STA 431s17 Assignment Eleven[1]

These problems are preparation for the quiz on Wednesday March 29th, and are not to be handed in.

1. Refer to the following as the *original model.* It's really not quite original because it has been centered. Let

$$
\begin{aligned}
D_1 &= \lambda_1 F_1 + e_1 \\
D_2 &= \lambda_2 F_1 + e_2 \\
D_3 &= \lambda_3 F_1 + e_3 \\
D_4 &= \lambda_4 F_2 + e_4 \\
D_5 &= \lambda_5 F_2 + e_5 \\
D_6 &= \lambda_6 F_2 + e_6,
\end{aligned}
$$

where all expected values are zero, $Var(e_i) = \omega_i$ for $i = 1, \ldots, 6$, $cov\left(\begin{array}{c} F_1 \\ F_2 \end{array} \right) = \left(\begin{array}{cc} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{array} \right)$, the factors are independent of the error terms, and all the error terms are independent of each other. All the factor loadings are non-zero.

(a) Make a path diagram.

(b) Does this model pass the test of the parameter count rule? Answer Yes or No and give the numbers.

(c) Calculate or otherwise obtain the covariance matrix of the observable variables.

(d) Show that the parmeters of this model are not identifiable by giving specific λ_1', λ_2', λ_3', ϕ_{11}' and ϕ_{12}' (different from the original parameters) that yield the same covariance matrix as the parameters of the original model. The parameters not named are th same in both models.

(e) Show that λ_2/λ_1 is identifiable.

(f) Show that ω_4 is identifiable.

(g) Show that the reliability of D_1 is identifiable. If λ_1 is getting in your way, remember that the reliability is one minus the proportion of variance that is *not* error.

(h) When we set a factor loading to one in order to obtain identifiability, we are not really *assuming* something like $\lambda_1 = 1$. That would be wildly unrealistic. Instead,

---

we are transforming the latent variable to obtain $D_1 = F_1^* + e_1$. Let's do the same thing to F_2 as well (the standard trick) so that $D_4 = F_2^* + e_4$. Now we will see the change of variables produces a re-parameterization that changes the meaning of the other parameters in the model.

    i. Under the re-parameterized (surrogate) model, $D_2 = \lambda_2^* F_1^* + e_2$. What is λ_2^*? There's just a little work to show. Show it.

    ii. What is λ_6^*? There's just a little work to show. Show it.

    iii. What is ϕ_{22}^*? There's just a little work to show. Show it.

    iv. What is ϕ_{12}^*? There's just a little work to show. Show it.

    v. What is λ_1^*? λ_4^*?

(i) The other standard trick is to set the variances of the factors to one. For the resulting surrogate model,

    i. What are F_1^* and F_2^* in terms of the original model?

    ii. What is λ_2^* in terms of the original model?

    iii. Show that $\phi_{12}^* = Corr(F_1, F_2)$.

    iv. The parameters of this surrogate model will be identifiable provided that the sign of one factor loading is known for each factor. Assuming $\lambda_1^* > 0$ and $\lambda_4^* > 0$, give formulas for λ_1^* and λ_4^* in terms of σ_{ij} quantities. You can use your answer to Question 1c, modifying the notation a bit in your mind.

    v. Using your answer to the last question, give a formula for ϕ_{12}^* in terms of σ_{ij} quantities.

    vi. Now take that same function of the σ_{ij} quantities assuming the *original* model. When you solve for ϕ_{12}^*, what are you really identifying in terms of the original model? The beauty of this is that ϕ_{12}^* is the correlation between factors under the original model and the surrogate model.

(j) Continuing with the two-factor example in which variances of the factors equal one under the surrogate model, suppose that in the original model, $F_2 = \gamma F_1 + \epsilon$.

    i. Give the 2×2 matrix $\boldsymbol{\Phi}$ assuming the original model.

    ii. For the surrogate latent model $F_2^* = \gamma^* F_1^* + \epsilon^*$, express γ^* and ψ^* in terms of the parameters of the original model.

    iii. What are F_1^*, F_2^* and ϵ^* in terms of the original model?

    iv. What is $Var(\epsilon^*)$ n terms of the parameters of the original model?

2. Now consider an original (except it's centered) single-factor model in which

$$\begin{aligned} D_1 &= \lambda_1 F + e_1 \\ D_2 &= \lambda_2 F + e_2 \\ D_3 &= \lambda_3 F + e_3 \\ D_4 &= \lambda_4 F + e_4 \end{aligned}$$

where $e_1, \ldots, e_4, F$ are all independent, $Var(e_j) = \omega_j$, $Var(F) = \phi$ and $\lambda_j \neq 0$. Write Σ in terms of the model parameters. This is mostly just for later use. You have already done almost all the work in Question 1c.

3. Now set $\lambda_1 = 1$ in the model of Question 2, resulting in a surrogate model.

   (a) Show that $\lambda_2, \lambda_3, \lambda_4$ and ϕ are identifiable by solving for them explicitly in terms of σ_{ij} quantities. The ω_j are identifiable too, but don't bother with them.

   (b) Now substitute your solutions for the parameters from Question 3a back into the six covariance structure equations for σ_{ij} where $i \neq j$. The result is two model-induced *equality constraints* on the covariances. What are they? By the way, this is a general method for deriving equality constraints (the null hypothesis of the goodness of fit test), but the parameters have to be identifiable.

   (c) But all is not lost. Verify that the constraints you obtained for the surrogate model also are true of the original model of Question 2. It is interesting and useful that models with non-identifiable parameters can imply testable constraints on the covariance matrix.

   (d) Are the constraints also true under the other surrogate model with $\phi = 1$?

4. For the following path diagram, assume that any arrow unmarked by a symbol has the coefficient one. When you give model equations and matrices below, please use the symbols from the path diagram.
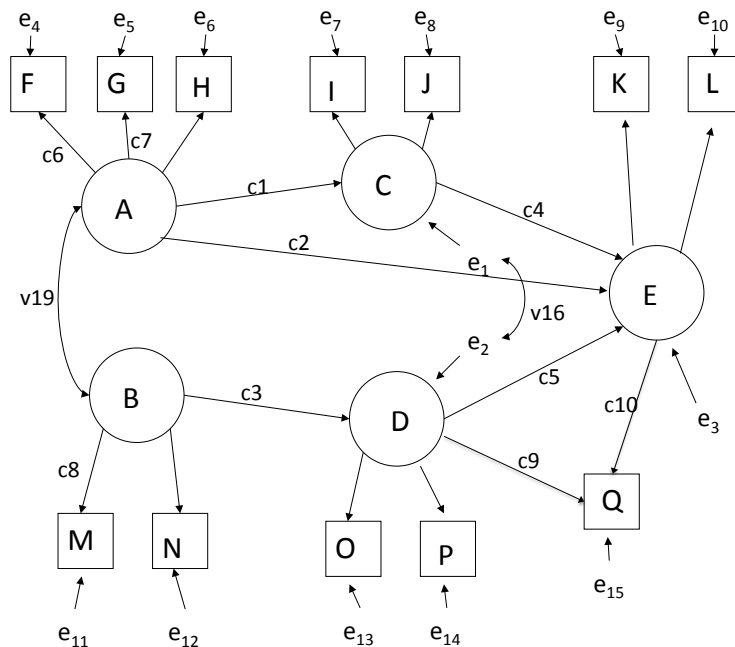

   Explain why the model parameters are all identifiable, making specific reference to the identifiability rules on the reference sheet.

5. Write symbols on the arrows of the path diagram below, selecting a surrogate model whose parameters are identifiable. Cite the rule you are using by name, letter and number. What else must you assume?
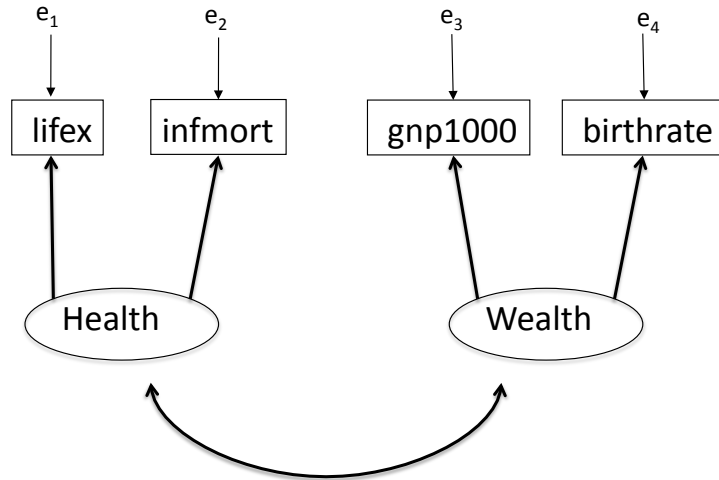
6. Take a look at the path diagram below.
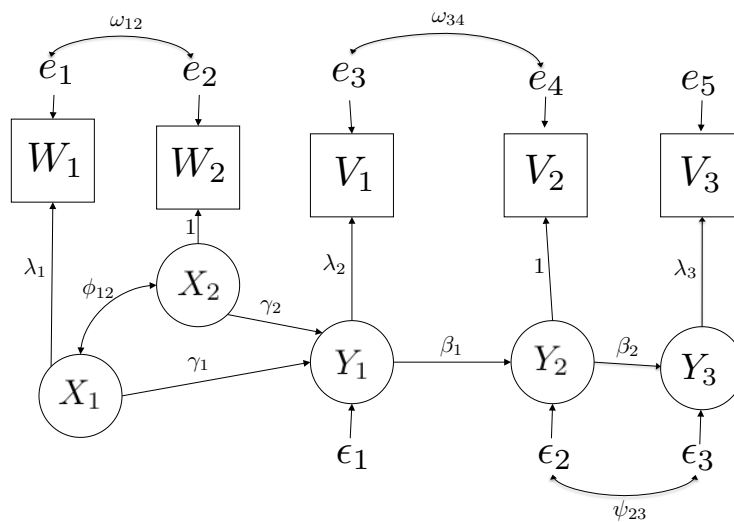
(a) At a glance, do the parameters look identifiable to you? How come? You don't have to prove anything.

(b) Now examine the latent variable model. How does the acyclic rule fail?

7. Make a path diagram with two factors that illustrates both the three-variable rule for unstandardized factors and the crossover rule. Write symbols on the arrows. If an arrow does not have a symbol it means the coefficient equals one. Make the parameters identifiable.

8. In a reaction time study, subjects are seated at a screen. A light flashes on the screen, and they press a key as fast as they can; the time between the light flash and the key press is recorded automatically.

   After some warmup trials, the subjects do the task 50 times, so 50 reaction times are recorded. The 50 times are divided randomly into two sets of 25, and then the median is calculated for each set. In the end, each subject produces two median reaction times.

   The scientists locate sample of university student volunteers whose parents and grandparents are also available to do the experiment. When all the data have been collected, there is a data file with n lines of data. Each line of data has 14 numbers. There are two median reaction times for each of the following individuals:

   - The student
   - Mother
   - Father
   - Maternal grandmother (mother's mother)
   - Maternal grandfather (mother's father)
   - Paternal grandmother (father's mother)
   - Paternal grandfather (father's father)

   As in most applications of statistical methods to real data, your job is to translate this flood of words into a statistical model.

   (a) Make a path diagram. Write symbols on the arrows, making it a surrogate model with identifiable parameters.

   (b) Explain why the parameters are identifiable. Cite the rules you are using by name, letter and number.

9. Make a path diagram that illustrates the acyclic rule, the error-free rule and double measurement. Write symbols on the arrows. If an arrow does not have a symbol it means the coefficient equals one. Keep it fairly simple and make the parameters identifiable.

10. In a study of maternal behaviour in cats, mother cats with new litters of kittens were injected daily with estrogen, a female sex hormone. The cats were randomly assigned to different dosages (amounts) of estrogen. There are lots of different dosages, so dosage may be treated as a continuous variable. Because the exact amount injected is known, the variable Dosage is observed without error.

    After three days, the amount of estrogen in the animal's bloodstream is measured, once. Of course it is measured with error. Then for the next seven days, the following maternal behaviours are recorded.

    - Nursing time in total minutes.

    - Licking in total number of times the cat licked one of her kittens.

    - Retrievals: The mother cat picks up one of her kittens by the skin on the back of its neck, and carries it somewhere.

    (a) Make a path diagram. Write symbols on the arrows, making it a surrogate model with identifiable parameters.

    (b) Explain why the parameters are identifiable. Cite the rules you are using by name, letter and number.

11. This question outlines a different approach to identifying the parameters of a measurement model. Recall that when latent variables are measured with error and the error terms are not correlated, only the variances of Σ are affected by measurement error. The covariances are untouched.

    (a) Just to remind yourself of this fact, calculate Σ for the following model. Independently for $i = 1, \ldots, n$, let

    $$
    \begin{array}{ll}
    D_{i,1} = F_{i,1} + e_{i,1} \\
    D_{i,2} = F_{i,2} + e_{i,2}
    \end{array}
    \quad
    cov\left(\begin{array}{c} F_{i,1} \\ F_{i,2} \end{array}\right) = \left(\begin{array}{cc} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{array}\right)
    \quad
    cov\left(\begin{array}{c} e_{i,1} \\ e_{i,2} \end{array}\right) = \left(\begin{array}{cc} \omega_1 & 0 \\ 0 & \omega_2 \end{array}\right)
    $$

    (b) Based on a random sample of (D_1, D_2) pairs, do we have $\widehat{\Sigma} \xrightarrow{p} \Phi$? Answer Yes or No and briefly justify your answer.

    (c) Denote the reliability of $D_{i,1}$ as a measure of $F_{i,1}$ by r_1, and denote the reliability of $D_{i,2}$ as a measure of $F_{i,2}$ by r_2. Suppose you have good (consistent) estimates of r_1 and r_2 from another source; say $\widehat{r}_1 \xrightarrow{p} r_1$ and $\widehat{r}_2 \xrightarrow{p} r_2$. Give a consistent estimator of Φ. Show your work.

    The point of this question is that sometimes you can use "auxiliary" (outside) information to identify the parameters of a measurement model and rescue the analysis of data that were not collected with latent variables in mind.