STA 431s09 Assignment 9

Do this assignment in preparation for the quiz on Friday, March 20th. For Question 4, bring your log and list files to the quiz. The other questions are practice for the quiz and are not to be handed in.

1. Consider the following confirmatory factor analysis model:

$$
\begin{aligned}
X_1 &= \lambda_1 F_1 + e_1 \\
X_2 &= \lambda_2 F_1 + e_2 \\
X_3 &= \lambda_3 F_1 + e_3 \\
X_4 &= F_2 + e_4 \\
X_5 &= F_2 + e_5,
\end{aligned}
$$

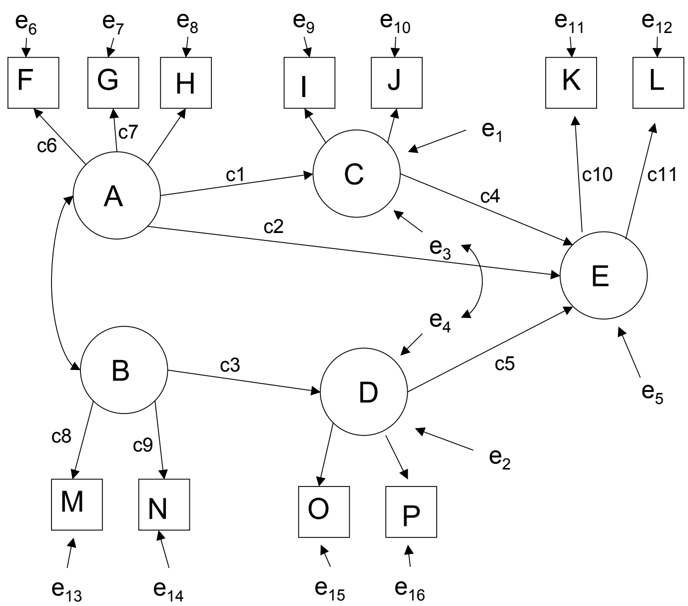
   where F_1 and F_2 are latent variables (factors), $e_1, \ldots e_5$ are error terms independent of F_1 and F_2 and each other, $Var(e_j) = \psi_j$ for $j = 1, \ldots, 5$,

$$
V \left[\begin{array}{c} F_1 \\ F_2 \end{array} \right] = \left[\begin{array}{cc} \phi_{1,1} & \phi_{1,2} \\ \phi_{1,2} & \phi_{2,2} \end{array} \right],
$$

   λ_1, λ_2 and λ_3 are all non-zero, and everything is multivariate normal with expected value zero.

   (a) Draw a path diagram for this model. Put coefficients on the arrows.
   (b) What is the parameter vector $\boldsymbol{\theta}$ for this model?
   (c) Does this model pass the test of the counting rule?
   (d) Calculate $\boldsymbol{\Sigma}$, the variance-covariance matrix of the observed variables.
   (e) Is this model identified? Answer Yes or No and prove your answer.
   (f) If the whole model is not identified, identification of some parameters is still a possibility. If this is the case, say what parameters are identified.
   (g) What (one) constraint would *you* introduce to make this model identified? Prove that the model is identified with the constraint you propose.

2. In this question, you will exercise (and demonstrate) your understanding of the LISREL model notation; see Christine's handout. In the path diagram below, $A \ldots P$ are random variables and $v_1, \ldots, v_{20}$ and $c_1, \ldots c_{11}$ are constants. Let $Var(e_j) = v_j$ for $j = 1, \ldots, 16$, $Cov(e_3, e_4) = v_{17}$, $Var(A) = v_{18}$, $Var(B) = v_{19}$, and $Cov(A, B) = v_{20}$. If a straight arrow is not marked with a coefficient, the coefficient equals one. All expected values are zero.



(a) In terms of the model shown above, give the following constants: m, n, p, q.

(b) In terms of the model shown above, give the following matrices. They *must* be of the right dimensions for you to get any marks. Check! Is it possible to do the matrix multiplications and additions indicated in the LISREL model?

$$
\begin{array}{cccc}
\boldsymbol{\eta} & \boldsymbol{\xi} & \boldsymbol{\zeta} & \\
\boldsymbol{\epsilon} & \boldsymbol{\delta} & \mathbf{x} & \mathbf{y} \\
\boldsymbol{\beta} & \boldsymbol{\Gamma} & \boldsymbol{\Lambda}_y & \boldsymbol{\Lambda_x} \\
\boldsymbol{\alpha} & \boldsymbol{\nu}_y & \boldsymbol{\nu}_x & \boldsymbol{\kappa} \\
\boldsymbol{\Phi} & \boldsymbol{\Psi} & \boldsymbol{\Theta}_\epsilon & \boldsymbol{\Theta}_\delta
\end{array}
$$

3. The following model is based on what we will call the *double measurement design*. In the double measurement design, two parallel sets of measurements are taken on a collection of latent variables (endogenous, exogenous or some of each). The two sets of measurements are usually collected on different occasions and in two

different ways. Great pains are taken to ensure that the measurement errors are uncorrelated between sets; correlation *within* each set is allowed. For example, the cases could be companies, with one set of measurements consisting of information provided by management. If the same information were independently determined by a forensic audit, it might be safe to assume that the measurement errors were uncorrelated across occasions, though errors of measurement within sets (especially by the company) would almost certainly be correlated.

This not very interesting as a factor analysis model, because all the factor loadings equal one. But it is useful as a measurement model for more general structural equation models with latent variables. In the following, all the latent variables are collected into a "factor' $\mathbf{F}$, and the observable variables are collected into $\mathbf{D}_1$ (measured by method one), and $\mathbf{D}_2$ (measured by method two). Let

$$\begin{aligned} \mathbf{D}_1 &= \mathbf{F} + \mathbf{e}_1 \\ \mathbf{D}_2 &= \mathbf{F} + \mathbf{e}_2, \end{aligned}$$

where all expected values are zero as usual, $V(\mathbf{F}) = \mathbf{\Phi}$, $V(\mathbf{e}_1) = \mathbf{\Omega}_1$ and $V(\mathbf{e}_2) = \mathbf{\Omega}_2$; all the covariance matrices are positive definite. The random vectors $\mathbf{F}$, $\mathbf{e}_1$ and $\mathbf{e}_2$ (all k by 1) are independent. Notice that the variance-covariance matrices of the error terms need *not* be diagonal. This is a big improvement on the usual measurement model, in which *all* the measurement errors are assumed to be independent.

(a) Write $\mathbf{\Sigma}$ as a 2×2 partitioned matrix (a matrix of matrices).

(b) Prove that the parameter matrices $\mathbf{\Phi}, \mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ are identified.

(c) Is the model just-identified, or is it over-identified?

(d) How many over-identifying restrictions are there? Your answer is a single number that depends on k.

4. The file `littlemath.data` (see link on the course web page in case the one from this document does not work) contains data from a study of students taking first-year calculus on one campus of a large North American university. The students took a diagnostic test with two parts: pre-calculus and calculus. Marks in high school calculus were available, as were marks in first-year university calculus. The question is whether these data are unidimensional. That is, do data arise from a single underlying factor ("Math Skill" or something)?

   Recall that while the factor loadings from an exploratory factor analysis are usually worthless, the test for number of factors is pretty good. So, use `proc factor` to check for unidimensionality (because it's a lot easier than `proc calis`). What you are seeking is two numbers: a test statistic and a p-value. Know what each one means. Are the data unidimensional? Answer Yes or No.