Chapter 9: Confirmatory Factor Analysis

Prerequisites: Chapter 5, Sections 3.9, 3.10, 4.3

9.1 *The Confirmatory Factor Analysis Model*

The difference between the models discussed in this section, and the regression model introduced in Chapter 5 is in the nature of the independent variables, and the fact that we have multiple dependent variables. The independent variables are unobserved constructs, also known as *factors*, dimensions or *latent variables*. At this point the student might ask, how scientific is it to speak of unobserved variables in a model? We will soon see that if the model of unobserved independent variables is correct, it makes a strong prediction about the structure of the covariances among the observed dependent variables. For this reason, these models are a special case of models known as *covariance structure models*.

Given that we are dealing with unobserved variables, it will be useful to shift our notation somewhat. In regression, we look at a particular variable as a column vector that displays the individual observations which comprise the rows. In factor analysis, the individual observations cannot be fully observed since the right hand side variables, the factors, are not observed. Instead, we will propagate our model using a typical observation, call it observation i, but leaving off the subscript i. What's more, instead of arranging our matrices such that the each column is a different variable and each row is a different observation, we will be looking at the transpose.

Of course, this is in contrast to the notation employed in Chapters 5 through 8. In that later chapter, we study the model

$$\mathbf{Y} = \mathbf{XB} + \boldsymbol{\varepsilon}$$

where the columns of **Y** (and the parameter vector **B** as well as the error matrix $\boldsymbol{\varepsilon}$) represent the p different dependent variables. If we were to take the transpose of both sides of that model we would have

$$\mathbf{Y}' = \mathbf{B}'\mathbf{X}' + \boldsymbol{\varepsilon}'.$$

You will note that, since the product of a transpose is the transpose of the product in reverse order [Equation (1.34)], **B** and **X** are now reversed. Also, the data matrices $\mathbf{Y}'$ and $\mathbf{X}'$ now have a row for each variable, instead of a column as before. Next, as described above, rather than look at every subject, we look at a typical observation, for example, number i:

$$\mathbf{y}_{i\cdot} = \mathbf{B}'\mathbf{x}_{i\cdot} + \boldsymbol{\varepsilon}_{i\cdot}.$$

The dot, which is a subscript reduction operator, is mentioned in Section 1.1. One final change is convenient. If we totally drop the subscripts from $\mathbf{y}_{i\cdot}$, $\mathbf{x}_{i\cdot}$ and $\boldsymbol{\varepsilon}_{i\cdot}$, we would just have

$$\mathbf{y} = \mathbf{B}'\mathbf{x} + \boldsymbol{\varepsilon}.$$

This is how we will describe the model in this chapter. We will call the regression weights λ's instead of β's and the independent variables will be η's instead of x's.

We start out with a scalar representation of the situation:

$$
\begin{aligned}
y_1 &= \lambda_{11}\eta_1 + \lambda_{12}\eta_2 + \cdots + \lambda_{1m}\eta_m + \varepsilon_1 \\
y_2 &= \lambda_{21}\eta_1 + \lambda_{22}\eta_2 + \cdots + \lambda_{2m}\eta_m + \varepsilon_2 \\
\cdots &= \quad \cdots \\
y_p &= \lambda_{p1}\eta_1 + \lambda_{p2}\eta_2 + \cdots + \lambda_{pm}\eta_m + \varepsilon_p \ .
\end{aligned}
\tag{9.1}
$$

The left hand side shows p different variables. Perhaps y_1 through y_3 represent three measures of consumer "greenness", that is, a tendency to buy environmental friendly products. Perhaps y_4 through y_6 represent three different measures of innovativeness. In any case, the point is that the y's are p *manifest* or observed variables. As has been mentioned, we are representing the data from a typical subject, the i-th, but the subscript i is left off according to the traditions in this area. On the right hand side, you have regression coefficients, the λ_{ij}, which are basically β weights. In the context of factor analysis, regression weights are called factor *loadings*. The reason that they have two subscripts is that you need one subscript to keep track of the dependent variable, or the equation, and another subscript to keep track of the independent variable. And speaking of which, these are the η values of which there are m. The η's are the *common factors* which explain much of the behavior of the y's, at least the part of their behavior that they have in common – the covariances. Finally, we have the ε_i which are called *unique factors*. This is not exactly the same thing as the error in a regression model. In regression, the error is an *error-in-equations*, also called *specification error.* That is to say, unless a regression model has an R^2 of 1, the model is missing some explanatory independent variables or is otherwise *mispecified*. In factor analysis, the ε's are *errors-in-variables*, or *measurement error*. The three variables we devised to measure "greenness", for example, might not do so perfectly. We generally assume that the part that the three variables have in common, as quantified by their covariances, must be due to the fact that all three are at least partially measuring what they are supposed to be measuring. But each one of the three has some variance that is unique to it. That is what the ε_i account for.

We can write the model in matrix terms,

$$
\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_p \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pm} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \cdots \\ \eta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_p \end{bmatrix}.
$$

$$
\mathbf{y} = \mathbf{\Lambda\eta} + \boldsymbol{\varepsilon} \ . \tag{9.2}
$$

By all rights, in addition to the $\mathbf{y}$ vector, the $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ vectors should have a subscript i since they are random variables, sampled from the population for which this model holds. On the other hand, $\mathbf{\Lambda}$ is a constant matrix, holding parameters that describe this population.

So how does this model with unobserved variables make contact with reality? In order to show how it does so, we need to start with some assumptions and some definitions. We will assume that $E(\mathbf{y}) = \mathbf{0}$, a p by 1 null vector. This does not reduce the generality of the model at all, since covariances are not affected by the addition or subtraction of a constant [see Theorem (4.8)]. In order to estimate the model, we will make the assumptions that

$$
\boldsymbol{\eta} \sim N(\mathbf{0}, \mathbf{\Psi}),
$$

$$
\boldsymbol{\varepsilon} \sim \mathbf{N(0, \Theta)}
$$

and that

$$\text{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{\eta}) = \mathbf{0}.$$

Like the $\mathbf{y}$ vector, $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ are mean-centered. We will also see quite a bit of the coviarance matrices for $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$, with $V(\boldsymbol{\eta}) = \boldsymbol{\Psi}$ and $V(\boldsymbol{\varepsilon}) = \boldsymbol{\Theta}$. At this point, we are ready to see what the covariance matrix of the y's should look like. We have by the definition of variance in Equation (4.7)

$$\begin{aligned} V(\mathbf{y}) \equiv \boldsymbol{\Sigma} &= E(\mathbf{y}\mathbf{y}') \\ &= E\left[(\boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon})(\boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon})'\right] \\ &= \boldsymbol{\Lambda}\, E(\boldsymbol{\eta}\boldsymbol{\eta}')\, \boldsymbol{\Lambda}' + \boldsymbol{\Lambda}\, E(\boldsymbol{\eta}\boldsymbol{\varepsilon}') + E(\boldsymbol{\varepsilon}\boldsymbol{\eta}')\, \boldsymbol{\Lambda}' + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'), \end{aligned}$$

but of the four components from left to right, pieces two and three vanish since $\text{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{\eta}) = \mathbf{0}$. We have made use of Equation (4.5) and (4.6). We can rewrite $E(\boldsymbol{\eta}\boldsymbol{\eta}') = \boldsymbol{\Psi}$, which was defined above as the covariance matrix of the η's when we were talking about assumptions. In piece four we have $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Theta}$ which was also defined above as the variance of the unique factors. Putting all of these conclusions together, we end up with the fact that the variance of $\mathbf{y}$ is

$$V(\mathbf{y}) = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}\,. \tag{9.3}$$

9.2 *A Confirmatory Factor Analysis Example*

Now is the section of the chapter where we look at an example confirmatory factor analysis that is just complicated enough to be a valid example, but is simple enough to be, well; a silly example. Lets say we have devised three questionnaire items which measure the consumers' attitude towards Beer B, and three other items that measure attitudes towards Beer C. Our six item survey then contains the variables listed in the table:

Variables	**Description**
y_1	Measurement 1 of B
y_2	Measurement 2 of B
y_3	Measurement 3 of B
y_4	Measurement 1 of C
y_5	Measurement 2 of C
y_6	Measurement 3 of C

To finish describing the model, we will hypothesize that there are two factors, B (η_1) and C (η_2). Our model would then look like

$$\mathbf{y} = \mathbf{\Lambda\eta} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix}$$

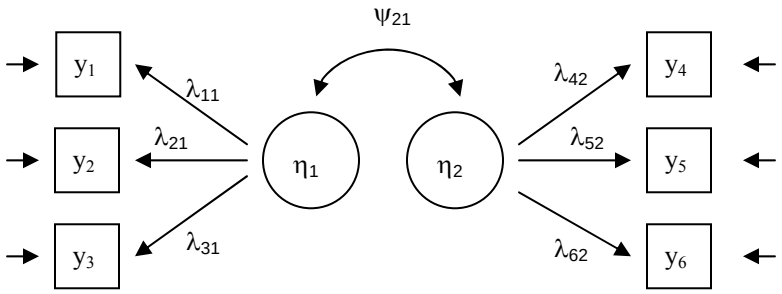
Again, remember that the $\mathbf{y}$, $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ vectors are random variables, but $\mathbf{\Lambda}$ is a parameter matrix and the unknowns in it must be estimated from the sample. To fully estimate the model, we also have two other parameter matrices,

$$\mathbf{\Psi} = \begin{bmatrix} \psi_{11} & - \\ \psi_{21} & \psi_{22} \end{bmatrix} \text{ and}$$

$$\mathbf{\Theta} = \begin{bmatrix} \theta_{11} & 0 & \cdots & 0 \\ 0 & \theta_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \theta_{pp} \end{bmatrix}.$$

Note that the $\mathbf{\Psi}$ matrix is symmetric, being a covariance matrix and so we do not need to enumerate the upper triangular part of it. And by the definition of what we mean by a unique factor, the ε_i are independent which means that the variance matrix of the ε_i, $\mathbf{\Theta}$, is diagonal. As a general rule in covariance structure models, we need to specify variances and covariances of right hand side random variables, and we need to specify regression weights between right hand and left hand side variables.

Below you can see what we call the *Path Diagram* for this model:



A path diagram is a very common way of representing a covariance structure model, and there are a set of conventions that go along with this type of figure. Single-headed arrows represent directional causal pathways, and two-headed arrows are used to represent covariation. Unique

factors, and other sorts of error terms, are usually indicated by single-headed arrows without labels. Circles are used to convey the presence of latent variables, and boxes convey observed variables.

9.3 *Setting a Metric for Latent Variables*

The model as it has been presented so far cannot be uniquely identified. To illustrate this, lets pretend we have a single variable and a single factor. In that case everything boils down to scalars, and the model is $y = \lambda\eta + \varepsilon$ and from Equation (9.3), $V(y) = \lambda^2\psi + \theta$. Now define $\eta^* = a\cdot\eta$ so that $V(\eta^*) = a^2\psi = \psi^*$. Also, define $\lambda^* = \lambda/a$. In that case,

$$y = \lambda^*\eta^* + \varepsilon = \frac{\lambda}{a}\cdot a\eta + \varepsilon \quad \text{and also} \tag{9.4}$$

$$V(y) = \lambda^{*2}\,\psi^* + \theta = \frac{\lambda^2}{a^2}\cdot a^2\psi + \theta\,. \tag{9.5}$$

What this means is that if I have a model with parameters λ^* and ψ^*, and you have a model with parameters λ and ψ, both models would fit equally well and there would be no logical way to decide which was better. In fact, they would be completely equivalent. The source of this ambiguity lies in the fact that η is unobserved, and it is at most an interval scale. To further identify the model we must set intervals for it, a process called setting its metric. We can do this in one of two ways. We can fix one loading per factor to a constant, such as 1.0, or we can fix the variance of each factor to 1.0. Returning to our two factor example, the first method would yield

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}, \quad \text{and} \quad \mathbf{\Psi} = \begin{bmatrix} \psi_{11} & - \\ \psi_{21} & \psi_{22} \end{bmatrix}$$

while the second approach would give

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}, \quad \text{and} \quad \mathbf{\Psi} = \begin{bmatrix} 1 & - \\ \psi_{21} & 1 \end{bmatrix}.$$

These two methods are equivalent, yielding the same Chi Square values, but the first method is slightly more general, being applicable in certain situations where the second method cannot be used. The first method ties the metric of each factor to the first variable that measures it. The second method turns the factors into z-scores, and the factor covariance matrix $\mathbf{\Psi}$ can then be interpreted as a correlation matrix. For both methods, the $\mathbf{\Theta}$ matrix has p free parameters.

9.4 *Degrees of Freedom for a Confirmatory Factor Analysis Model*

Factor analysis does not look directly at raw data. The input data for this technique are the elements of the sample covariance matrix **S**, which is a p by p symmetric matrix. Therefore **S** contains

$$\frac{p(p+1)}{2} \tag{9.6}$$

"data points", those being the p variances and the p(p-1)/2 unique covariances. For our 6 variable example, this would total 21. In our model, assuming we use the first method to fix the metric of the two factors, we have

4 λ's
3 ψ's
6 θ's
---
13 parameters

The degrees of freedom for the model are equal to the number of data points minus the number of unique free parameters that are estimated from those data. In our case, we have 21 – 13 = 8 degrees of freedom. We will be able to reject the model (or not as the case may be) using a χ^2 test with 8 degrees of freedom. In terms of hypotheses, we will be testing

$$H_0: \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \tag{9.7}$$

against the general alternative

$$H_A: \boldsymbol{\Sigma} = \mathbf{S} . \tag{9.8}$$

In some ways this pair of hypotheses is very similar to hypotheses that we saw in Chapter 6 with regression. However, here we have a different sort of emotional attachment to the hypotheses. In regression, which encompasses everything from the basic *t*-test through more complex possibilities, we are generally motivated to "hope for" H_A and hope against H_0. Here, our model is H_0, so in an emotional sense, the roles of the Type I and II errors are reversed. The truth is that the current situation is actually more natural, if we can use that word. In regression, the hypothesis we are testing is a sort of "straw man" that no one believes in anyway, and that we set up just to knock down. We will talk more about the "emotional reversal" of H_0 and H_A later when we discuss goodness of fit measures (that is, measures other than the traditional χ^2). But first, it is time to understand how we estimate the parameters of the model and come up with a χ^2 value to test it. That is the topic of the next two sections. We will be using an estimation philosophy known as Maximum Likelihood. In order to explore this topic, we will be returning to the much simpler regression model. Then we will venture forth and look at estimation for confirmatory factor analysis models.

9.5 *Maximum Likelihood Estimators for Factor Analysis*

Maximum likelihood is discussed in general in Section 3.10 and within the context of the regression model in Section 5.4. ML for factor analysis begins with the probability of observation i under the confirmatory factor analysis model. Here we have the multivariate normal distribution

[see Equation (4.17)] to deal with since we have p variables, not just one as we did with regression. We have

$$\Pr(\mathbf{y}_{i\cdot}) = \frac{1}{(2\pi)^{p/2}\,|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}\mathbf{y}_i'\boldsymbol{\Sigma}^{-1}\mathbf{y}_i\right] \tag{9.9}$$

for the p variables on observation i. For the whole sample we have

$$\ell_0 = \prod_i^n \Pr(\mathbf{y}_i) = \frac{1}{(2\pi)^{np/2}\,|\boldsymbol{\Sigma}|^{n/2}} \exp\left[-\frac{1}{2}\sum_i^n \mathbf{y}_i'\boldsymbol{\Sigma}^{-1}\mathbf{y}_i\right]. \tag{9.10}$$

The summation in the exponent of the above equation makes sense if you keep in mind that $e^a \cdot e^b = e^{a+b}$. Now, to get ready for the next equation note that from Equation (1.27) and Section 1.7

$$\sum_i^n \mathbf{y}_i'\boldsymbol{\Sigma}^{-1}\mathbf{y}_i' = \mathrm{tr}[n\mathbf{S}\boldsymbol{\Sigma}^{-1}] \tag{9.11}$$

because

$$\sum_i^n \mathbf{y}_i'\boldsymbol{\Sigma}^{-1}\mathbf{y}_i = \mathrm{Tr}\left[\sum \mathbf{y}_i'\boldsymbol{\Sigma}^{-1}\mathbf{y}_i\right] = \mathrm{Tr}\sum_i^n \mathbf{y}_i\mathbf{y}_i'\boldsymbol{\Sigma}^{-1} = \mathrm{Tr}[n\mathbf{S}\boldsymbol{\Sigma}^{-1}].$$

This is so since a scalar is equal to its trace, and the trace of a product is invariant to the sequence of that product assuming conformability. We now take the log of the likelihood in Equation (9.10) but substitute the identity from Equation (9.11) to end up with

$$\begin{aligned} \ln \boldsymbol{\ell}_0 = L_0 &= -\frac{1}{2}np\,\ln(2\pi) - \frac{1}{2}n\ln|\boldsymbol{\Sigma}| - \frac{1}{2}n\,\mathrm{tr}\,(\mathbf{S}\boldsymbol{\Sigma}^{-1}) \\ &= \text{constant} - \frac{1}{2}n\left[\ln|\boldsymbol{\Sigma}| + \mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1})\right]. \end{aligned} \tag{9.12}$$

The term "constant" above represents $-\frac{1}{2}np\,\ln(2\pi)$ which doesn't impact the optimal solution one way or the other since it does not depend on the parameters and so will not figure into the derivative. Now suppose I look at the likelihood under H_A: $\boldsymbol{\Sigma} = \mathbf{S}$. We will call that log likelihood L_A and we find that

$$L_A = \text{constant} - \frac{1}{2}n\left[\ln|\mathbf{S}| + p\right]. \tag{9.13}$$

Now we have two log likelihoods, one; L_0 which reflects the confirmatory factor analysis model, and another that gives us the log likelihood under the general alternative that $\boldsymbol{\Sigma}$ exhibits no particular structure, which is to say it is arbitrary. In other words, it is what it is.

It turns out that under very general conditions,

$$-2\ \ln\left[\frac{\ell_0}{\ell_A}\right] = -2\left[L_0 - L_A\right] \sim \chi^2(m), \tag{9.14}$$

where m represents the difference in the number of parameters estimated under the two models; the null (0) and the alternative (A). As we have already described, the alternative model estimates $\frac{p(p+1)}{2}$ parameters while the number of parameters in the null model depends on the specific theory as expressed in the matrices $\mathbf{\Lambda}$, $\mathbf{\Psi}$ and $\mathbf{\Theta}$. Plugging Equations (9.12) and (9.13) into Equation (9.14), the χ^2 value is then

$$\hat{\chi}^2 = n\left[\ln|\mathbf{\Sigma}| - \ln|\mathbf{S}| + tr(\mathbf{S}\mathbf{\Sigma}^{-1}) - p\right]. \tag{9.15}$$

As can be seen, as $\mathbf{\Sigma} \to \mathbf{S}, \hat{\chi}^2 \to 0$. Thus the closer the match between $\mathbf{\Sigma}$ and $\mathbf{S}$, the smaller the value of χ^2. But it is also true that as $n \to \infty, \hat{\chi}^2 \to \infty$, and conversely, as $n \to 0, \hat{\chi}^2 \to 0$. This means that all things being equal, it becomes easier to reject H_0 the larger the sample size, and it becomes harder to reject H_0 the smaller the sample size. This is how all efficient statistics function, but since we have an emotional attachment to H_0 instead of H_A, this would seem to have certain consequences both for individual researchers, and for the development of marketing as a whole.

It is necessary that we pick values for the unknowns in the matrices $\mathbf{\Lambda}$, $\mathbf{\Psi}$ and $\mathbf{\Theta}$ at the minimum value of Equation (9.15). Equation (9.15) is obviously nonlinear in the unknowns so this will entail nonlinear optimization as discussed in Section 3.9. For now we note that any computer algorithm that finds the minimum of Equation (9.15) will utilize the derivatives of that function to determine "which way is down". Any such algorithm, however, requires rational starting values to avoid ending up in a local, rather than the global, minimum of the function. As such, you should do the best job that you can by manually inserting starting values into whatever program you use to estimate the confirmatory factor model. Certainly, under any circumstances, you should be able to get the sign right for any loadings in the matrix $\mathbf{\Lambda}$. Diagonal elements of $\mathbf{\Theta}$ could be seeded with small positive values. Diagonal elements of $\mathbf{\Psi}$ are likely to resemble the variances of the measures, while off-diagonal elements could be smaller than the diagonal, and of appropriate sign. Of course, it is also important that any fixed elements in the matrices $\mathbf{\Lambda}$, $\mathbf{\Psi}$ and $\mathbf{\Theta}$ have appropriate starting values, as these will also end up as the final values!

9.6 *Special Case: The One Factor Model*

Consider a confirmatory factor model with one factor:

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_p \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdots \\ \lambda_p \end{bmatrix} \eta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_p \end{bmatrix} .$$

If we fix $V(\eta) = \psi_{11} = 1$, the expression for the covariance matrix is simply

$$\mathbf{\Sigma} = \mathbf{\lambda}\mathbf{\lambda}' + \mathbf{\Theta} \tag{9.16}$$

and our measures y_1 , y_2 , …, y_p are called *congeneric tests*. In this context the single η is called a *true score*. As you might guess, this terminology comes from the field of educational and psychological measurement. If we further specialize the model so that all lambdas are equal, i. e.

$$\lambda_1 = \lambda_2 = \cdots = \lambda_p = \lambda ,$$

we have the model of *τ-equivalent tests*. Congeneric tests have p λ's and p θ's, but τ-equivalent tests have only one λ and p θ's. Finally, the model of *parallel tests* includes the additional restriction that

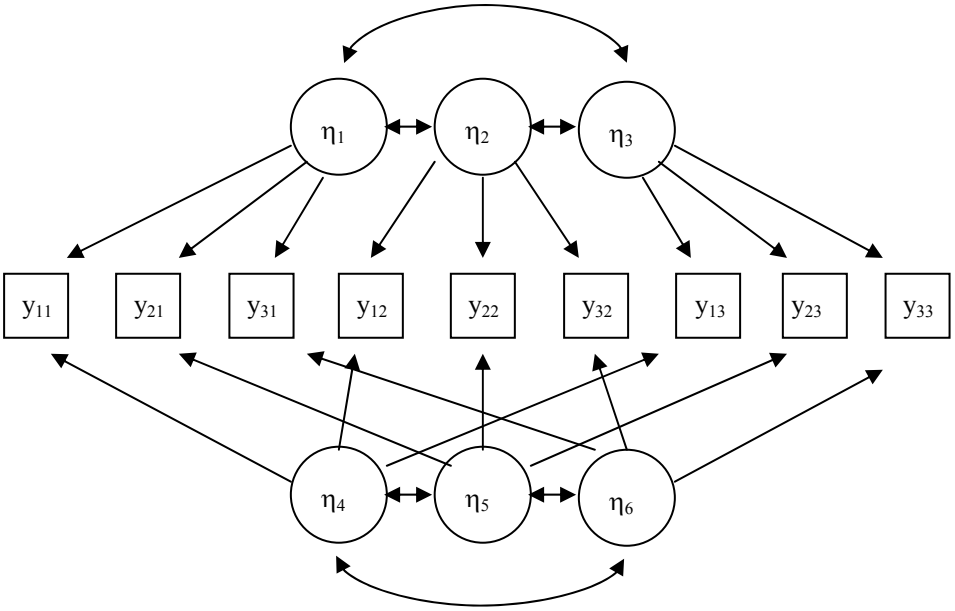
$$\theta_{11} = \theta_{22} = \cdots = \theta_{pp} = \theta .$$

Congeneric tests involve 2p free parameters to be estimated from the sample covariances, τ-equivalent tests have p + 1 parameters, and parallel tests have only 2 unknown parameters. Thus the model of parallel tests makes a very strong prediction about the structure of the covariance matrix using only 2 parameters. Having only 2 parameters means that the model has a larger number of degrees of freedom than τ-equivalence and especially congeneric tests. The degrees of freedom of the model represent restrictions that must be met in the covariance matrix. As such, parallel tests places many more restrictions on the covariance matrix which is shown below:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \lambda \\ \lambda \\ \cdots \\ \lambda \end{bmatrix} \begin{bmatrix} \lambda & \lambda & \cdots & \lambda \end{bmatrix} + \begin{bmatrix} \theta & 0 & \cdots & 0 \\ 0 & \theta & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \theta \end{bmatrix} .$$

9.7 *The Multi-Trait Multi-Method Model*

We sometimes have an opportunity to measure a set of traits using a common set of methods. For example we might measure the consumer's attitude towards a set of products repeating the same items to measure each product. With three traits (products) and three methods (items) we would have a path diagram as below. Note that to simplify an already complicated diagram, the unique factors were left off, as were the labels on the arrows.



and then the model would appear as

$$\begin{bmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{12} \\ y_{22} \\ y_{32} \\ y_{13} \\ y_{23} \\ y_{33} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 & 0 & \lambda_{14} & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 & \lambda_{25} & 0 \\ \lambda_{31} & 0 & 0 & 0 & 0 & \lambda_{36} \\ 0 & \lambda_{42} & 0 & \lambda_{44} & 0 & 0 \\ 0 & \lambda_{52} & 0 & 0 & \lambda_{55} & 0 \\ 0 & \lambda_{62} & 0 & 0 & 0 & \lambda_{66} \\ 0 & 0 & \lambda_{73} & \lambda_{74} & 0 & 0 \\ 0 & 0 & \lambda_{83} & 0 & \lambda_{85} & 0 \\ 0 & 0 & \lambda_{93} & 0 & 0 & \lambda_{96} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \eta_6 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{31} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{32} \\ \varepsilon_{13} \\ \varepsilon_{23} \\ \varepsilon_{33} \end{bmatrix},$$

where η_1, η_2 and η_3 are trait factors and η_4, η_5 and η_6 are method factors. To finish specifying the model, we note that $V(\boldsymbol{\varepsilon}) = \text{Diag}(\theta_{11}\ \theta_{22} \cdots \theta_{99})$, meaning that the nine unique elements of $\boldsymbol{\Theta}$ are arrayed on it's diagonal, and that

$$V(\boldsymbol{\eta}) = \boldsymbol{\Psi} = \begin{bmatrix} 1 & & & & & \\ \alpha_{21} & 1 & & & — & \\ \alpha_{31} & \alpha_{32} & 1 & & & \\ 0 & 0 & 0 & 1 & & \\ 0 & 0 & 0 & \beta_{21} & 1 & \\ 0 & 0 & 0 & \beta_{31} & \beta_{32} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\alpha} & - \\ \mathbf{0} & \boldsymbol{\beta} \end{bmatrix}.$$

The three by three section of zeroes in $\boldsymbol{\Psi}$ is null because trait and method factors are assumed independent, an assumption that we would be testing when we look at the χ^2 for the model. Note that we have called the correlations among the trait factor α's and the correlations among the method factors β's. This does not change anything of course. This is just a confirmatory factor analysis model in which certain values in the $\boldsymbol{\Psi}$ matrix are playing slightly different roles from other values.

9.8 Goodness of Fit, Root Mean Square Error, and Other Output from the Model

With a large enough sample size, one can statistically reject even fairly good models. Conversely, with a small sample size it is possible to fail-to-reject models that are patently incorrect. Given that state of affairs, Bentler and Bonet (1980) proposed that in addition to comparing H_0 vs H_A, that we introduce a truly null hypothesis. I will call this latest hypothesis H_S for "straw man" hypothesis. Specifically we have

$$H_A: \boldsymbol{\Sigma} = \mathbf{S}$$

$$H_0: \boldsymbol{\Sigma} = \boldsymbol{\Lambda\Psi\Lambda}' + \boldsymbol{\Theta}$$

$$H_S: \boldsymbol{\Sigma} = \boldsymbol{\Psi} \text{ (with } \boldsymbol{\Psi} \text{ diagonal)}$$

For the straw man hypothesis, H_S, we have further restricted H_0 such that $\boldsymbol{\Lambda} = \mathbf{I}$, $\boldsymbol{\Theta} = \mathbf{0}$, and $\boldsymbol{\Psi}$ is diagonal. We have three hypotheses. For hypothesis j, with degrees of freedom df_j, we define

$$Q_j = \frac{\hat{\chi}^2}{df_j}$$

and then we define

$$\rho_{s0} = \frac{Q_S - Q_0}{Q_S - 1} \tag{9.17}$$

as one possible measure and

$$\Delta_{s0} = \frac{\hat{\chi}_s^2 - \hat{\chi}_0^2}{\hat{\chi}_s^2} \tag{9.18}$$

as another measure of *goodness of fit*. This latter index, Δ_{s0}, where the subscripts s and 0 highlight the fact that we are comparing hypotheses s and 0, represents the percent improvement in $\hat{\chi}^2$ from hypothesis s to hypothesis 0. The quantity 1 - Δ_{s0} gives us the remaining improvement that would be possible for H_A.

Joreskög has proposed an index simply termed *GFI* that consists of

$$\text{GFI} = 1 - \frac{\text{tr}\left[\boldsymbol{\Sigma}^{-1}\mathbf{S} - \mathbf{I}\right]^2}{\text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S})}$$

and an adjusted version,

$$\text{AGFI} = 1 - \frac{p(p+1)}{2 \cdot df_0}(1 - \text{GFI})$$

We should also mention that there exists a traditional measure of fit for any sort of model, the root mean square error, or

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{i}(s_{ij} - \sigma_{ij})^2}{p(p+1)/2}} \quad .$$

Note that the double summation operators in the numerator run through each of the unique elements in the covariance matrix. The RMSE gives you the average error across the elements of $\boldsymbol{\Sigma}$ as compared with $\mathbf{S}$.

We can also look at lack of fit for any individual fixed parameter. Of course, any free parameter estimated from the sample covariance matrix $\mathbf{S}$ does not contribute to lack of fit. It is the fixed parameters, generally the 0's in $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ that are being tested in H_0 and it is these elements that cause a model to not fit. Given that we are picking free parameters in such a way that the derivative of Chi Square with respect to those parameters is 0, or assuming all of our free parameters are in the vector $\boldsymbol{\alpha}'$, we have solved for the free parameters when

$$\frac{\partial \hat{\chi}^2}{\partial \boldsymbol{\alpha}'} = \mathbf{0}'$$

because when the derivatives are zero, Chi Square is minimized. But this suggests a way to judge the fixed parameters. For any fixed parameter, say π, in general

$$\frac{\partial \hat{\chi}^2}{\partial \pi} \neq 0 \, .$$

These first derivatives provide a clue as to which parameter can be changed from fixed to free for the maximal benefit to $\hat{\chi}^2$. All that remains is that we scale the first derivative with the second derivative and we have what is called a *modification index*, or MI:

$$\text{MI} = \frac{\frac{n}{2}\left(\frac{\partial \hat{\chi}^2}{\partial \pi}\right)^2}{\frac{\partial (\hat{\chi}^2)^2}{\partial \pi \, \partial \pi}} \, .$$

General information on the second order derivative is given in Section 3.3 and its role in ML is discussed in Section 3.10.

References

Bentler, Peter and Douglas Bonett (1980) Significance Tests and Goodness-of-Fit in the Analysis of Covariance Structures. *Psychological Bulletin*, 88, 588-606.

Cole, David A and Scott E. Maxwell (1985) Multitrait-Multimethod Comparisons Across Populations: A Confirmatory Factor Analytic Approach. *Multivariate Behavioral Research*. 2, 389-417.

Jöreskog, Karl Gustav (1969) A General Approach to Confirmatory Maximum Likelihood Factor Analysis. *Psychometrika*. 34 (2), 183-200.