The Tetrad Criterion

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

The Tetrad Criterion

Introduction

- Spearman realized early in the 20th century that the existence of the general intelligence factor "g" would be supported by a single common factor model.
- Moreover, it became clear to him that this factor (as he envisioned explaining the correlations among mental abilities "in the partial correlation sense"), would fit the data if and only if the correlation matrix of the observations could be written in the form

$$\mathbf{R} = \mathbf{f}\mathbf{f}' + \mathbf{U}^2 \tag{1}$$

- The question immediately arose: How could one test this hypothesis empirically?

Computing the Implications of a Model

- Spearman eventually discovered something very interesting.
- If one expressed the data as a function of the model, one could compute the implications of the model for data.
- If the data agreed with those implications, the model withstood its empirical test.
- To understand how Spearman accomplished this, let's take a *very* simple example.

Computing the Implications of a Model

Model Equations and Model Constraints

- Before we begin discussing the relationship between model and data, we need to discuss what a model is.
- In structural equation modeling textbooks, you see statements to the effect that a structural equation model states that $\mathbf{\Sigma}$, the population covariance matrix, can be expressed as $\mathbf{M}(\boldsymbol{\theta})$, where $\mathbf{M}()$ is a *model matrix function* carrying the list of parameters in $\boldsymbol{\theta}$ into a covariance matrix (hopefully equal to $\mathbf{\Sigma}$).
- Unless the model is tautological, this implies a more general concept of a model, which is it is a statement about the parameters of the population distribution that restrict them to a subspace of the parameter space.

Computing the Implications of a Model

Model Equations and Model Constraints

- When a model is falsifiable, it imposes *constraints* on data. That is, not all data can fit the model, so the data must satisfy some constraint conditions in order to fit the model.
- Here is a simple example.
- Suppose you have 3 data values, a, b, and c.
- You have a model that says these data values can be explained in terms of 2 parameters, x and y, via the equations

$$a = x + y \tag{2}$$

$$b = x - y \tag{3}$$

$$c = 2x \tag{4}$$

Computing the Implications of a Model

Model Equations and Model Constraints

- We eliminate the parameters from the system to see what it "tells us about the data."
- If we add the first two equations together, we get

$$a + b = 2x \tag{5}$$

- But the third equation states that $c = 2x$, so for the data to fit this model, it *must* be the case that

$$a + b = c \tag{6}$$

Computing the Implications of a Model

Using Constraints to Calculate Parameters

- Suppose a data set actually does fit the model. What is the solution for x and y?
- Substituting $a+b$ for c in the original model equations, it is rather easy to see that

$$x = \frac{a+b}{2} = \frac{c}{2} \qquad (7)$$

$$y = \frac{a-b}{2} \qquad (8)$$

- Software like *Mathematica* does such operations in an eyeblink. For example

```
Eliminate[
{x + y == a, x - y == b, 2x == c},
{x, y}
]

{c == a + b}
```

Spearman's Tetrad Criterion

- Charles Spearman followed essentially the same process shown above in order to deduce the constraints that the correlation matrix must follow in order to fit the data.
- Let's start by expressing the model in terms of the covariance matrix $\boldsymbol{\Sigma}$.

$$\mathbf{f} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix}, \qquad \mathbf{U}^2 = \begin{pmatrix} \theta_5 & 0 & 0 & 0 \\ 0 & \theta_6 & 0 & 0 \\ 0 & 0 & \theta_7 & 0 \\ 0 & 0 & 0 & \theta_8 \end{pmatrix} \tag{9}$$

Spearman's Tetrad Criterion

We get (showing only the lower triangle)

$$\boldsymbol{\Sigma} = \mathbf{f}\mathbf{f}' + \mathbf{U}^2 = \begin{pmatrix} \theta_1^2 + \theta_5 & & & \\ \theta_2\theta_1 & \theta_2^2 + \theta_6 & & \\ \theta_3\theta_1 & \theta_3\theta_2 & \theta_3^2 + \theta_7 & \\ \theta_4\theta_1 & \theta_4\theta_2 & \theta_4\theta_3 & \theta_4^2 + \theta_8 \end{pmatrix} \tag{10}$$

We have a set of 10 equations, the first being $\sigma_{1,1} = \theta_1^2 + \theta_5$. We can eliminate the parameters by hand, but it is tedious. Mathematic does it, as follows

Spearman's Tetrad Criterion

```
Eliminate[
  {θ1^2 + θ5 == σ1,1,
   θ1 θ2 == σ2,1,
   θ2^2 + θ6 == σ2,2,
   θ1 θ3 == σ3,1,
   θ2 θ3 == σ3,2,
   θ3^2 + θ7 == σ3,3,
   θ1 θ4 == σ4,1,
   θ2 θ4 == σ4,2,
   θ3 θ4 == σ4,3,
   θ4^2 + θ8 == σ4,4},
  {θ1, θ2, θ3, θ4, θ5, θ6, θ7, θ8}]
```

Spearman's Tetrad Criterion

- *Mathematica* returns the following solution

$$\sigma_{3,1}\,\sigma_{4,2} == \sigma_{3,2}\,\sigma_{4,1}\ \&\&\ \sigma_{2,1}\,\sigma_{4,3} == \sigma_{3,2}\,\sigma_{4,1}$$

- The "&&" means "and" in *Mathematica* language.
- These "tetrad' equations are the conditions that the data must satisfy to fit a 1-factor model.
- Notice that the variances of the variables never appear in these 4 equations.
- What does that tell you?

Spearman's Tetrad Criterion

- Realizing that the data can be rescaled without affecting the tetrad difference equations, we can rewrite the formulas in this case to be

$$\rho_{3,1}\rho_{4,2} - \rho_{3,2}\rho_{4,1} = 0 \quad (11)$$
$$\rho_{2,1}\rho_{4,3} - \rho_{3,2}\rho_{4,1} = 0 \quad (12)$$

- In this form they became known as the tetrad difference equations.
- Researchers started examining the values of the tetrad differences, and comparing them to zero as a test of Spearman's single factor theory.

Spearman's Tetrad Criterion

- Once the tetrad formulas are added to the original model equations, they become solveable.
- *Mathematica* gives two solutions with opposite signs on the loadings.
- I'll reproduce just one.

Spearman's Tetrad Criterion

```
Solve[
 {θ1^2 + θ5 == σ1,1,
  θ1 θ2 == σ2,1,
  θ2^2 + θ6 == σ2,2,
  θ1 θ3 == σ3,1,
  θ2 θ3 == σ3,2,
  θ3^2 + θ7 == σ3,3,
  θ1 θ4 == σ4,1,
  θ2 θ4 == σ4,2,
  θ3 θ4 == σ4,3,
  θ4^2 + θ8 == σ4,4,
  σ3,1 σ4,2 == σ3,2 σ4,1,
  σ2,1 σ4,3 == σ3,2 σ4,1
 },
 {θ1, θ2, θ3, θ4, θ5, θ6, θ7, θ8}]
```

Spearman's Tetrad Criterion

$$\left\{\theta_5 \to \sigma_{1,1} - \frac{\sigma_{2,1}^2\,\sigma_{3,1}^2\,\sigma_{4,2}\,\sigma_{4,3}}{\sigma_{3,2}^3\,\sigma_{4,1}^2},\right.$$

$$\theta_6 \to \sigma_{2,2} - \frac{\sigma_{2,1}^2\,\sigma_{4,2}\,\sigma_{4,3}}{\sigma_{3,2}\,\sigma_{4,1}^2},$$

$$\theta_7 \to \sigma_{3,3} - \frac{\sigma_{3,1}^2\,\sigma_{4,2}\,\sigma_{4,3}}{\sigma_{3,2}\,\sigma_{4,1}^2},$$

$$\theta_8 \to -\frac{\sigma_{4,2}\,\sigma_{4,3}}{\sigma_{3,2}} + \sigma_{4,4},$$

$$\theta_1 \to \frac{\sigma_{2,1}\,\sigma_{3,1}\,\sqrt{\sigma_{4,2}}\,\sqrt{\sigma_{4,3}}}{\sigma_{3,2}^{3/2}\,\sigma_{4,1}},$$

$$\theta_2 \to \frac{\sigma_{2,1}\,\sqrt{\sigma_{4,2}}\,\sqrt{\sigma_{4,3}}}{\sqrt{\sigma_{3,2}}\,\sigma_{4,1}},$$

$$\theta_3 \to \frac{\sigma_{3,1}\,\sqrt{\sigma_{4,2}}\,\sqrt{\sigma_{4,3}}}{\sqrt{\sigma_{3,2}}\,\sigma_{4,1}},$$

$$\left.\theta_4 \to \frac{\sqrt{\sigma_{4,2}}\,\sqrt{\sigma_{4,3}}}{\sqrt{\sigma_{3,2}}}\right\}\}$$

Spearman's Tetrad Criterion

- As I mentioned, Spearman's ingenious approach met with some immediate objections.
- One problem was the issue of sampling error, although the multivariate delta method makes it straightforward to derive an estimated standard error of a tetrad difference based on asymptotics.
- Another point was that, even if the tetrad differences are small, they are very unlikely ever to be precisely zero in the sample. This raises the practical question of which estimates are optimal. Modern maximum likelihood methods now are in wide use, and Spearman's approach has been almost completely forgotten.
- But in passing, we noticed something that ultimately proved valuable in a modern context.
- Any ideas what that might be?