PSYC 402: Structural Equation Modelling

By Jack Zhou

1 Algebra Review

Let capital letters be matrices, lowercase letters be random vectors, and greek lowercase letters be constant vectors. We proceed with some definitions.

1.1 Matrix Algebra

TBD

1.2 Expectation Algebra

An expectation operator is denoted as $\mathbb{E}[\cdot]$, it has the following properties:

- $\mathbb{E}[Az] = A\mathbb{E}[z]$
- $\mathbb{E}[\alpha] = \alpha$
- $\mathbb{E}[z + r] = \mathbb{E}[z] + \mathbb{E}[r]$

1.3 Variance/Covariance Algebra

The covariance operator is denoted as $\text{Cov}[\cdot]$. It is defined as follows:

$$\begin{aligned}\text{Cov}[x, y] &= \mathbb{E}\left[(x - \mathbb{E}[x])(y - \mathbb{E}[y])^T\right] \\ &= \mathbb{E}\left[xy^T\right] - \mathbb{E}[x]\mathbb{E}[y]^T\end{aligned}$$

If only one argument is given to $\text{Cov}[\cdot]$, then it is understood that both arguments are the same (ex. $\text{Cov}[a] = \text{Cov}[a, a] = \text{Var}[a]$).
Below are some properties of the covariance operator:

- $\text{Cov}[Az] = A\text{Cov}[z]A^T$
- $\text{Cov}[x + y] = \text{Cov}[x] + \text{Cov}[x, y] + \text{Cov}[y, x] + \text{Cov}[y]$
- $\text{Cov}[x, y + z] = \text{Cov}[x, y] + \text{Cov}[x, z]$
- $\text{Cov}[x, \alpha] = 0$

The following result is particularly useful:

$$\begin{aligned}&\text{Cov}[Az + \dagger] \\ =&\text{Cov}[Az] + \text{Cov}[Az, \dagger] + \text{Cov}[\dagger, Az] + \text{Cov}[\dagger] \\ =&A\text{Cov}[z]A^T + A\text{Cov}[z, \dagger] + \text{Cov}[\dagger, z]A^T + \text{Cov}[\dagger]\end{aligned}$$

$\dagger$ can either be a random vector or constant. If it is a constant, then $A\text{Cov}[z, \dagger] + \text{Cov}[\dagger, z]A^T + \text{Cov}[\dagger] = 0$.

2 Three classes of models

Structural Equations can be classified by 3 different classes.

2.1 Path Model

A path models can also be referred to as a multivariate regression model for observed cariables, or a causal model. A path model can contain p dependent (*endogenous*) variables and q independent (*exogenous*) variables, denoted as the following:

- $\tilde{\mathbf{y}}$: a $p \times 1$ vector of *endogenous* observed variables
- $\tilde{\mathbf{x}}$: a $q \times 1$ vector of *exogenous* observed variables
- $\tilde{\zeta}$: a $p \times 1$ vector of residuals

As suggested by the match of dimensionality between $\tilde{\mathbf{y}}$ and $\tilde{\zeta}$, each of the entries in $\tilde{\mathbf{y}}$ corresponds to a value in $\tilde{\zeta}$ with the same index. That is, for each dependent variable $\tilde{\mathbf{y}}_i$, we have a corresponding residual $\tilde{\zeta}_i$ that the model is unable to account for.
It is important to note that path models do not contain any latent (unobserved) variables, as we can see in the following equation:

$$\tilde{\mathbf{y}} = [\mathbf{B}\tilde{\mathbf{y}} + \mathbf{\Gamma}\tilde{\mathbf{x}}] + \tilde{\zeta}$$

The general interpretation is that with this model, the set of endogenous variables are to be described in terms of a sum between the structural model $[\mathbf{B}\tilde{\mathbf{y}} + \mathbf{\Gamma}\tilde{\mathbf{x}}]$ and residuals $\tilde{\zeta}$.
The model consists of two parts, the $\mathbf{B}\tilde{\mathbf{y}}$ term describes the regression of endogenous variables onto themselves, and the $\mathbf{\Gamma}\tilde{\mathbf{x}}$ term describes the regression of exogenous variables onto the endogenous variables, in combined effort to best account for the endogenous variables.
One glaring issue is that in order to solve for $\tilde{\mathbf{y}}$, it should only be on one side of the equation, not both. We can address this with algebra:

$$\begin{aligned}\tilde{\mathbf{y}} &= [\mathbf{B}\tilde{\mathbf{y}} + \mathbf{\Gamma}\tilde{\mathbf{x}}] + \tilde{\zeta} \\ \tilde{\mathbf{y}} - \mathbf{B}\tilde{\mathbf{y}} &= \mathbf{\Gamma}\tilde{\mathbf{x}} + \tilde{\zeta} \\ (I - \mathbf{B})\tilde{\mathbf{y}} &= \mathbf{\Gamma}\tilde{\mathbf{x}} + \tilde{\zeta} \\ \tilde{\mathbf{y}} &= (I - \mathbf{B})^{-1}(\mathbf{\Gamma}\tilde{\mathbf{x}} + \tilde{\zeta})\end{aligned}$$

The equation $\tilde{\mathbf{y}} = (I - \mathbf{B})^{-1}(\mathbf{\Gamma}\tilde{\mathbf{x}} + \tilde{\zeta})$ is known as the reduced form, which is easier to work with mathematically. Below are descriptions to all the matrices in a path model:

- $\mathbf{B}$: $p \times p$ regression coefficients for $\tilde{\mathbf{y}}$
- $\mathbf{\Gamma}$: $p \times q$ regression coefficients for $\tilde{\mathbf{x}}$
- $\mathbf{\Phi}$: $q \times q$ covariance matrix of $\tilde{\mathbf{x}}$
- $\mathbf{\Psi}$: $p \times p$ covariance matrix of $\tilde{\zeta}$

Note that $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are matrices that aren't seen in the model.
For any covariance matrix $\mathbf{M}$, any $\mathbf{M}_{ii}$ entry is the variance of the ith variable and any $\mathbf{M}_{ij}$ entry is the covariance between the ith and jth variable.
$\mathbf{\Phi}$ can be considered a parameter for $\tilde{\mathbf{y}}$. Therefore, we consider it as a starting point.
Often times we see correlations between residuals (ex. in longitudinal models), so $\mathbf{\Psi}$ does not necessarily have to be diagonal.

2.1.1 Kinds of Path Models

- **Standard Multiple Regression**

  Suppose we have an endogenous variable y_1, its corresponding ζ_1, and a set of exogenous variables $\{x_1, ..., x_q\}$. A Standard Multiple Regression can be seen as a digraph with the following properties:

  - An edge (ζ_1, y_1) exists; the residual impacts the endogenous variable
  - For all x_i in the set of exogenous variables, an edge (x_i, y_1) exists; each of the exogenous varibales impacts the endogenous variable
  - For all $x_i \neq x_j$, an edge (x_i, x_j) exists to indicate correlations between exogenous variables

  Let $\tilde{\mathbf{x}}$ be a $p \times 1$ vector of exogenous variables, Γ be a $1 \times q$ vector of regression coefficients for $\tilde{\mathbf{x}}$, then such graph can be written as an equation as follows:

  $$y_1 = \Gamma\tilde{\mathbf{x}} + \zeta_1$$

- **Standard Multivariate Regression**

  In this case, there are greater than $p > 1$ endogenous variables as well as $q > 1$ exogenous variables.

  The setting is similar to the Standard Multiple Regression case, but there is more than one y_i, but no edges between any $y_i \neq y_j$ (no correlation between any endogenous variables).

  In addition to the notation used in standard multiple regression, we let $\tilde{\mathbf{y}}$ be the $p \times 1$ vector of endogenous variables, and $\mathbf{B}$ be a $p \times p$ matrix as defined before. As the model asserts that there are no correlations between any endogenous variables, $\mathbf{B}$ is a zero matrix, and the covariance matrix of $\tilde{\mathbf{y}}$, $\mathbf{\Psi}$, is taken to be diagonal.

- **First Order Autoregressive Time Series**

  A chain of measurements are collected in sequential order, imposing a temporal structure.

  The first measurement is considered exogenous, so we would name it x_1. From there on, each measurement that comes after are all endogenous variables y_i with residuals ζ_i.

  Each variable is linearly dependent on the variable before, causing a chain of dependencies.

2.2 Linear Factor/Measurement Models

In the basic treatment, we have the following variables.

- $\tilde{\mathbf{x}}$: a $q \times 1$ vector of *indicators* (observed variables)
- $\tilde{\xi}$: a $n \times 1$ vector of *latent*, common-factor variables
- $\tilde{\delta}$: a $q \times 1$ vector of measurement error variables

In a parallel alternative form, we have these instead:

- $\tilde{\mathbf{y}}$: a $q \times 1$ vector of *indicators* (observed variables)
- $\tilde{\eta}$: a $n \times 1$ vector of *latent*, common-factor variables
- $\tilde{\epsilon}$: a $q \times 1$ vector of measurement error variables

Note that here we have a set of observed variables and a set of latent variables. The observed variables play the role in trying to determine the latent variables. Imagine measuring intelligence (abstract and latent) with test scores (indicators; concrete and observable).
Under Spearman, all $\tilde{\mathbf{x}}_i$ is causally (regression) dependent upon a $\tilde{\xi}_i$ for some i. Moreover, upon regressing $\tilde{\mathbf{x}}$ on $\tilde{\xi}$, within-correlations in $\tilde{\mathbf{x}}$ no longer exist.
In other words, the reason why different $\tilde{\mathbf{x}}_i$ could be correlated is because they arise from a common latent variable; $\tilde{\xi}$ can be taken as an explanation as to why within-correlations in $\tilde{\mathbf{x}}$ exist.
We can summarize this model with the following formula:

$$\tilde{\mathbf{x}} = \mathbf{\Lambda_x}\tilde{\xi} + \tilde{\delta}$$

Where $\mathbf{\Lambda_x}$ is $q \times n$ matrix called factor loadings, linking the indicators to the latent variables.
There is also a $n \times n$ matrix $\mathbf{\Phi}$, which is the covariance matrix of $\tilde{\xi}$. Typically there are many restriction on this matrix, for example we would assert it to be a diagonal matrix for orthogonality.
Finally, we also have covariance for $\tilde{\delta}$ as $\mathbf{\Theta}_\delta$, a $q \times q$ matrix. Recall that upon regression, the indicators become uncorrelated, suggesting $\mathbf{\Theta}_\delta$ is typically diagonal.

2.2.1 Correlation accross sets of indicator/latent variables

Sometimes we are interested in the correlation between multiple latent variables, and the only access we have towards them are their indicators, so the correlation would have to run on some linear weighted composite of the indicators per grouping.
The resulting correlation, which is used to indicate the correlation between the actual latent variables, is known to be less than the actual correlation between the latent variables.
To account for this issue, a mechanism of disattenuation is used. For two latent variables τ_1, τ_2, their correlation ρ_{τ_1,τ_2} can be determined by a function of observed composites c_1, c_2 (respectively) as follows:

$$\rho_{\tau_1,\tau_2} = \frac{\rho_{c_1,c_2}}{\sqrt{\rho_{c_1,c_1'} \cdot \rho_{c_2,c_2'}}}$$

Where $\rho_{c_1,c_1'}$ denotes reliability for c_1, and likewise for c_2. This is known as the disattenuating correlation formula. With structural equation modelling, we can add an edge to the path diagram that joins the latent variables and work from there, without the need for disattenuation.

2.3 Multivariate Regression Model for Latent Variables

Often known as the full model, a multivariate regression model for latent works like a causal model in latent terms. We have the following variables:

- $\tilde{\eta}$: a $m \times 1$ vector of latent endogenous variables
- $\tilde{\xi}$: a $n \times 1$ vector of latent exogenous variables
- $\tilde{\zeta}$: a $m \times 1$ vector of residuals
- $\tilde{\mathbf{x}}$: a $q \times 1$ vector of indicators of $\tilde{\xi}$
- $\tilde{\mathbf{y}}$: a $p \times 1$ vector of indicators of $\tilde{\eta}$
- $\tilde{\delta}$: a $q \times 1$ vector of measurement errors on $\tilde{\mathbf{x}}$
- $\tilde{\epsilon}$: a $p \times 1$ vector of measurement errors on $\tilde{\mathbf{y}}$

In terms of equations, we have the following:

- **Regresion Model For Unobservables**
  $$\tilde{\eta} = \mathbf{B}\tilde{\eta} + \mathbf{\Gamma}\tilde{\xi} + \tilde{\zeta}$$
- **Measurement Model for $\tilde{\mathbf{x}}$**
  $$\tilde{\mathbf{x}} = \mathbf{\Lambda_x}\tilde{\xi} + \tilde{\delta}$$
- **Measurement Model for $\tilde{\mathbf{y}}$**
  $$\tilde{\mathbf{y}} = \mathbf{\Lambda_y}\tilde{\eta} + \tilde{\epsilon}$$

In line with the notation as previously introduced before, we have the following parameter matrices:

- $\mathbf{B}$: a $m \times m$ regression matrix for $\tilde{\eta}$ on $\tilde{\eta}$

  As it makes no sense to say that an $\tilde{\eta}_i$ is dependent on $\tilde{\eta}_i$ itself, the diagonal elements of $\mathbf{B}$ are zero.
- $\mathbf{\Gamma}$: a $m \times n$ regression matrix for $\tilde{\xi}$ on $\tilde{\eta}$

  The entries in $\mathbf{\Gamma}$ are usually handpicked to be zero or nonzero. It essentially spells out our beliefs of how $\tilde{\eta}$ relates to $\tilde{\xi}$.
- $\mathbf{\Phi}$: a $n \times n$ covariance matrix of $\tilde{\xi}$
- $\mathbf{\Psi}$: a $m \times m$ covariance matrix of $\tilde{\zeta}$
- $\mathbf{\Lambda_x}$: a $q \times n$ matrix of loadings
- $\mathbf{\Theta}_\delta$: a $q \times q$ covariance matrix of $\tilde{\mathbf{x}}$
- $\mathbf{\Lambda_y}$: a $p \times m$ matrix of loadings
- $\mathbf{\Theta}_\epsilon$: a $p \times p$ covariance matrix of $\tilde{\mathbf{y}}$

3 Aims of Modelling

A model is a candidate explanation of the joint distribution of a set of observable varibles. We place restrictions on parameter matrices in a model, reducing its complexity and also targetting specific aspects of the joint distribution. Typically targetting a set of association parameters within the distribution. In structural equation modelling, we are interested in the covariances $\mathbf{\Sigma}$ between the variables at the population level.

Placing restrictions effectively reduces the amount of parameters to estimate. Typically we want the number of parameters to estimate t to be lower than the amount of input data elements r, so there needs to be lots of restrictions placed on the model.

By using a model with parameter matrices, we can denote the set of parameters to estimate as $\tilde{\mathbf{\Theta}}$. We use $\tilde{\mathbf{\Theta}}$ to estimate $\mathbf{\Sigma}$, denoted as $\hat{\mathbf{\Sigma}}(\tilde{\mathbf{\Theta}})$, a function of $\tilde{\mathbf{\Theta}}$.

Once $\tilde{\mathbf{\Theta}}$ is determined, we can find $\hat{\mathbf{\Sigma}}(\tilde{\mathbf{\Theta}})$ and determine its likeliness to a sample estimate s of $\tilde{\mathbf{\Sigma}}$.

4 Approach

To begin, for a model from any of the three classes previously discussed, we proceed in a sequence of steps. We will discuss solely in terms of a observed variable path model, where no latent variables are considered.

4.1 Rendering

The rendering stage effectively generates a path diagram after specifying the observed and latent variables. The relationships between the variables are discussed and established and a diagram is drawn.

The convention for drawing diagrams suggests drawing ellipsis for latent variables and rectangles for observed variables. Residuals are usually drawn with no enclosing shape.

A single headed arrow denotes a regression dependency, where a double headed arrow denotes a correlational relationship.

Suppose a variable a has a regression dependency on another variable b, and b has a regression dependency on a, then a feedback loop is created and the entire model is said to be non-recursive.

4.2 Mathematization

The mathematization stage takes in a path diagram and re-expresses the diagram in terms of mathematical formulas.
Conventionally we have the following values to denote the number of variables in a model:

- NX: Number of $\mathbf{x}$ variables
- NY: Number of $\mathbf{y}$ variables
- NK: Number of ξ variables
- NE: Number of η variables
- NZ: Number of ζ variables

Note that since we are working with observed value path models, NK=NE=0.
Given these values, we can determine the dimensionality of $\mathbf{B}$, $\mathbf{\Gamma}$, $\mathbf{\Phi}$, $\mathbf{\Psi}$ and fill in entries based on connections in the path diagram.
Often the parameters in the path diagram have subscripts that determine which index it belongs in for the corresponding matrix. For example, γ_{13} indicates that it belongs to row 1 column 3 for $\mathbf{\Gamma}$.
Usually for $\mathbf{\Phi}$ and $\mathbf{\Psi}$ the parameters may not be indicated explicitly, but as a general rule the diagonal elements (the variances) are by default nonzero.
From this information, we can generate a model implied covariance matrix $\hat{\mathbf{\Sigma}}(\tilde{\mathbf{\Theta}})$, note that it is a matrix for the joint distribution, which means its dimensionality is $(NX + NY) \times (NX + NY)$. We can partition $\hat{\mathbf{\Sigma}}(\tilde{\mathbf{\Theta}})$ as follows:

$$\hat{\mathbf{\Sigma}}(\tilde{\mathbf{\Theta}}) = \left[\begin{array}{c|c} \hat{\mathbf{\Sigma}}_y(\tilde{\mathbf{\Theta}}) & \hat{\mathbf{\Sigma}}_{yx}(\tilde{\mathbf{\Theta}}) \\ \hline \hat{\mathbf{\Sigma}}_{xy}(\tilde{\mathbf{\Theta}}) & \mathbf{\Phi} \end{array}\right]$$

Where

$$\hat{\mathbf{\Sigma}}_y(\tilde{\mathbf{\Theta}}) = (I - \mathbf{B})^{-1} \left[\mathbf{\Gamma}\mathbf{\Phi}\mathbf{\Gamma}^T + \mathbf{\Psi}\right] ((I - \mathbf{B})^{-1})^T$$

and

$$\hat{\mathbf{\Sigma}}_{yx}(\tilde{\mathbf{\Theta}}) = (I - \mathbf{B})^{-1}\mathbf{\Gamma}^T\mathbf{\Phi}$$

4.3 Identification

TBD