Being negative can be good for your test

Leonard Vanbrabant

Rens van de Schoot

Yves Rosseel

4th Constrained Statistical Inference Meeting
March 4, 2016 – Tilburg University

lavaan 0.3-1 (first public version, May 2010)

```
Model converged normally after 35 iterations using ML

  Minimum Function Chi-square                    85.306
  Degrees of freedom                                 24
  P-value                                        0.0000

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  visual =~
    x1                1.000
    x2                0.554    0.100    5.554    0.000
    x3                0.729    0.109    6.685    0.000
  textual =~
    x4                1.000
    x5                1.113    0.065   17.014    0.000
    x6                0.926    0.055   16.703    0.000
  speed =~
    x7                1.000
    x8                1.180    0.165    7.152    0.000
    x9                1.082    0.151    7.155    0.000

Latent covariances:
  visual ~~
    textual           0.408    0.074    5.552    0.000
    speed             0.262    0.056    4.660    0.000
  textual ~~
```

```
    speed             0.173    0.049    3.518    0.000

 Latent variances:
    visual            0.809    0.145    5.564    0.000
    textual           0.979    0.112    8.737    0.000
    speed             0.384    0.086    4.451    0.000

 Residual variances:
    x1                0.549    0.114    4.833    0.000
    x2                1.134    0.102   11.146    0.000
    x3                0.844    0.091    9.317    0.000
    x4                0.371    0.048    7.778    0.000
    x5                0.446    0.058    7.642    0.000
    x6                0.356    0.043    8.277    0.000
    x7                0.799    0.081    9.823    0.000
    x8                0.488    0.074    6.573    0.000
    x9                0.566    0.071    8.003    0.000
```

lavaan 0.4-10 (Oct 2011)

```
Lavaan (0.4-10) converged normally after 41 iterations

  Number of observations                           301

  Estimator                                         ML
  Minimum Function Chi-square                   85.306
  Degrees of freedom                                24
  P-value                                        0.000

Parameter estimates:

  Information                                 Expected
  Standard Errors                             Standard

                   Estimate  Std.err  Z-value  P(>|z|)
Latent variables:
  visual =~
    x1                1.000
    x2                0.553    0.100    5.554    0.000
    x3                0.729    0.109    6.685    0.000
  textual =~
    x4                1.000
    x5                1.113    0.065   17.014    0.000
    x6                0.926    0.055   16.703    0.000
  speed =~
    x7                1.000
```

```
    x8                1.180    0.165    7.152    0.000
    x9                1.082    0.151    7.155    0.000

 Covariances:
  visual ~~
    textual           0.408    0.074    5.552    0.000
    speed             0.262    0.056    4.660    0.000
  textual ~~
    speed             0.173    0.049    3.518    0.000

 Variances:
    x1                0.549    0.114
    x2                1.134    0.102
    x3                0.844    0.091
    x4                0.371    0.048
    x5                0.446    0.058
    x6                0.356    0.043
    x7                0.799    0.081
    x8                0.488    0.074
    x9                0.566    0.071
    visual            0.809    0.145
    textual           0.979    0.112
    speed             0.384    0.086
```

lavaan 0.5-19 (Oct 2015)

```
lavaan (0.5-19) converged normally after  35 iterations

  Number of observations                           301

  Estimator                                         ML
  Minimum Function Test Statistic               85.306
  Degrees of freedom                                24
  P-value (Chi-square)                           0.000

Parameter Estimates:

  Information                                 Expected
  Standard Errors                             Standard

Latent Variables:
                   Estimate  Std.Err  Z-value  P(>|z|)
  visual =~
    x1                1.000
    x2                0.554    0.100    5.554    0.000
    x3                0.729    0.109    6.685    0.000
  textual =~
    x4                1.000
    x5                1.113    0.065   17.014    0.000
    x6                0.926    0.055   16.703    0.000
  speed =~
    x7                1.000
```

```
    x8                1.180    0.165    7.152    0.000
    x9                1.082    0.151    7.155    0.000

Covariances:
                   Estimate  Std.Err  Z-value  P(>|z|)
  visual ~~
    textual           0.408    0.074    5.552    0.000
    speed             0.262    0.056    4.660    0.000
  textual ~~
    speed             0.173    0.049    3.518    0.000

Variances:
                   Estimate  Std.Err  Z-value  P(>|z|)
    x1                0.549    0.114    4.833    0.000
    x2                1.134    0.102   11.146    0.000
    x3                0.844    0.091    9.317    0.000
    x4                0.371    0.048    7.779    0.000
    x5                0.446    0.058    7.642    0.000
    x6                0.356    0.043    8.277    0.000
    x7                0.799    0.081    9.823    0.000
    x8                0.488    0.074    6.573    0.000
    x9                0.566    0.071    8.003    0.000
    visual            0.809    0.145    5.564    0.000
    textual           0.979    0.112    8.737    0.000
    speed             0.384    0.086    4.451    0.000
```

asymptotic null distribution of likelihood-based tests

- simple hypothesis: $H_0 : \theta = \theta_0$ versus $H_a : \theta \neq \theta_0$
- Wald test, Score test, LR test with test statistic T_W, T_S, T_{LR}
- it can be shown that 'if regularity conditions hold' all three test statistics follow a chi-square distribution under H_0 if $n \rightarrow \infty$
- for the LRT, the proof dates back to Wilks 1938
- what are these 'regularity conditions':
  - they are needed to establish the asymptotic normality of ML estimates
  - the one that matters here is:

    the true parameter θ (scalar or vector) must be an interior point of the parameter space Θ
  - in other words: θ should NOT be a boundary point
- see for example Boos & Stefanski (2013): page 284 (scalar), page 286 (vector), page 291 (discussion)

two nonstandard settings

- the alternative hypothesis is order-restricted
  - for example:

$$H_0 : \mu_1 = \mu_2 = \mu_2 \quad \text{versus} \quad H_a : \mu_1 < \mu_2 < \mu_3$$

  - we need different test statistics; see Silvapulle and Sen (2005)
- the null hypothesis value, θ_0, lies on the boundary of the parameter space, say $\theta_0 = b$
  - the parameter space is restricted a priori (eg. $\theta \geq b$)
  - we violate the boundary assumption: $\theta_0 = b$ is not an interior point
  - original reference: Chernoff (1954)
  - one-sided tests are a logical choice:

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_a : \theta > \theta_0$$

a typical example: inference on variance components

- consider a random-intercepts model:

$$Y_{ij} = x_{ij}^T \beta + b_i + \epsilon_{ij}$$

  where Y_{ij} is the response for member $j = 1, \ldots, n_i$ of cluster $i = 1, \ldots, N$; x_{ij} is a vector of known covariate values; β is a vector of unknown regression coefficients; and $b_i \sim N(0, \tau^2)$ is a cluster-specific random effect, independently distributed from the residual-error components $\epsilon_{ij} \sim N(0, \sigma^2)$

- two-sided test:

$$H_0 : \tau^2 = 0 \quad \text{versus} \quad H_a : \tau^2 \neq 0$$

  – if we assume the parameter space is restricted ($\tau^2 \geq 0$), then 0 is on the boundary, and we can not use the classical test statistics

- one-sided test:

$$H_0 : \tau^2 = 0 \quad \text{versus} \quad H_a : \tau^2 > 0$$

  – we can use special one-sided test statistics; Wald, Score and LR type statistics have been available for a while now

interlude

- suppose you are fitting a random-intercept model (or any mixed model) using your favorite (but non-Bayesian) software
- you are interested in inference about variance components
- do you get a proper one-sided test?

question

- what if . . . variance components could be negative?

negative variance components

- some of you may reject this immediately on mathematical (or religious, or philosophical) grounds
- but what if we treat those variance components in a more pragmatic way: as model parameters that are just part of a bigger story
- references in the world of SEM:
  - Kolenikov and Bollen (2012). Testing negative error variances: is a Heywood case a symptom of misspecification? *Sociological Methods & Research, 41*, 124–167
  - Savalei and Kolenikov (2008). Constrained versus unconstrained estimation in structural equation modeling. *Psychological Methods, 13*, 150–170.
- in the world of mixed models, negative variance (components) have been discussed by Nelder (1954), Thompson (1962), Searle, Casella and McCulloh (1992), Verbeke and Molenberghs (2000)

mixed models: hierarchical versus marginal view

- the hierchical model:

$$Y_{ij} = x_{ij}^T \beta + b_i + \epsilon_{ij}, \quad \text{where} \quad b_i \sim N(0, \tau^2) \quad \text{and} \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

  - variances must be positive

- the marginal model:
  - collecting all Y_{ij} from the same cluster into a (long) vector Y_i
  - the marginal dstribution of Y_i is given by

$$Y_i \sim N(X_i \beta, V_i) \quad \text{where} \quad V_i = \tau^2 J_{n_i} + \sigma^2 I_{n_i}$$

  - negative values for τ^2 (or σ^2) are allowed, as long as V_i is positive definite
  - but in this case, a hierarchical interpretation is no longer possible as there is no random-effects structure that would yield such a marginal model

but there is more

- the hierarchical model can naturally be estimated using Bayesian estimation
- the hierarchical model can not be estimated in a frequentist framework: the random effects are treated as unobserved (latent) variables, and they must be integrated out
- so although we may 'specify' the model, as if it has a hierachical structure, we always transform it to a marginal model
- consequently, at least from an estimation point of view, we can allow for negative variance components
- if we opt for this unconstrained, marginal approach
  - we may get negative variances
  - but standard tests (Wald, Score, LR) are no problem
  - negative variances can be informative (misspecification? negative intra-class correlation?)

software aspects

- even if you adopt the marginal view, your software package may decide otherwise
- many popular software packages for fitting mixed models enforce strictly positive variance components (by default)
- if you are lucky, you can change the default, and ask for an unconstrained solution
- most SEM packages use unconstrained estimation per default
- as a user, you should be aware of the default options

software defaults (variances)

software	constrained	unconstrained	user-controllable
HLM 6	x		
R lme	x		
R lmer	x		
MLwin	x (>level 1)	x (level 1)	x
SAS proc mixed	x		x
SPSS mixed	x		
Stata (xt)mixed	x		
AMOS		x	
EQS	x		x
LISREL		x	x
lavaan		x	x
Mplus		x	x
Stata sem	x		

empirical example

- Alzheimer data (Hand & Taylor, 1987), but modified to get a negative variance for the random intercept
- 100 patients, 5 repeated measures of a score (number or words that a patient could correctly recall from a list); time is coded as 0,1,2,3,4
- model: linear growth curve with random intercept and random slope
- we can treat this as 2-level data (repeated measures nested within patients)
- to mimic MLM software, we constrained the residual variances of the 5 measures to be the same over time when using SEM software

lavaan syntax

```
library(lavaan); Data <- read.csv("heywood_wide.csv")

model <- '
    i =~ 1*t1 + 1*t2 + 1*t3 + 1*t4 + 1*t5
    s =~ 0*t1 + 1*t2 + 2*t3 + 3*t4 + 4*t5

    t1 + t2 + t3 + t4 + t5 ~ 0*1

    # residual variance
    t1~~a*t1
    t2~~a*t2
    t3~~a*t3
    t4~~a*t4
    t5~~a*t5

    # fixed effects
    i ~ 1
    s ~ 1

    # variance components
    s ~~ s
    i ~~ i
    s ~~ i
'
fit <- sem(model, data = Data)
summary(fit)
```

lavaan output

```
lavaan (0.5-19) converged normally after  42 iterations

  Number of observations                           100

  Estimator                                         ML
  Minimum Function Test Statistic               10.127
  Degrees of freedom                                14
  P-value (Chi-square)                           0.753

Parameter Estimates:

  Information                                 Expected
  Standard Errors                             Standard

Latent Variables:
                   Estimate  Std.Err  Z-value  P(>|z|)
  i =~
    t1                1.000
    t2                1.000
    t3                1.000
    t4                1.000
    t5                1.000
  s =~
    t1                0.000
    t2                1.000
    t3                2.000
```

```
    t4                  3.000
    t5                  4.000

Covariances:
                    Estimate  Std.Err  Z-value  P(>|z|)
  i ~~
    s                  0.329    0.183    1.796    0.073

Intercepts:
                    Estimate  Std.Err  Z-value  P(>|z|)
    t1                  0.000
    t2                  0.000
    t3                  0.000
    t4                  0.000
    t5                  0.000
    i                   9.761    0.154   63.208    0.000
    s                   3.076    0.093   33.000    0.000

Variances:
                    Estimate  Std.Err  Z-value  P(>|z|)
    t1          (a)     5.297    0.432   12.247    0.000
    t2          (a)     5.297    0.432   12.247    0.000
    t3          (a)     5.297    0.432   12.247    0.000
    t4          (a)     5.297    0.432   12.247    0.000
    t5          (a)     5.297    0.432   12.247    0.000
    s                   0.339    0.130    2.603    0.009
    i                  -0.794    0.425   -1.865    0.062
```

lmer input

```
> library(lme4); Data <- read.table("heywood.dat", header = TRUE)
> fit <- lmer(y ~ 1 + time + (1 + time | id), data = Data, REML = FALSE)
> summary(fit)

Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: y ~ 1 + time + (1 + time | id)
   Data: Data

     AIC      BIC   logLik deviance df.resid
  2368.3   2393.6  -1178.2   2356.3      494

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.64050 -0.64722 -0.03457  0.61837  3.01296

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 id       (Intercept) 0.007527 0.08676
          time        0.442050 0.66487  1.00
 Residual             4.962683 2.22771
Number of obs: 500, groups:  id, 100

Fixed effects:
            Estimate Std. Error t value
(Intercept)  9.76050    0.17278   56.49
time         3.07582    0.09687   31.75
```

overview

	β_{00}	β_{10}	σ^2_ϵ	τ^2_0	τ^2_1	τ^2_{01}
HLM	9.761 (.173)	3.076 (.098)	4.973 (np)	0.015 (np)	.448 (np)	0.057
lme	9.761 (.173)	3.076 (.098)	4.975 (np)	0.008 (np)	.459 (np)	0.043
lmer	9.761 (.173)	3.076 (.097)	4.975 (np)	0.008 (np)	.448 (np)	0.058
MLwin	9.761 (.173)	3.076 (.099)	4.966 (.351)	0.000 (-)	.481 (.092)	0.000
SAS	9.761 (.173)	3.076 (.097)	4.979 (.353)	0.000 (-)	.434 (.134)	0.081
SPSS	9.761 (.173)	3.076 (.096)	4.971 (.352)	0.000 (-)	.433 (.127)	0.083
Stata	9.761 (.173)	3.076 (.097)	4.963 (.353)	0.008 (.005)	.442 (.088)	0.058
AMOS	9.761 (.155)	3.076 (.094)	5.297 (.435)	-0.794 (.428)	0.339 (.131)	0.329
EQS	9.761 (.174)	3.076 (.097)	5.016 (.412)	0.000 (.494)	0.432 (.139)	0.081
lavaan	9.761 (.154)	3.076 (.093)	5.297 (.432)	-0.794 (.425)	0.339 (.130)	0.329
Mplus	9.761 (.154)	3.076 (.093)	5.297 (.432)	-0.794 (.425)	0.339 (.130)	0.329
LISREL	9.761 (.156)	3.076 (.094)	5.350 (.439)	-0.801 (.432)	0.343 (.132)	0.332
Stata	9.761 (.173)	3.076 (.098)	4.963 (.348)	0.001 (.000)	0.465 (.010)	0.026

discussion

- is this old news?
- at least Molenberghs & Verbeke have discussed this in several papers in the (bio)statistics literature
- in the SEM world, we have a few papers (Savalei & Kolenikov, Kolenikov & Bollen)
- the discussion of ‘inference for variance components’ in handbooks written for applied users in psychology are confusing at best (if not plain wrong)
- should we (read: Leonard) write a paper about this?