PSYC 402: Structural Equation Modelling

By Jack Zhou

1 Linear Algebra Review

TBD

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 $\tilde{\mathbf{y}}$
- $\mathbf{\Gamma}$: $q \times q$ 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 $\tilde{\zeta}$

$\mathbf{\Phi}$ and $\mathbf{\Psi}$ are matrices that aren't seen in the model.