Inference with Heywood cases

Stas Kolenikov

Joint work with Kenneth Bollen (UNC) and Victoria Savalei (UBC)
NSF support from SES-0617193 with funds from SSA

September 18, 2009

What is a Heywood case?

Heywood (1931) considered characterization of the correlation matrices:

$$\begin{pmatrix} 1 & r_1 r_2 & r_1 r_3 & \dots & r_1 r_n \\ r_1 r_2 & 1 & r_2 r_3 & \dots & r_2 r_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_1 r_n & r_2 r_n & r_3 r_n & \dots & 1 \end{pmatrix}, r_1 \geq r_2 \geq \dots \geq r_n \geq 0$$

He showed that for this scheme

$$r_1^2 \leq 1 + \frac{1}{\frac{r_2^2}{1-r_2^2} + \frac{r_3^2}{1-r_3^2} + \dots \frac{r_n^2}{1-r_n^2}}$$

and hence r_1 can be greater than 1, which was an extension on the earlier results by Spearman and Garnett.

What is a Heywood case?

- Factor analysis literature of early 1960s: correlations greater than 1?
- Modern SEM usage: improper solutions, correlation estimates greater than 1, variance estimates less than 0

The current project deals mostly with variance estimates.

- Indicator variances
- Factor variances

Causes of Heywood cases

- Outliers (Bollen 1987).
- Non-convergence and underidentification (Van Driel 1978, Boomsma & Hoogland 2001).
- Empirical underidentification (Rindskopf 1984).
- Structurally misspecified models (Van Driel 1978, Dillon, Kumar & Mulani 1987, Sato 1987, Bollen 1989, Kolenikov & Bollen 2008).
- Sampling fluctuations (Van Driel 1978, Boomsma 1983, Anderson & Gerbing 1984).

Typology of Heywood cases

Inference problem:

$H_+ : \sigma^2 > 0$	"Normal" situation
$H_\emptyset : \sigma^2 = 0$	Perfect indicator (!?) or no factor present
$H_- : \sigma^2 < 0$	Misspecified model

Population value	Sample estimate $\hat\sigma^2 > 0$	Sample estimate $\hat\sigma^2 < 0$
$\sigma^2 > 0$	☺	??
$\sigma^2 = 0$	☺	??
$\sigma^2 < 0$	Type I error	Misspecification detected!

Breakdown of the papers

Savalei & Kolenikov (2008)

- Heywood case with an error variance.
- True situation is $H_\emptyset$.
- Should we consider H_+ or $H_+ \cup H_-$ as an alternative?

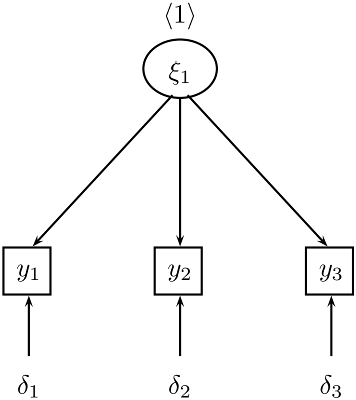
Kolenikov & Bollen (2008)

- Heywood case with an error variance.
- True situation is H_-; the model is structurally misspecified.
- Does it cause problems for inference?

Work in progress with Vika

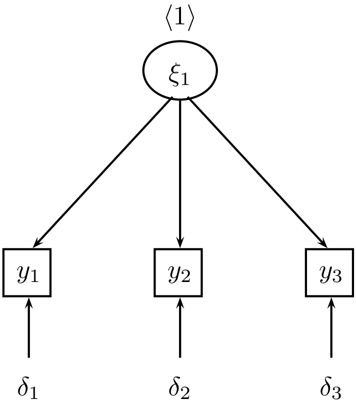
- Heywood case with an factor variance.
- True situation is $H_\emptyset$, and there are even more serious regularity condition violations.
- Does it cause problems for inference?

Is that really a variance?



$$
\begin{aligned}
\sigma_{11} &= \lambda_1^2 + \theta_1 \\
\sigma_{12} &= \lambda_1 \lambda_2 \\
\sigma_{13} &= \lambda_1 \lambda_3 \\
\sigma_{22} &= \lambda_2^2 + \theta_2 \\
\sigma_{23} &= \lambda_2 \lambda_3 \\
\sigma_{33} &= \lambda_3^2 + \theta_3
\end{aligned}
$$

Is that really a variance?



$$\hat\lambda_1 = \sqrt{s_{12}s_{13}/s_{23}}$$
$$\hat\lambda_2 = \sqrt{s_{12}s_{23}/s_{13}}$$
$$\hat\lambda_3 = \sqrt{s_{13}s_{23}/s_{12}}$$
$$\hat\theta_1 = s_{11} - s_{12}s_{13}/s_{23}$$
$$\hat\theta_2 = s_{22} - s_{12}s_{23}/s_{13}$$
$$\hat\theta_3 = s_{33} - s_{13}s_{23}/s_{12}$$

Questions?

Savalei and Kolenikov (2008)

Savalei, V. and Kolenikov, S. (2008), ‘Constrained vs. unconstrained estimation in structural equation modeling’, *Psychological Methods* **13**, 150–170.

Research question: if the truth is $H_\emptyset$, should we test against H_+ or against $H_- \cup H_+$?

Constrained estimation

- Irregular problem: the standard regularity condition of interior point is violated.
- Chernoff (1954): test with normal data of $\mu = 0$ vs. $\mu > 0$.
- Shapiro (1985): geometry of constrained parameter spaces in SEM.
- Stram & Lee (1994): variance components in mixed models.
- Jamshidian & Bentler (1994): algorithms for constrained estimation for SEM.
- Andrews (1999, 2001) established the most general results.
- Stoel, Garre, Dolan & van den Wittenboer (2006): computational procedure for testing $H_\emptyset$ vs. H_+.

Constrained estimation

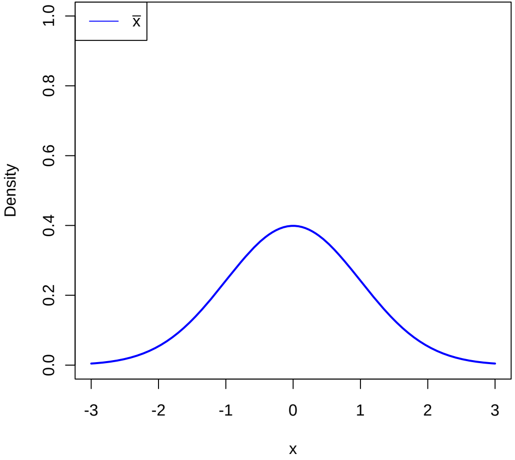
If there are k parameters and r boundaries present, the (asymptotic) distribution is

$$T \sim \sum_{j=k-r}^{k} w_j \chi^2_j, \quad \text{where} \sum_{j=k-r}^{k} w_j = 1$$

with weights w_j that depend on information matrix, i.e., covariances between parameters.

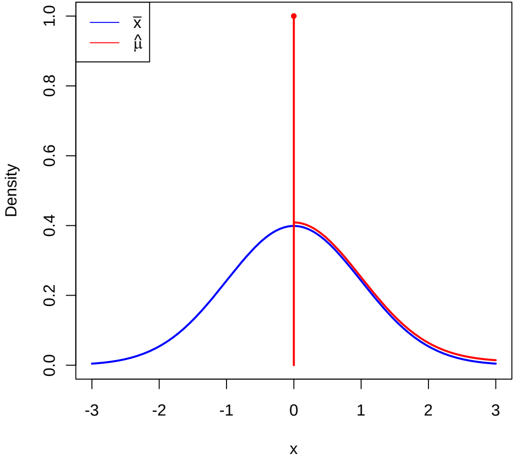
Constrained estimation

Chernoff (1954): $X \sim N(\mu, 1)$;
distribution of the test $H_\emptyset$ vs. H_+ is $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$



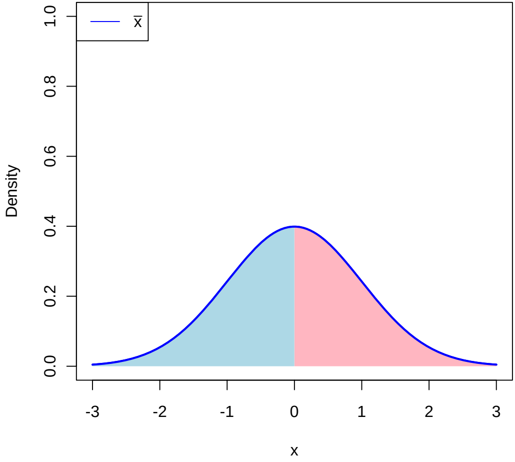
Constrained estimation

Chernoff (1954): $X \sim N(\mu, 1)$;
distribution of the test $H_\emptyset$ vs. H_+ is $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$



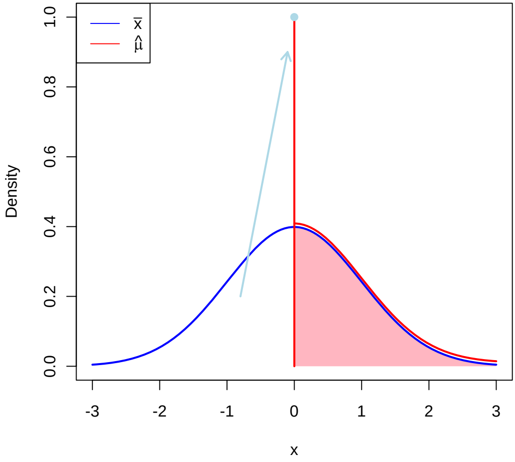
Constrained estimation

Chernoff (1954): $X \sim N(\mu, 1)$;
distribution of the test $H_\emptyset$ vs. H_+ is $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$



Constrained estimation

Chernoff (1954): $X \sim N(\mu, 1)$;
distribution of the test $H_\emptyset$ vs. H_+ is $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$



Heywood cases

Constrained?

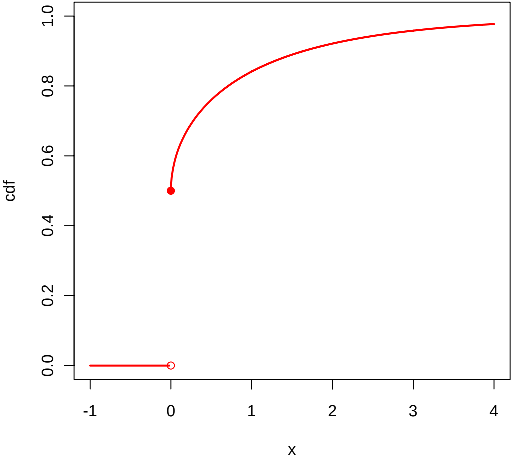
Misspecified?

Factor Heywood cases

References

Constrained estimation

Chernoff (1954): $X \sim N(\mu, 1)$;
distribution of the test $H_\emptyset$ vs. H_+ is $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$



Constrained estimation: cons

Savalei & Kolenikov (2008):

- Mixtures arise *only* when a combination of conditions about true parameters (unknown to the researcher) and estimation procedures (constrained estimation) occur together.
- Test of overall fit has distribution which is impossible to characterize (due to unknowns). Conservative upper bound: χ^2_k.
- Constrained estimation is only internally consistent when all the procedures are laid out ahead rather than followed the data!
- Ad hoc procedure of resetting Heywood cases to zero: effectively equivalent to constrained estimation, but type I error is likely inflated.
- Software implementation: explicit constraints are required; a far more difficult optimization procedure.
- Non-normal data and not asymptotically robust situations: just forget it.

Conditional approach

Dijkstra (1992): start with the first stage estimation; if Heywood cases occur:

1. restrict parameter(s);
2. release degree(s) of freedom;
3. re-estimate the model and report p-value only. The test statistic itself is not meaningful.

In this conditional approach, the *unconditional* distribution is again the mixture. It also leads to somewhat improved finite sample approximation when the parameter is *near* the boundary.

Implementation: SAS PROC CALIS

Unconstrained estimation

Savalei & Kolenikov (2008):

- Overall fit test has a known distribution, same for all points in parameter space!
- Decomposition of χ^2 into the fit and effect of the boundary components.
- Easier to implement than constrained estimation.
- More informative about sources of misfit.
- Provides power against a broader range of alternatives.

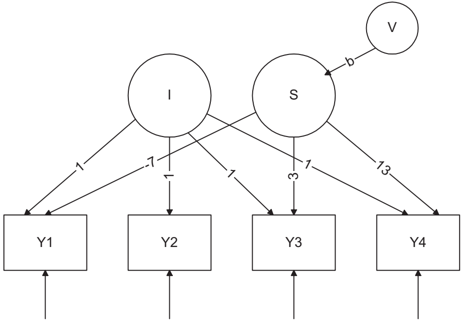
Software implementation

Savalei & Kolenikov (2008) compared AMOS 7.0, EQS 6.1, LISREL 8.8, Mplus 4.2, and SAS PROC CALIS in SAS 9.1.

- Unconstrained estimation is default in all software packages except EQS.
- Release one d.f. for conditional inference: only in SAS PROC CALIS.
- Constrained Wald test: no software does that.
- Numerical and conceptual discrepancies between packages.
- Convergence diagnostic/missing s.e. problems with LISREL.
- Should the standard error on the constrained parameter be zero?
- Factor correlations > 1: parameterization may matter.

Phantom variables

Rindskopf (1983): device to impose inequality constraints.



- $\mathbb{V}[S] = \psi_S = b^2$
- When $\psi_S = 0$, $b = 0$, and $\partial l / \partial \psi_S = 2b \cdot \partial l / \partial b = 0$, so information matrix is not invertible!
- Phantom variables substitute one irregularity for another.

Recommended procedures

- Commit to restricting the estimates to be in their proper range; use conditional approach to overall fit test and mixture distribution for χ^2 difference tests.
- Leave the choice to the software and don't intervene (EQS: constrained estimation)
- Allow inadmissible solutions and test them for significance. Use Wald tests, confidence intervals or χ^2 difference tests that all have their "regular" distributions.

Questions?

Kolenikov and Bollen (2008)

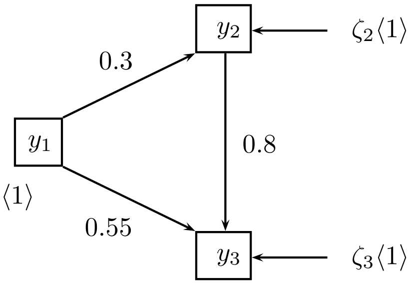
Kolenikov, S. and Bollen, K. A. (2008), 'Testing negative error variances: Is a Heywood case a symptom of misspecification?' Under review.

Research question: testing specification, as expressed via

$$H_{\emptyset} \cup H_{+} \text{ against } H_{-}$$

Population Heywood case

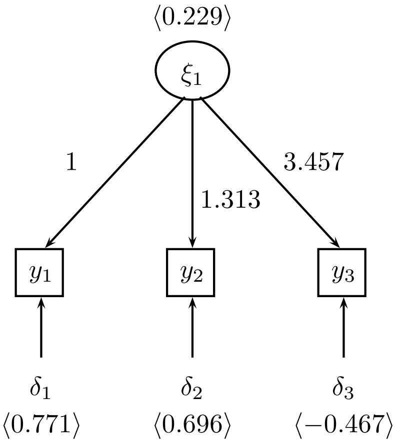
Data generating process



Jump to simulations

Population Heywood case

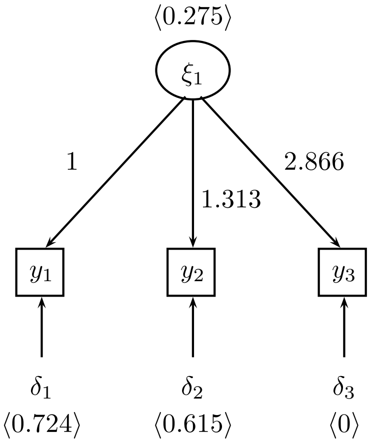
CFA model fit



Jump to simulations

Population Heywood case

CFA with restricted variance



Jump to simulations

Misspecified models

- Distributional misspecification: the distribution of the data is not normal, likelihood methods are not fully applicable (Satorra 1990, Satorra & Bentler 1994).
- Structural misspecification: the structure of the model, the number of latent variables, the relations between the variables in the model are not specified correctly (Yuan, Marshall & Bentler 2003).
- Heywood case: impossible value in the population; evidence of structural misspecification (or what??)
- If the model is misspecified, doesn't everything just fall apart??

Misspecified models

- Huber (1967)
  - Point estimates are consistent for the minimizer of the population fit function:

$$\arg\min F(S,\Sigma(\theta)) \to \arg\min F(\Sigma,\Sigma(\theta)) \geq 0,$$
$$F(\Omega,\Sigma(\theta)) = \ln|\Sigma(\theta)| + \Sigma(\theta)^{-1}\Omega - \ln|S| - \dim\Sigma(\theta) =$$
$$\overset{as.eq.}{=} (\omega - \sigma(\theta))' \, V \, (\omega - \sigma(\theta)),$$
$$\omega = \mathrm{vech}\,\Omega, \quad \sigma(\theta) = \mathrm{vech}\,\Sigma(\theta)$$

  - Uncertainty about these estimates is given by an asymptotic covariance matrix (sandwich estimator):

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, A^{-1}BA^{-T}),$$
$$A = \mathbb{E}\,\partial\psi(X,\theta_0), \quad B = \mathbb{E}\,\psi(X,\theta_0)\psi(X,\theta_0)^T,$$
$$\psi(X,\theta_0) = -\dot{\sigma}(\theta)' V (s - \sigma(\theta))$$

Misspecified models

Special cases:

- Eicker (1967) and White (1980), linear regression with heteroskedastic errors:

$$\hat{\beta} = (X'X)^{-1}X'y, \quad \hat{v}(\hat{\beta}) = (X'X)^{-1}(X'ee'X)(X'X)^{-1}$$

- Browne (1974): least squares estimator for SEM.
- White (1982): somewhat milder regularity conditions that are easier to check in practice; application to some common econometric models.
- Arminger & Schoenberg (1989): an econometrics paper in *Psychometrika*.
- Satorra & Bentler (1994): model based sandwich estimator with explicit expression for matrix B (the meat of the sandwich) as a function of the model parameters and the fourth order moments of data.
- Yuan & Hayashi (2006): comparison of empirical sandwich and the bootstrap standard errors for SEM.

Can standard errors be trusted?

Distributional specification	Structural specification: Correct	Structural specification: Incorrect
Correct	I, S-B, ES, EB, BSB	I, ES, EB
Incorrect	S-B, ES, EB, BSB	ES, EB

Analytic standard errors:

- **I** observed or expected information matrix
- **S-B** Satorra-Bentler standard errors
- **ES** empirical (Huber) sandwich

Resampling standard errors:

- **EB** empirical bootstrap
- **BSB** Bollen-Stine bootstrap with data rotation

Tests of $H_\emptyset \cap H_+$ vs. H_-

Likelihood ratio type tests

- Overall fit
- χ^2-difference tests (from $\theta_k = 0$)
- Signed root tests for simple χ^2 difference and scaled Satorra & Bentler (2001) difference

$$r(\theta_0) = \mathrm{sign}[\hat{\theta} - \theta_0]\sqrt{\Delta T}$$

Wald type tests, using

- Information matrix standard errors
- Satorra-Bentler standard errors
- Huber empirical sandwich standard errors

Simulation study

Saturated model with 3 variables and 6 parameters: $\chi^2_0 \equiv 0$, no way to test the model fit. . . unless you hit a Heywood case!

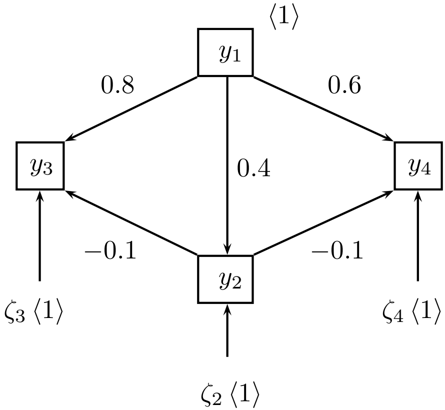
Jump to the Heywood case in population example

Main results:

- HUGE biases of the Heywood case in small samples.
- Information matrix standard errors are biased leading to undercoverage of CIs based on them when the data are non-normal.

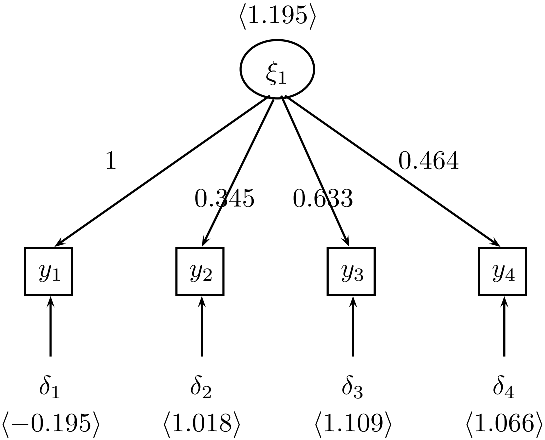
Simulation study

Larger model with 4 variables, 8 parameters and 2 d.f. for model fit test.
Data generating process



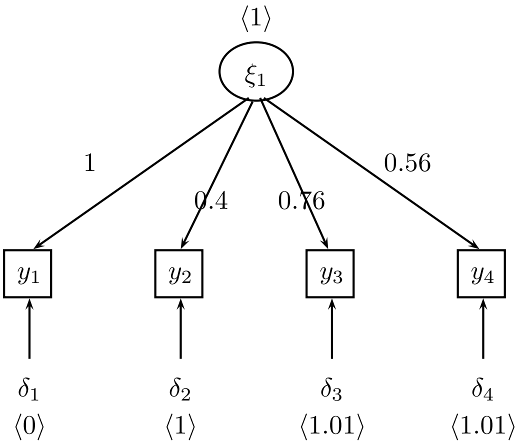
Simulation study

Larger model with 4 variables, 8 parameters and 2 d.f. for model fit test.
CFA model fit



Simulation study

Larger model with 4 variables, 8 parameters and 2 d.f. for model fit test.
CFA with restricted variance



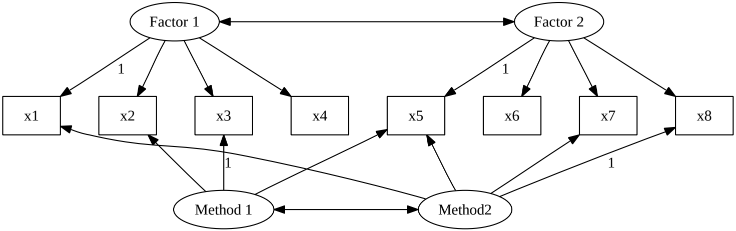
Simulation study

Larger model with 4 variables, 8 parameters and 2 d.f. for model fit test.
Main results:

- HUGE biases of the Heywood case in small samples.
- Information matrix standard errors are biased leading to undercoverage of CIs based on them when the data are non-normal.
- Satorra-Bentler standard errors are biased down leading to undercoverage *even when the data are normal!*
- Huber empirical sandwich standard errors are asymptotically accurate.
- Empirical bootstrap standard errors are even more accurate in smaller samples, and better centered to the true variability.
- Signed root of Satorra-Bentler scaled difference is the most accurate test in terms of the size, and has the greatest power among accurate tests!

Zero factor variances

Koleinkov and Savalei: work in progress.
Question: testing *factor* variances, as opposed to *error* variances.



Zero factor variances

Koleinkov and Savalei: work in progress.
Question: testing *factor* variances, as opposed to *error* variances.
Complications:

- If a factor variance is zero, what happens to its covariances with other factors and the loadings of observed variables?
- How many parameters are we actually testing?
- What are degrees of freedom?

Zero factor variances

Technically speaking, zero factor variances lead to underidentified models.

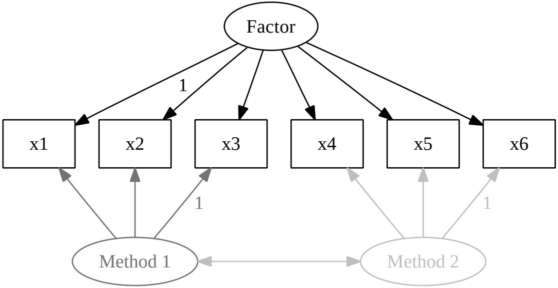
$\phi_k = 0 \Rightarrow \mathrm{Corr}(\xi_k, \xi_j)$ is underID, λ_{lk} is underID

Technical research into underidentified models (Davies 1977, Davies 1987, Andrews & Ploberger 1994, Hansen 1996):

- Asymptotic distribution still exists!
- It is characterized by max $\chi^2(\theta)$ over a range of θ
- Analytical work is very limited
- Simulation approach seem attractive

Simulation study

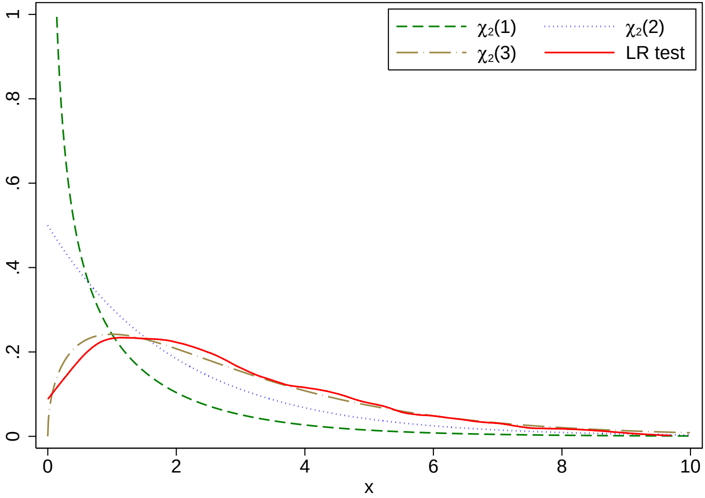
- Data generation: 1 factor multivariate normal model, 6 variables.
- Fitted models: extra one or two factors, loading on half of the variables each.



- Numeric stability issues: hundreds of maximization iterations; ridge-like likelihoods; absurd parameter estimates and standard errors.

Test statistic distribution

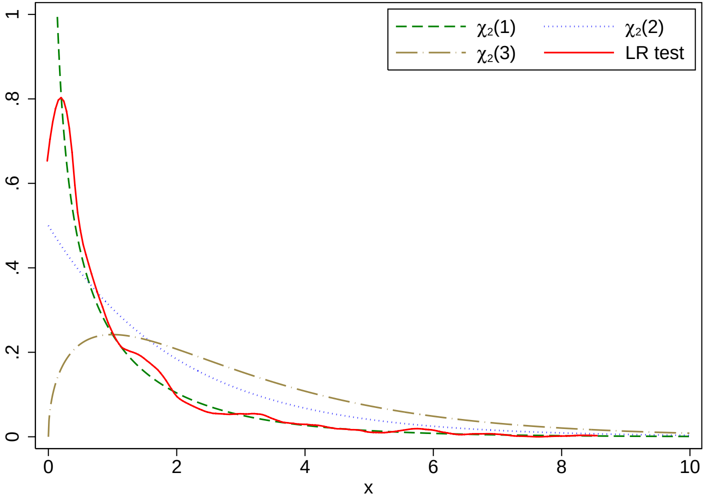
One extraneous factor, free loadings, $N = 2000$



Kolmogorov-Smirnov test vs. χ^2_3: p-value = 0.022.

Test statistic distribution

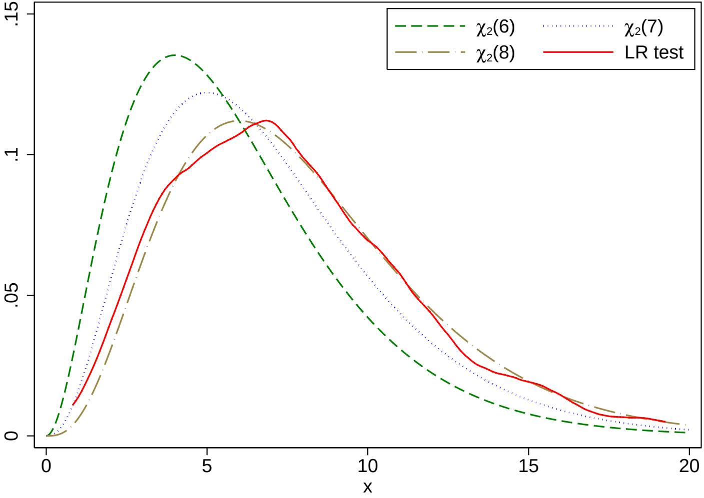
One extraneous factor, fixed loadings, $N = 2000$



Kolmogorov-Smirnov test vs. χ_1^2: p-value = 0.115.

Test statistic distribution

Two extraneous factors, free loadings, $N = 2000$



Kolmogorov-Smirnov test vs. χ^2_8: p-value = 0.047.

References I

Anderson, J. C. & Gerbing, D. (1984), 'The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis', *Psychometrika* **49**, 155–173.

Andrews, D. W. K. (1999), 'Estimation when a parameter is on a boundary', *Econometrica* **67**(6), 1341–1383.

Andrews, D. W. K. (2001), 'Testing when a parameter is on the boundary of the maintained hypothesis', *Econometrica* **69**(3), 683–734. doi:10.1111/1468-0262.00210.

Andrews, D. W. K. & Ploberger, W. (1994), 'Optimal tests when a nuisance parameter is present only under the alternative', *Econometrica* **62**(6), 1383–1414.

Arminger, G. & Schoenberg, R. J. (1989), 'Pseudo maximum likelihood estimation and a test for misspecification in mean and covariance structure models', *Psychometrika* **54**, 409–426.

Bollen, K. A. (1987), 'Outliers and improper solutions: A confirmatory factor analysis example', *Sociological Methods and Research* **15**, 375–384.

References II

Bollen, K. A. (1989), *Structural Equations with Latent Variables*, Wiley, New York.

Boomsma, A. (1983), *On the Robustness of LISREL (Maximum Likelihood Estimation) Against Small Sample Size and Nonnormality*, Sociometric Research Foundation, Amsterdam, the Netherlands.

Boomsma, A. & Hoogland, J. J. (2001), 'The robustness of LISREL modeling revisited', Structural Equation Modeling: Present and Future.

Browne, M. W. (1974), 'Generalized least squares estimators in the analysis of covariances structures', *South African Statistical Journal* **8**, 1–24.

Chernoff, H. (1954), 'On the distribution of the likelihood ratio', *The Annals of Mathematical Statistics* **25**(3), 573–578.

Davies, R. B. (1977), 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika* **64**(2), 247–254.

Davies, R. B. (1987), 'Hypothesis testing when a nuisance parameter is present only under the alternatives', *Biometrika* **74**(1), 33–43.

References III

Dijkstra, T. K. (1992), 'On statistical inference with parameter estimates on the boundary of the parameter space', *British Journal of Mathematical and Statistical Psychology* **45**, 289–309.

Dillon, W. R., Kumar, A. & Mulani, N. (1987), 'Offending estimates in covariance structure analysis: Comments on the causes and solutions to Heywood cases', *Psychological Bulletin* **101**, 126–135.

Eicker, F. (1967), Limit theorems for regressions with unequal and dependent errors, *in* 'Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability', Vol. 1, University of California Press, Berkeley, pp. 59–82.

Hansen, B. E. (1996), 'Inference when a nuisance parameter is not identified under the null hypothesis', *Econometrica* **64**(2), 413–430.

Heywood, H. B. (1931), 'On finite sequences of real numbers', *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **134**(824), 486–501.

References IV

Huber, P. (1967), The behavior of the maximum likelihood estimates under nonstandard conditions, *in* 'Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability', Vol. 1, University of California Press, Berkeley, pp. 221–233.

Jamshidian, M. & Bentler, P. M. (1994), 'Gramian matrices in covariance structure models', *Applied Psychological Measurement* **18**(1), 79–94.

Kolenikov, S. & Bollen, K. A. (2008), Testing negative error variances: Is a Heywood case a symptom of misspecification? under review in *Sociological Methods and Research*.

Rindskopf, D. (1983), 'Parameterizing inequality constraints on unique variances in linear structural models', *Psychometrika* **48**(1), 73–83.

Rindskopf, D. (1984), 'Structural equation models: Empirical identification, Heywood cases, and related problems', *Sociological Methods and Research* **13**, 109–119.

Sato, M. (1987), 'Pragmatic treatment of improper solutions in factor analysis', *Annals of the Institute of Statistics and Mathematics, part B* **39**, 443–455.

References V

Satorra, A. (1990), ‘Robustness issues in structural equation modeling: A review of recent developments’, *Quality and Quantity* **24**, 367–386.

Satorra, A. & Bentler, P. (2001), ‘A scaled difference chi-square test statistic for moment structure analysis’, *Psychometrika* **66**(4), 507–514.

Satorra, A. & Bentler, P. M. (1994), Corrections to test statistics and standard errors in covariance structure analysis, *in* A. von Eye & C. C. Clogg, eds, ‘Latent variables analysis’, Sage, Thousands Oaks, CA, pp. 399–419.

Savalei, V. & Kolenikov, S. (2008), ‘Constrained vs. unconstrained estimation in structural equation modeling’, *Psychological Methods* **13**, 150–170.

Shapiro, A. (1985), ‘Asymptotic distribution of test statistic in the analysis of moment structures under inequality constraints’, *Biometrika* **72**(1), 133–144.

References VI

Stoel, R. D., Garre, F. G., Dolan, C. & van den Wittenboer, G. (2006), 'On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints', *Psychological Methods* **11**(4).

Stram, D. O. & Lee, J. W. (1994), 'Variance components testing in the longitudinal mixed effects model', *Biometrics* **50**(4), 1171–1177.

Van Driel, O. P. (1978), 'On various causes of improper solutions in maximum likelihood factor analysis', *Psychometrika* **43**, 225–43.

White, H. (1980), 'A heteroskedasticity-consistent covariance-matrix estimator and a direct test for heteroskedasticity', *Econometrica* **48**(4), 817–838.

White, H. (1982), 'Maximum likelihood estimation of misspecified models', *Econometrica* **50**(1), 1–26.

Yuan, K.-H. & Hayashi, K. (2006), 'Standard errors in covariance structure models: Asymptotics versus bootstrap', *British Journal of Mathematical and Statistical Psychology* **59**, 397–417.

References VII

Yuan, K. H., Marshall, L. L. & Bentler, P. M. (2003), 'Assessing the effect of model misspecifications on parameter estimates in structural equation models', *Sociological Methodology* **33**(1), 241–265.