χ^2	*df*	$\Delta\chi^2$	Δdf
All parameters free	112.947	68		
Loadings only equal	136.569	78	23.6222[a]	10

[a] p < .10, * p < .05, ** p < .01, *** p < .001

Sample Write-up

A multigroup structural equation modeling approach was used to compare men and women on the factor loadings of the positive and negative affect scale. To test for weak factorial invariance (Meredith, 1993) across groups, the chi-square from a model with all parameters allowed to be unequal across groups was compared to the chi-square from a model with only the loadings constrained to be equal across groups. No means or intercepts were estimated in these models. The model with all parameters freely estimated in the two groups, fit the data well (CFI = .971, SRMR = .042), according to fit criteria suggested by Hu and Bentler (1999), although the overall chi-square was significant, $\chi^2(68) = 112.947$ $p < .001$. The weak invariance model with loadings constrained to be equal across groups had fit that was significantly poorer, $\chi^2(78) = 136.569$, $p < .001$, $\Delta\chi^2(10) = 23.622$, $p < .01$. This difference was small in magnitude, however, $w = .064$, $\Delta Mc = .011$. The Comparative Fit Index for this model indicated good fit (CFI = .962), but the Standardized Root Mean Square Residual (SRMR = .143) suggested the fit could be improved. Further analyses are required to determine which loadings may differ across groups. The findings suggest that the measurement of the two-factor positive and negative affect scale differs across groups, and, thus, caution may be warranted in comparing these groups.[2]

[2] In an actual analysis, the researcher should not stop here. Identifying the source of the difference is important for possible modifications.

lavaan
Loadings Only Constrained Equal Across Groups *(some output deleted)*

```
> #listwise deletion
> stack1 <- stack1[complete.cases(stack1), ]
> model = '
+         posaff =~ NA*panas1 + panas2 + panas3 + panas4 + panas5
+       negaff =~ NA*panas6 + panas7 + panas8 + panas9 + panas10
+
+       panas1~~panas1
+         panas2~~panas2
+         panas3~~panas3
+         panas4~~panas4
+         panas5~~panas5
+       panas6~~panas6
+         panas7~~panas7
+         panas8~~panas8
+         panas9~~panas9
+         panas10~~panas10
+
+       posaff~~1*posaff
+       negaff~~1*negaff
+       posaff~~negaff '
>
> #I resort back to listwise deletion to simplify examples and LR tests
> fit <- sem(model=model, data = stack1, estimator="ML", group="widow", group.equal=c("loadings"))
> summary(object=fit,fit.measures=TRUE, rsquare=TRUE, standardized=TRUE)
lavaan (0.5-22) converged normally after  44 iterations

  Number of observations per group
  0                                               159
  1                                                40

  Estimator                                         ML
  Minimum Function Test Statistic              136.569
  Degrees of freedom                                78
  P-value (Chi-square)                           0.000

Chi-square for each group:
  0                                             80.163
  1                                             56.406

Model test baseline model:

  Minimum Function Test Statistic            1643.772
  Degrees of freedom                               90
  P-value                                       0.000

User model versus baseline model:
  Comparative Fit Index (CFI)                    0.962
  Tucker-Lewis Index (TLI)                       0.957

Loglikelihood and Information Criteria:
  Loglikelihood user model (H0)              -1970.251
  Loglikelihood unrestricted model (H1)      -1901.967

  Number of free parameters                        52
  Akaike (AIC)                               4044.502
  Bayesian (BIC)                             4215.754
```

```
  Sample-size adjusted Bayesian (BIC)         4051.015

Root Mean Square Error of Approximation:
  RMSEA                                          0.087
  90 Percent Confidence Interval          0.062  0.111
  P-value RMSEA <= 0.05                          0.010

Standardized Root Mean Square Residual:
  SRMR                                           0.129

Parameter Estimates:
  Information                                 Expected
  Standard Errors                             Standard

Group 1 [0]:
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  posaff =~
    panas1  (.p1.)    0.858    0.050   17.035    0.000    0.858    0.901
    panas2  (.p2.)    0.979    0.065   15.071    0.000    0.979    0.842
    panas3  (.p3.)    0.854    0.051   16.632    0.000    0.854    0.909
    panas4  (.p4.)    0.906    0.054   16.915    0.000    0.906    0.919
    panas5  (.p5.)    0.856    0.057   15.015    0.000    0.856    0.859
  negaff =~
    panas6  (.p6.)    0.868    0.051   17.098    0.000    0.868    0.941
    panas7  (.p7.)    0.694    0.059   11.672    0.000    0.694    0.723
    panas8  (.p8.)    0.786    0.054   14.484    0.000    0.786    0.853
    panas9  (.p9.)    0.396    0.050    7.973    0.000    0.396    0.552
    panas10 (.10.)    0.816    0.076   10.688    0.000    0.816    0.697

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  posaff ~~
    negaff           -0.711    0.043  -16.386    0.000   -0.711   -0.711

Group 2 [1]:

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  posaff =~
    panas1  (.p1.)    0.858    0.050   17.035    0.000    0.858    0.963
    panas2  (.p2.)    0.979    0.065   15.071    0.000    0.979    0.897
    panas3  (.p3.)    0.854    0.051   16.632    0.000    0.854    0.906
    panas4  (.p4.)    0.906    0.054   16.915    0.000    0.906    0.910
    panas5  (.p5.)    0.856    0.057   15.015    0.000    0.856    0.834
  negaff =~
    panas6  (.p6.)    0.868    0.051   17.098    0.000    0.868    0.905
    panas7  (.p7.)    0.694    0.059   11.672    0.000    0.694    0.757
    panas8  (.p8.)    0.786    0.054   14.484    0.000    0.786    0.800
    panas9  (.p9.)    0.396    0.050    7.973    0.000    0.396    0.481
    panas10 (.10.)    0.816    0.076   10.688    0.000    0.816    0.612

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  posaff ~~
    negaff           -0.405    0.139   -2.922    0.003   -0.405   -0.405
```

Illustration of equality constraints for just one parameter (loading for panas5 constrained equal across groups)[3]

```
model = '
       posaff =~ NA*panas1 + panas2 + panas3 + panas4 + c(l5,l5)*panas5
      negaff =~ NA*panas6 + panas7 + panas8 + panas9 + panas10

      panas1~~panas1
       panas2~~panas2
       panas3~~panas3
       panas4~~panas4
       panas5~~panas5
      panas6~~panas6
       panas7~~panas7
       panas8~~panas8
       panas9~~panas9
       panas10~~panas10

      posaff~~1*posaff
      negaff~~1*negaff
      posaff~~negaff '

#I resort back to listwise deletion to simplify examples and LR tests
fit <- sem(model=model, data = stack1, estimator="ML", group="widow")
summary(object=fit,fit.measures=TRUE, rsquare=TRUE, standardized=TRUE)

Group 1 [0]:

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  posaff =~
    panas1            0.753    0.053   14.254    0.000    0.753    0.873
    panas2            0.874    0.070   12.533    0.000    0.874    0.808
    panas3            0.799    0.053   14.989    0.000    0.799    0.898
    panas4            0.864    0.056   15.525    0.000    0.864    0.915
    panas5    (l5)    0.847    0.057   14.965    0.000    0.847    0.861
Group 2 [1]:

Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  posaff =~
    panas1            1.048    0.096   10.965    0.000    1.048    0.980
    panas2            1.202    0.128    9.424    0.000    1.202    0.929
    panas3            0.995    0.106    9.391    0.000    0.995    0.928
    panas4            1.012    0.109    9.316    0.000    1.012    0.925
    panas5    (l5)    0.847    0.057   14.965    0.000    0.847    0.832
```

[3] Please note that I am illustrating this individual loading constraint just to show how individual parameters can be constrained in lavaan. I do not recommend such tests for measurement invariance, because the choice of referent and the particular constraints can interact due to the interdependence of the indicator and factor variances (see Cheung & Lau, 2012; Cheung & Rensvold, 2002, for discussion of the issues).