Modeling Moderated Variance Ratios in Lavaan

Michael Smithson
The Australian National University

April 20, 2012

Introduction

The lavaan package (Rosseel, Y., 2012) allows restrictions to be imposed on variables that are functions of variances and covariances. Thus, for a categorical moderator, the equal variances ratio (EVR) test described in Smithson (2012) can be incorporated into a multi-groups structural equations model (SEM) via appropriate model comparisons. These SEMs also can incorporate tests of homogeneity of (error) variance and, of course, moderation of correlations and moderation of slopes.

EVR models are implemented in correlation SEMs of the kind described in Smithson (2012). The basic syntax for such an SEM is given below. The EVR component is initiated by defining the relevant variance ratios. Suppose the variance ratios of interest are for the i^{th} and j^{th} categories of the moderator. These variance ratios must first be defined (e.g., by the `ratioi := ...` statement). We can then compare a model that leaves them unconstrained against a model that imposes equality between them (via the statement `ratioi == ratioj`). Clearly multiple hypotheses of this kind can be tested simultaneously. Lavaan has the usual extractor functions such as `summary()`. Model comparisons can be made in the same fashion as many other packages in R.

```
# This function sets up an EVR model:
mod1 <- '
fy =~ c(sy1,sy2,...,syi,...syj,...syJ)*y
fx =~ c(sx1,sx2,...,sxi,...sxj,...sxJ)*x
fx ~~ fy
fx ~~ fx
fy ~~ fy
x ~ 1
y ~ 1
ratioi := sxi^2/syi^2
ratioj := sxj^2/syj^2
...
ratioi == ratioj
...
```

```
'
#
# Here is the default estimation method for the model.
cov1 <- sem(mod1, data = dataframe, group = "moderator name", std.lv = TRUE,
estimator = "ML", likelihood = "wishart")
```

Homogeneity of variance also can be tested in the correlation SEMs. For example, homogeneity of variances for x and y in the above model could be modeled by this command:

```
cov1 <- sem(modc1, data = eqdat, group = "z", std.lv = TRUE, group.equal
= "loadings", estimator = "ML", likelihood = "wishart")
```

The `group.equal = "loadings"` subcommand restricts the fx and fy parameters to be equal across all categories of the moderator. Of course, this also imposes EVR. Homogeneity of variance for x or y alone (which is incompatible with EVR) can be tested via the model syntax shown next, which illustrates this for homogeneity of variance in x.

```
modc1 <- '
fy =~ y
fx =~ c(sx1,sx1, ..., sx1)*x
fx~~fy
fx~~fx
fy~~fy
x ~ 1
y ~ 1
'
```

Regression SEMs can include tests for homogeneity of error variance. Syntax for doing so is shown below.

```
mod1r<- '
y ~ x
x ~ 1
y ~ 1
y ~~ c(se1, se1,..., se1)*y
x ~~ x
'
```

Regression SEMs also can incorporate EVR tests, but this matter is best left to illustration in the two examples to follow.

Examples of EVR-testing SEMs in lavaan

Two-Category Moderator

This is a two-category moderator example in which the null hypothesis of EVR is true. It uses an artificial data-set sampled from two bivariate normal distributions, associated with one category of a binary moderator variable z which takes values -1 and +1. For the first moderator category $\sigma_x^2 = 1$ and $\sigma_y^2 = 2$, while for the second category $\sigma_x^2 = 4$ and $\sigma_y^2 = 8$. Thus, the population variance ratio in both moderator categories is 1/2. The population covariances are 1 and the means are 0 for both categories. The data-file is eqdat.txt. There are 500 observations in each category. In the first category, $s_x^2 = 0.883$ and $s_y^2 = 1.958$, so $s_x^2/s_y^2 = 0.451$. In the second category, $s_x^2 = 4.068$ and $s_y^2 = 7.515$, so $s_x^2/s_y^2 = 0.541$.

We first fit a saturated model permitting unequal variance ratios:

```
modc0 <- '
fy =~ c(sy1,sy2)*y
fx =~ c(sx1,sx2)*x
fx~~fy
fx~~fx
fy~~fy
ratio1 := sx1^2/sy1^2
ratio2 := sx2^2/sy2^2
'
cov0 <- sem(modc0, data = eqdat, group = "z", std.lv = TRUE, likelihood = "wishart")
```

The output reproduces the sample variance ratios, 0.451 and 0.541.

```
summary(cov0)
Lavaan (0.4-10) converged normally after 36 iterations
  Number of observations per group
  -1                                                 500
  1                                                  500
  Estimator                                           ML
  Minimum Function Chi-square                      0.000
  Degrees of freedom                                   0
  P-value                                          1.000

Chi-square for each group:
  -1                                               0.000
  1                                                0.000

Parameter estimates:
  Information                                   Expected
  Standard Errors                               Standard

Group 1 [-1]:
                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  fy =~
    y       (sy1)     1.399    0.044   31.591    0.000
  fx =~
    x       (sx1)     0.940    0.030   31.591    0.000
```

```
Covariances:
  fy ~~
    fx                0.737    0.020   36.015    0.000
Intercepts:
    x                 0.033    0.042    0.777    0.437
    y                 0.049    0.063    0.787    0.431
    fy                0.000
    fx                0.000
Variances:
    fx                1.000
    fy                1.000
    y                 0.000
    x                 0.000
Defined parameters:
    ratio1            0.451    0.027   16.521    0.000
    ratio2            0.541    0.048   11.300    0.000

Group 2 [1]:
                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  fy =~
    y       (sy2)     2.741    0.087   31.591    0.000
  fx =~
    x       (sx2)     2.017    0.064   31.591    0.000
Covariances:
  fy ~~
    fx                0.152    0.044    3.469    0.001
Intercepts:
    x                -0.008    0.090   -0.092    0.927
    y                -0.052    0.123   -0.423    0.672
    fy                0.000
    fx                0.000
Variances:
    fx                1.000
    fy                1.000
    y                 0.000
    x                 0.000
Defined parameters:
    ratio1            0.451    0.027   16.521    0.000
    ratio2            0.541    0.048   11.300    0.000
```

The next model restricts the variance ratios to be equal:

```
modc1 <- '
fy =~ c(sy1,sy2)*y
fx =~ c(sx1,sx2)*x
fx~~fy
```

```
fx~~fx
fy~~fy
ratio1 := sx1^2/sy1^2
ratio2 := sx2^2/sy2^2
ratio1 == ratio2
'
cov1 <- sem(modc1, data = eqdat, group = "z", std.lv = TRUE, estimator = "ML", likelihood =
```

Its output shows that this model cannot be rejected ($\chi^2(1) = 2.914, p = .088$), and estimates the variance-ratio to be 0.478.

```
summary(cov1)
lavaan (0.4-13) converged normally after 170 iterations
  Number of observations per group
  -1                                                 500
  1                                                  500
  Estimator                                           ML
  Minimum Function Chi-square                      2.914
  Degrees of freedom                                   1
  P-value                                          0.088
Chi-square for each group:
  -1                                               0.929
  1                                                1.986
Parameter estimates:
  Information                                   Expected
  Standard Errors                               Standard
Group 1 [-1]:
                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  fy =~
    y       (sy1)     1.379    0.042   32.810    0.000
  fx =~
    x       (sx1)     0.954    0.029   32.811    0.000
Covariances:
  fy ~~
    fx                0.737    0.020   35.963    0.000
Intercepts:
    y                 0.049    0.062    0.798    0.425
    x                 0.033    0.043    0.766    0.444
    fy                0.000
    fx                0.000
Variances:
    fx                1.000
    fy                1.000
    y                 0.000
    x                 0.000
```

```
Group 2 [1]:
                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  fy =~
    y       (sy2)     2.831    0.073   38.673    0.000
  fx =~
    x       (sx2)     1.957    0.051   38.675    0.000
Covariances:
  fy ~~
    fx                0.151    0.044    3.462    0.001
Intercepts:
    y                -0.052    0.127   -0.410    0.682
    x                -0.008    0.088   -0.095    0.924
    fy                0.000
    fx                0.000
Variances:
    fx                1.000
    fy                1.000
    y                 0.000
    x                 0.000

Defined parameters:
    ratio1            0.478    0.024   20.009    0.000
    ratio2            0.478    0.024   20.009    0.000
Constraints:                               Slack (>=0)
    ratio1 - (ratio2)                                0.000
```

Regression SEMs also can incorporate EVR tests. The variance ratios have to be derived from the error-variances, whence the definitions of `ratio1` and `ratio2` in the syntax below.

```
modr1 <- '
y ~ c(b1,b2)*x
x ~~ c(sx1,sx2)*x
y ~~ c(se1,se2)*y
ratio1 := sx1/(se1 + sx1*(b1^2))
ratio2 := sx2/(se2 + sx2*(b2^2))
ratio1 == ratio2
'
reg1 <- sem(modr1, data = eqdat, group = "z", std.lv = TRUE, fixed.x=FALSE, estimator = "ML'
```

This type of model provides estimates of variances rather than standard deviations, which is slightly confusing when comparing its estimates with those of the correlation model. Nonetheless, its estimates are consistent with the correlation model; the ratio of the estimated correlations in the correlation model and regression coefficients in this model are identical (up to roundoff error): $.737/.151 = 0.488$ and $1.066/.219 = 0.487$.

```
lavaan (0.4-13) converged normally after 111 iterations
  Number of observations per group
  -1                                                 500
  1                                                  500
  Estimator                                           ML
  Minimum Function Chi-square                      2.914
  Degrees of freedom                                   1
  P-value                                          0.088
Chi-square for each group:

  -1                                               0.929
  1                                                1.986
Parameter estimates:
  Information                                   Expected
  Standard Errors                               Standard
Group 1 [-1]:
                   Estimate  Std.err  Z-value  P(>|z|)
Regressions:
  y ~
    x        (b1)     1.066    0.040   26.748    0.000
Intercepts:
    y                 0.014    0.042    0.346    0.729
    x                 0.033    0.043    0.766    0.444
Variances:
    x       (sx1)     0.909    0.055
    y       (se1)     0.871    0.053

Group 2 [1]:
                   Estimate  Std.err  Z-value  P(>|z|)
Regressions:
  y ~
    x        (b2)     0.219    0.064    3.449    0.001
Intercepts:
    y                -0.050    0.125   -0.400    0.689
    x                -0.008    0.088   -0.095    0.924
Variances:
    x       (sx2)     3.829    0.198
    y       (se2)     7.831    0.405

Defined parameters:
    ratio1            0.478    0.024   20.008    0.000
    ratio2            0.478    0.024   20.010    0.000
Constraints:                                Slack (>=0)
    ratio1 - (ratio2)                              0.000
```

SEM Example in Smithson (2012)

The data-set for this example is as described in Smithson (2012), and the file is semex.txt. Starting with SEMs for regression coefficients, we begin with a model that assumes HoV for X but otherwise leaves the other parameters free.

```
modr0 <- '
y ~ c(b1,b2)*x
x ~ 1
x ~~ c(sx1,sx1)*x
y ~~ c(se1,se2)*y
ratio1 := sx1/(se1 + sx1*(b1^2))
ratio2 := sx1/(se2 + sx1*(b2^2))
'
reg0 <- sem(modr0, data = ex5, group = "z", std.lv = TRUE, fixed.x=FALSE, estimator = "ML",
```

The results can be obtained with the `summary(reg0)` command. As reported in Smithson (2012), the model fit is $\chi^2(1) = 0.369(p = .544)$. Next, we estimate a model that imposes equal variances ratios.

```
modr1 <- '
y ~ c(b1,b2)*x
x ~ 1
x ~~ c(sx1,sx2)*x
y ~~ c(se1,se2)*y
ratio1 := sx1/(se1 + sx1*(b1^2))
ratio2 := sx2/(se2 + sx2*(b2^2))
ratio1 == ratio2
'
reg1 <- sem(modr1, data = ex5, group = "z", std.lv = TRUE, fixed.x=FALSE, estimator = "ML",
```

This model yields a chi-square fit of $\chi^2(1) = 81.971(p < .0001)$, so we may reject the EVR hypothesis. We then fit an equivalent regression model that relaxes EVR but forces equality of regression coefficients.

```
modr2 <- '
y ~ c(b1,b1)*x
x ~ 1
x ~~ c(sx1,sx1)*x
y ~~ c(se1,se2)*y
ratio1 := sx1/(se1 + sx1*(b1^2))
ratio2 := sx1/(se2 + sx1*(b1^2))
'
reg2 <- sem(modr2, data = ex5, group = "z", std.lv = TRUE, fixed.x=FALSE, estimator = "ML",
```

This model can be rejected because $\chi^2(2) = 15.746(p = .0004)$, and the fit comparison between this model and the reg0 model yields $\chi^2(1) = 15.746 - 0.369 = 15.377(p < .0001)$, suggesting significantly worse fit. So we conclude that there is moderation of slopes.

Turning now to SEMs for correlations, as before we begin with a model that assumes HoV for X but otherwise leaves the other parameters free.

```
modc0 <- '
fy =~ c(sy1,sy2)*y
fx =~ c(sx1,sx1)*x
fx~~ c(r1,r2)*fy
fx~~fx
fy~~fy
x ~ 1
y ~ 1
ratio1 := sx1^2/sy1^2
ratio2 := sx1^2/sy2^2
'
cov0 <- sem(modc0, data = ex5, group = "z", std.lv = TRUE, estimator = "ML", likelihood = "w
```

The results can be obtained with the `summary(cov0)` command, and as with the corresponding regression SEM, the model fit is $\chi^2(1) = 0.369(p = .544)$. Next, a model restricting the variance ratios to equality is estimated:

```
modc1 <- '
fy =~ c(sy1,sy2)*y
fx =~ c(sx1,sx2)*x
fx~~ c(r1,r2)*fy
fx~~fx
fy~~fy
x ~ 1
y ~ 1
ratio1 := sx1^2/sy1^2
ratio2 := sx2^2/sy2^2
ratio1==ratio2
'
cov1 <- sem(modc1, data = ex5, group = "z", std.lv = TRUE, estimator = "ML", likelihood = "w
```

As with its regression counterpart, this model yields a chi-square fit of $\chi^2(1) = 81.971(p < .0001)$. Finally, we estimate an equal-correlation model.

```
modc2 <- '
fy =~ c(sy1,sy2)*y
fx =~ c(sx1,sx1)*x
fx~~ c(r1,r1)*fy
fx~~fx
fy~~fy
x ~ 1
y ~ 1
ratio1 := sx1^2/sy1^2
ratio2 := sx1^2/sy2^2
'
cov2 <- sem(modc2, data = ex5, group = "z", std.lv = TRUE, estimator = "ML", likelihood = "w
```

The results are as reported in Smithson (2012),with $\chi^2(2) = 0.451(p = .798)$. The model comparison result is $\chi^2(1) = 0.082(p = .775)$, suggesting that the equal-correlations model is a good fit to the data. The estimates can be displayed via the `summary(cov2)` command.

Four-Category Moderator

Now we consider a four-category moderator example, with EVR for the first two categories and for the second two, but not for both pairs of categories. We generate an artificial data-set sampled from four bivariate normal distributions, associated with a moderator variable z which takes values 1, 2, 3, and 4, with 300 observations in each category. The variance ratio for the first two moderator categories equals 1/2, whereas for the third and fourth categories the ratio is 1/6. The first two categories' correlations also are identical. The data-set is named fourdat.txt. We begin with a model that tests both EVR hypotheses and the equal-correlations hypothesis.

```
modc0 <- '
   fy =~ c(sy1,sy2,sy3,sy4)*y
   fx =~ c(sx1,sx2,sx3,sx4)*x
   fx ~~ c(r1,r1,r3,r4)*fy
   ratio1 := sx1^2/sy1^2
   ratio2 := sx2^2/sy2^2
   ratio3 := sx3^2/sy3^2
   ratio4 := sx4^2/sy4^2
   ratio1 == ratio2
   ratio3 == ratio4
'
cov0 <- sem(modc0, data = fourgp, group = "group", std.lv = TRUE, estimator = "ML", likeliho
```

The results are shown below. The model cannot be rejected ($\chi^2(3) = 2.242, p = .524$), and estimates the variance-ratios to be 0.506 and 0.176.

```
lavaan (0.4-13) converged normally after 342 iterations
  Number of observations per group
  1                                                300
  2                                                300
  3                                                300
  4                                                300
  Estimator                                         ML
  Minimum Function Chi-square                    2.242
  Degrees of freedom                                 3
  P-value                                        0.524
Chi-square for each group:
  1                                              1.118
  2                                              1.118
  3                                              0.002
```

```
  4                                             0.004
Parameter estimates:
  Information                                Observed
  Standard Errors                         First.order
Group 1 [1]:
                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  fy =~
    y       (sy1)     1.444    0.050   28.822    0.000
  fx =~
    x       (sx1)     1.027    0.038   27.259    0.000
Covariances:
  fy ~~
    fx       (r1)     0.734    0.019   37.859    0.000
Intercepts:
    y                -0.082    0.083   -0.980    0.327
    x                -0.040    0.060   -0.665    0.506
    fy                0.000
    fx                0.000
Variances:
    y                 0.000
    x                 0.000
    fy                1.000
    fx                1.000

Group 2 [2]:
                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  fy =~
    y       (sy2)     1.734    0.074   23.589    0.000
  fx =~
    x       (sx2)     1.233    0.049   25.307    0.000
Covariances:
  fy ~~
    fx       (r1)     0.734    0.019   37.859    0.000
Intercepts:
    y                 0.029    0.104    0.281    0.779
    x                -0.001    0.073   -0.014    0.989
    fy                0.000
    fx                0.000
Variances:
    y                 0.000
    x                 0.000
    fy                1.000
    fx                1.000
```

```
Group 3 [3]:
                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  fy =~
    y       (sy3)     2.399    0.096   25.023    0.000
  fx =~
    x       (sx3)     1.007    0.044   22.990    0.000
Covariances:
  fy ~~
    fx       (r3)     0.833    0.019   44.818    0.000
Intercepts:
    y                 0.135    0.139    0.965    0.335
    x                 0.005    0.059    0.093    0.926
    fy                0.000
    fx                0.000
Variances:
    y                 0.000
    x                 0.000
    fy                1.000
    fx                1.000

Group 4 [4]:
                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  fy =~
    y       (sy4)     3.326    0.144   23.126    0.000
  fx =~
    x       (sx4)     1.396    0.060   23.176    0.000
Covariances:
  fy ~~
    fx       (r4)     0.495    0.049   10.084    0.000
Intercepts:
    y                 0.282    0.194    1.456    0.145
    x                 0.032    0.081    0.390    0.697
    fy                0.000
    fx                0.000
Variances:
    y                 0.000
    x                 0.000
    fy                1.000
    fx                1.000

Defined parameters:
    ratio1            0.506    0.040   12.699    0.000
    ratio2            0.506    0.040   12.751    0.000
    ratio3            0.176    0.011   15.424    0.000
```

```
    ratio4            0.176    0.017   10.325    0.000

Constraints:                               Slack (>=0)
    ratio1 - (ratio2)                            0.000
    ratio3 - (ratio4)                            0.000
```

We then estimate a model that constrains all variance ratios to be equal. The syntax is shown below. This model can be rejected ($\chi^2(3) = 176.254, p < .0001$).

```
modc1 <- '
fy =~ c(sy1,sy2,sy3,sy4)*y
fx =~ c(sx1,sx2,sx3,sx4)*x
fx~~fy
fx~~fx
fy~~fy
x ~ 1
y ~ 1
ratio1 := sx1^2/sy1^2
ratio2 := sx2^2/sy2^2
ratio3 := sx3^2/sy3^2
ratio4 := sx4^2/sy4^2
ratio1 == ratio2
ratio1 == ratio3
ratio1 == ratio4
'
```

As noted earlier, is also possible to incorporate EVR tests in a regression SEM. In this example, the hypothesis that the correlations in the first two categories are identical is equivalent to the hypothesis that the corresponding regression coefficients are identical. The following syntax extracts the EVR hypotheses from the error-variances and variances for X.

```
modr1<- '
y ~ c(b1,b1,b3,b4)*x
x ~ 1
y ~~ c(se1,se2,se3,se4)*y
x ~~ c(sx1,sx2,sx3,sx4)*x
ratio1 := sx1/(se1 + (b1^2)*(sx1))
ratio2 := sx2/(se2 + (b1^2)*(sx2))
ratio3 := sx3/(se3 + (b3^2)*(sx3))
ratio4 := sx4/(se4 + (b4^2)*(sx4))
ratio1 == ratio2
ratio3 == ratio4
'
```

The output is shown next. As noted earlier this type of model provides estimates of variances rather than standard deviations, but the estimates are consistent with those in the correlation model. Most importantly, due to the EVR for the third and fourth categories of the moderator, we should expect the ratios of the estimated correlations and regression coefficients in those two groups to be the same, and indeed they are (within roundoff error): $.833/.495 = 1.984/1.181 = 1.68$.

```
lavaan (0.4-13) converged normally after 148 iterations
  Number of observations per group
  1                                                 300
  2                                                 300
  3                                                 300
  4                                                 300
  Estimator                                          ML
  Minimum Function Chi-square                     2.242
  Degrees of freedom                                  3
  P-value                                         0.524
Chi-square for each group:
  1                                               1.118
  2                                               1.118
  3                                               0.002
  4                                               0.004
Parameter estimates:
  Information                                  Expected
  Standard Errors                              Standard
Group 1 [1]:
                   Estimate  Std.err  Z-value  P(>|z|)
Regressions:
  y ~
    x        (b1)     1.032    0.039   26.437    0.000
Intercepts:
    x                -0.040    0.059   -0.668    0.504
    y                -0.041    0.057   -0.719    0.472
Variances:
    y       (se1)     0.961    0.068
    x       (sx1)     1.055    0.075

Group 2 [2]:
                   Estimate  Std.err  Z-value  P(>|z|)
Regressions:
  y ~
    x        (b1)     1.032    0.039   26.437    0.000
Intercepts:
    x                -0.001    0.071   -0.014    0.989
    y                 0.030    0.068    0.445    0.657
Variances:
    y       (se2)     1.386    0.098
```

```
    x       (sx2)     1.521    0.108

Group 3 [3]:
                   Estimate  Std.err  Z-value  P(>|z|)
Regressions:
  y ~
    x        (b3)     1.984    0.068   29.058    0.000
Intercepts:
    x                 0.005    0.058    0.093    0.926
    y                 0.124    0.077    1.611    0.107
Variances:
    y       (se3)     1.767    0.141
    x       (sx3)     1.014    0.081

Group 4 [4]:
                   Estimate  Std.err  Z-value  P(>|z|)
Regressions:
  y ~
    x        (b4)     1.181    0.109   10.855    0.000
Intercepts:
    x                 0.032    0.081    0.391    0.696
    y                 0.245    0.167    1.465    0.143
Variances:
    y       (se4)     8.348    0.584
    x       (sx4)     1.948    0.136

Defined parameters:
    ratio1            0.506    0.028   18.006    0.000
    ratio2            0.506    0.028   18.007    0.000
    ratio3            0.176    0.010   18.510    0.000
    ratio4            0.176    0.010   18.510    0.000

Constraints:                               Slack (>=0)
    ratio1 - (ratio2)                            0.000
    ratio3 - (ratio4)                            0.000
```

References

Rosseel, Y. (2012). lavaan: Latent variable analysis [Computer software manual]. Retrieved from http://CRAN.R-project.org/package=lavaan (R package version 0.4-13).

Smithson, M. (2012). A simple statistic for comparing moderation of slopes and correlations. The Australian National University, Canberra, Australia: Unpublished manuscript.