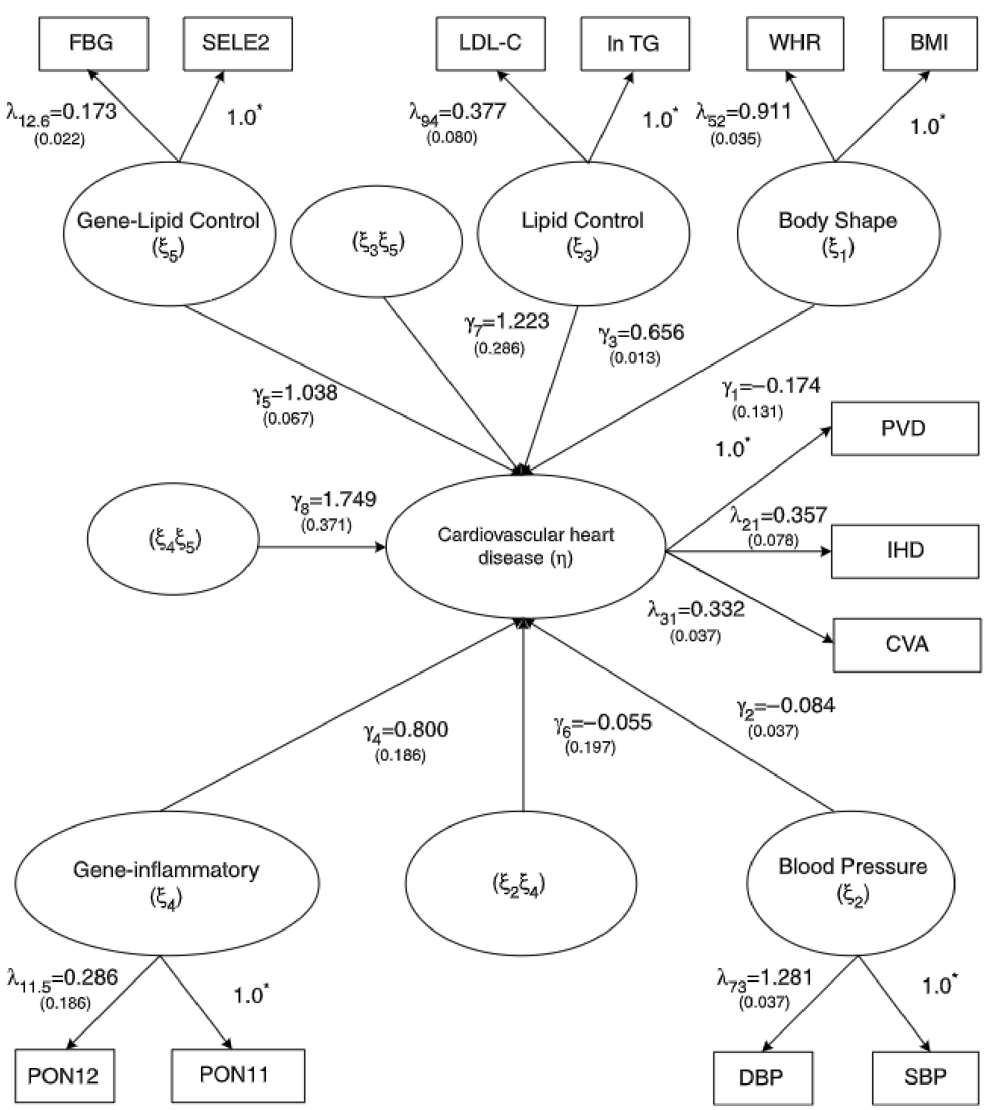
STAT 5020 : Topics in Multivariate Analysis

Assignment 1 (Due date: 1-Mar-2023)

Academic year 22/23, 2nd term

1. The path diagram of a structural equation model (SEM) is presented in the following figure, along with the estimates of important parameters. Assume that the response variables are continuous.



a. Write down the explicit form of the model.
b. Interpret the model in terms of the variables provided in the rectangles and ellipses.
c. Is the model identified? Why?
d. List and interpret the fixed and unknown parameters in the model.
e. Discuss all possible submodels of model (a).
f. Discuss the differences between SEMs and conventional regression models.

2. A non-linear SEM is defined as follows: for $i = 1, \ldots, n$,

$$
\begin{aligned}
y_{i1} &= \mu_1 + a_1 * c_i + \eta_{i1} + \epsilon_{i1},\\
y_{i2} &= \mu_2 + a_2 * c_i + \lambda_{21} * \eta_{i1} + \epsilon_{i2},\\
y_{i3} &= \mu_3 + a_3 * c_i + \eta_{i2} + \epsilon_{i3},\\
y_{i4} &= \mu_4 + a_4 * c_i + \lambda_{42} * \eta_{i2} + \epsilon_{i4},\\
y_{i5} &= \mu_5 + a_5 * c_i + \xi_{1i} + \epsilon_{i5},\\
y_{i6} &= \mu_5 + a_6 * c_i + \lambda_{63} * \xi_{i1} + \epsilon_{i6},\\
y_{i7} &= \mu_7 + a_7 * c_i + \lambda_{73} * \xi_{i1} + \epsilon_{i7},\\
y_{i8} &= \mu_8 + a_8 * c_i + \xi_{i2} + \epsilon_{i8},\\
y_{i9} &= \mu_9 + a_9 * c_i + \lambda_{94} * \xi_{i2} + \epsilon_{i9},\\
y_{i,10} &= \mu_{10} + a_{10} * c_i + \lambda_{10,4} * \xi_{i2} + \epsilon_{i,10},\\
\eta_{i1} &= b_1 * d_{i1} + b_2 * d_{i2} + b_3 * d_{i3} + \gamma_1 * \xi_{i1} + \gamma_2 * \xi_{i2} + \gamma_3 * \xi_{i1}^2 + \delta_{i1},\\
\eta_{i2} &= b_4 * d_{i1} + b_5 * d_{i2} + b_6 * d_{i3} + \gamma_4 * \eta_{i1} + \gamma_5 * \xi_{i1} + \gamma_6 * \xi_{i2} + \gamma_7 * \xi_{i1} * \xi_{i2} + \gamma_8 * \xi_{i2}^2 + \delta_{i2},
\end{aligned}
$$

where

- $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2})^T \sim N(0, \boldsymbol{\Phi})$ with $\boldsymbol{\Phi}$ being a general covariance matrix,
- $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{i,10})^T \sim N(0, \boldsymbol{\Psi}_\epsilon)$ with $\boldsymbol{\Psi}_\epsilon = diag(\psi_{\epsilon 1}, \ldots, \psi_{\epsilon 10})$,
- $\boldsymbol{\delta}_i = (\delta_{i1}, \delta_{i2})^T \sim N(0, \boldsymbol{\Psi}_\delta)$ with $\boldsymbol{\Psi}_\delta = diag(\psi_{\delta 1}, \psi_{\delta 2})$,

a. Write the model in a matrix form and draw the corresponding path diagram.
b. In a Bayesian analysis, what prior distributions do we usually assign to the model parameters? Why?
c. Derive the full conditional distributions in detail for the unknown parameters.