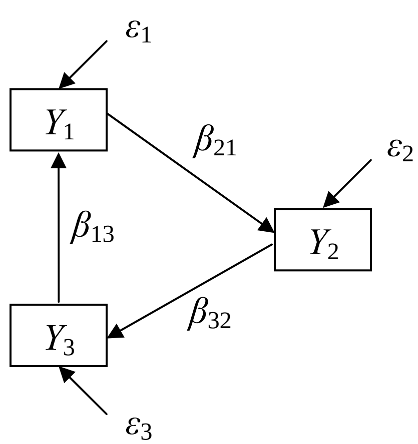
Covariance Structure and Factor Models

Example questions for the mid-term exam

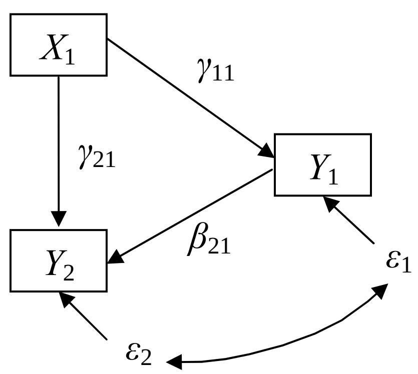
1. Describe the two conditions of the recursive path model, and provide two examples (as diagram models) which do not meet each one of the conditions.

<u>Answer:</u>
The first condition is that there is no feedback loop of causal influence. This condition is fulfilled if the **B** matrix (that defines causal effects of one dependent variable to another except for the DVs as indicators of latent variables) is a lower triangular matrix. An example that violates this condition is



where each of the 3 dependent variables indirectly influences itself through the other DVs.

The second condition of recursive models is that all error terms are mutually orthogonal (so that any covariances between dependent variables are fully explained by covariances between independent variables). An example that doesn't meet this condition is



2. (a) Describe the difference between the OLS (ordinary least squares) and ULS (unweighted least squares) estimators. In your description, use the multiple regression model as an example whose structural equation is written in the LISREL notation as:

$$y = \boldsymbol{\gamma}'\mathbf{x} + \zeta$$

where both y and the q-elements of $\mathbf{x}$ are in deviation form. (b) Show that the multiple regression model has zero degrees of freedom (i.e., data *df* = the number of free parameters in the model).

Answer:

(a) The OLS estimator of the regression weights $\boldsymbol{\gamma}$ minimizes the sum of squared residuals with respect to y (i.e., $y - \hat{y} = \zeta$, which is estimated as $\hat{\boldsymbol{\gamma}} = (\mathbf{x}\mathbf{x}')^{-1}\mathbf{x}y$). On the other hand, the ULS estimator minimizes the sum of squared residuals with respect to the covariance matrix (i.e., $\mathbf{S} - \hat{\boldsymbol{\Sigma}}$ of order $q + 1$ including 1 DV and q IVs).

The goodness of fit by the OLS is often indicated by R^2 that represents the proportion of variance of y accounted for by $\mathbf{x}$, which is less than 1 unless the variance of error term ζ is 0. In contrast, the ULS estimator of the multiple regression model always perfectly fits the data $\mathbf{S}$ in that the model implied covariance matrix $\hat{\boldsymbol{\Sigma}}$ includes the variance of the error term. (Note that the variances and covariance of $\mathbf{x}$ are identical in both $\mathbf{S}$ and $\hat{\boldsymbol{\Sigma}}$, and so that part does not contribute to the sum of squared residuals.)

(b) The data covariance matrix $\mathbf{S}$ has $(q+1)(q+2)/2$ distinctive elements. Under the multiple regression model, we have q explanatory variables that are allowed to be freely correlated (and so $q(q+1)/2$ free parameters for their variances and covariances), one error term that are not correlated with the explanatory variables (and so 1 free parameter for its variance), and q regression weights. Thus, the total number of free parameters is

$$\text{\# of free parameters} = \frac{q(q+1)}{2} + q + 1 = \frac{(q+1)(q+2)}{2} = \text{\# of distinctive data points}$$

which is the same as the number of distinctive elements in the data, showing that the multiple regression model has zero degrees of freedom from the perspective of covariance structure analysis.