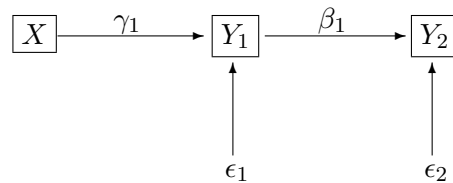


## STA 431s23 Assignment Eight<sup>1</sup>

The non-computer questions on this assignment are practice for the quiz, and will not be handed in. Please bring printouts for the R part of this assignment (Question 9) to the quiz. There may be one or more questions about the printouts, and you may be asked to hand them in with the quiz.

1. In the following model, assume that  $E(X) = \mu_x$ , and the regression equations *do* have intercepts.

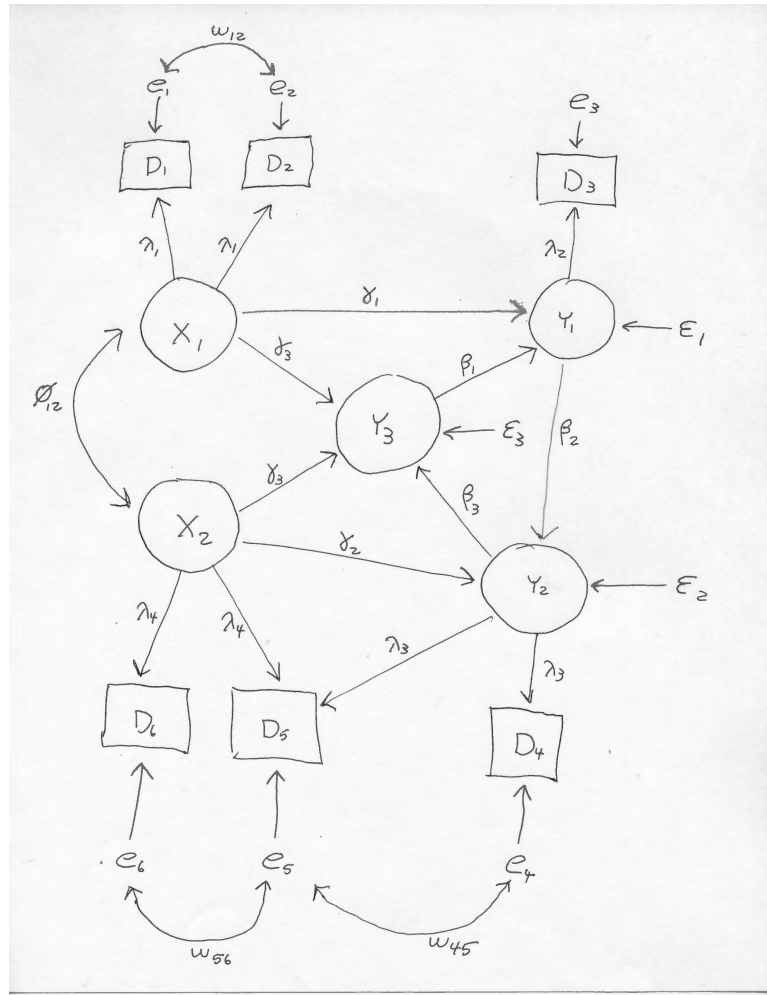


- (a) Classify all the random variables in the model (including error terms) as either Exogenous or Endogenous, and as either Observable or Latent.
- (b) Express the model as a set of equations. Please start by writing “Independently for  $i = 1, \dots, n, \dots$ ” and put a subscript  $i$  on all the random variables. Assume that all the exogenous variables are normal, and include this in the statement of the model. Make up your own symbols for parameters when necessary, but try to stay consistent with the notation being used in the course.
- (c) What is the parameter vector  $\theta$  for this model?
- (d) Write the mean vector  $\mu$  and variance-covariance matrix  $\Sigma$  of an observable data vector  $(X_i, Y_{i,1}, Y_{i,2})$  in terms of the model parameters; show your work. Each element of  $\mu$  and  $\Sigma$  should contain a formula in terms of quantities like  $\phi$ ,  $\beta$  and so on.
- (e) Are the parameters of the model identifiable? Answer Yes or No and prove it. If the answer is No, all you need is a simple numerical example of two distinct parameter vectors that yield the same mean and covariance matrix of the observable data.
- (f) Is this model saturated? Answer Yes or No.
- (g) Suppose that  $X$ ,  $Y_1$  and  $Y_2$  were all latent variables, and there were two independent measurements of each one. Independent means no covariance between any measurement errors. For simplicity, assume no intercepts and  $E(X) = 0$ .
  - i. Draw the path diagram for the new model.
  - ii. Are all the parameters of the new model identifiable? Answer yes or no and explain why. No detailed calculations are needed.

---

<sup>1</sup>This assignment was prepared by Jerry Brunner, Department of Statistical Sciences, University of Toronto. It is licensed under a [Creative Commons Attribution - ShareAlike 3.0 Unported License](https://creativecommons.org/licenses/by-sa/3.0/). Use any part of it as you like and share the result freely. The L<sup>A</sup>T<sub>E</sub>X source code is available from the course website: <http://www.utstat.toronto.edu/~brunner/oldclass/431s23>

2. Patients with high blood pressure are randomly assigned to different dosages of a blood pressure medication. There are lots of different dosages, so dosage may be treated as a continuous variable. Because the exact dosage is known, this variable is observed without error. After one month of taking the medication, the level of the drug in the patient's bloodstream is measured once (with error, of course), by an independent lab. Then, three independent measurements of the patient's blood pressure are taken. One is done by the lab that did the blood test, one is the average of 7 daily measurements taken at home by the patient, and one is done in the doctor's office. Notice that the same lab measures the blood level of the drug, and also does one of the blood pressure measurements. Do *not* assume that errors in the two measurements carried out by the lab are independent. **Make a path diagram. Do not bother to write coefficients on the arrows this time, but write brief labels ("Dose" etc.) in the boxes and ovals.**
3. Consider the following path diagram.



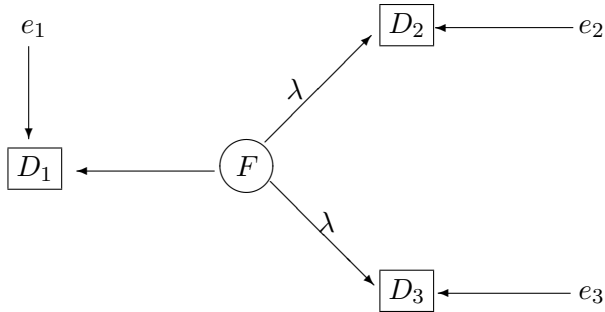
- (a) Give the model equations in scalar form, for a centered model with no intercepts or expected values.

- (b) Give the latent variable model equations in matrix form:  $\mathbf{y}_i = \boldsymbol{\beta}\mathbf{y}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\epsilon}_i$ . Use symbols from the path diagram. For example, you will write  $\boldsymbol{\beta}$  as a  $3 \times 3$  matrix of zeros and  $\beta_j$  symbols.
- (c) Give the measurement model equations in matrix form:  $\mathbf{d}_i = \boldsymbol{\Lambda}\mathbf{F}_i + \mathbf{e}_i$ . Use symbols from the path diagram.
- (d) Give the following matrices:  $\boldsymbol{\Phi}_x, \boldsymbol{\Psi}, \boldsymbol{\Omega}$ . Make sure the dimensions are correct. Some of the symbols you need are on the path diagram, but not all.
4. For the General Structural Equation Model (see formula sheet), calculate
- (a)  $cov(\mathbf{y}_i)$
- (b)  $cov(\mathbf{x}_i, \mathbf{y}_i)$
5. In your calculation of  $cov(\mathbf{y}_i)$  and  $cov(\mathbf{x}_i, \mathbf{y}_i)$ , you used the matrix  $(\mathbf{I} - \boldsymbol{\beta})^{-1}$ . As described in lecture, the existence of this matrix is implied by the model. Assume it does *not* exist. Then the rows of  $(\mathbf{I} - \boldsymbol{\beta})$  are linearly dependent, and there is a  $q \times 1$  vector  $\mathbf{v} \neq \mathbf{0}$  with  $\mathbf{v}^\top(\mathbf{I} - \boldsymbol{\beta}) = \mathbf{0}$ .
- (a) Under this assumption, show  $\mathbf{v}^\top \boldsymbol{\epsilon}_i = -\mathbf{v}^\top \boldsymbol{\Gamma}\mathbf{X}_i$ .
- (b) Show that the last equality contradicts  $\boldsymbol{\Psi}$  positive definite.
6. The following centered model has zero covariance between all pairs of exogenous variables, including error terms.

$$\begin{aligned} Y_1 &= \gamma_1 X + \epsilon_1 \\ Y_2 &= \beta Y_1 + \gamma_2 X + \epsilon_2 \\ W &= X + e_1 \\ V_1 &= Y_1 + e_2 \\ V_2 &= Y_2 + e_3 \end{aligned}$$

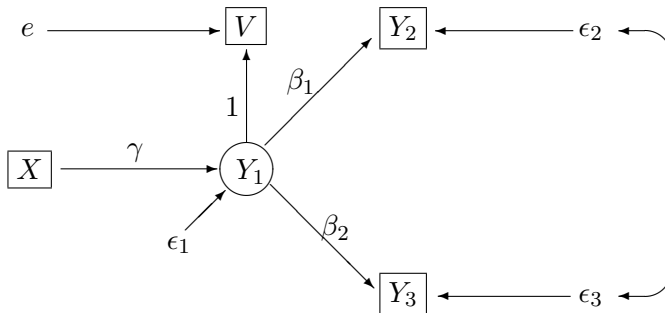
- (a) As the notation suggests, the observable variables are  $W, V_1$  and  $V_2$ . Draw the path diagram. Put a coefficient on each straight arrow that does not come from an error term, either the number one or a Greek letter. It is assumed that all straight arrows coming from error terms have a one.
- (b) Are the parameters of this model identifiable from the covariance matrix? Respond Yes or No and justify your answer.

7. Consider the following model.



- Write the model equations in centered form. Please start by writing “Independently for  $i = 1, \dots, n, \dots$ ” and put a subscript  $i$  on all the random variables.
- Let  $\theta$  denote the vector of parameters that appear in the covariance matrix of the observable data. What is  $\theta$ ?
- Does this model pass the test of the parameter count rule? Answer Yes or No and give the numbers.
- The parameters are identifiable at some points in the parameter space, but not at all points. For what points in the parameter space are the parameters identifiable? Show your work.
- In a test of model fit, what would the degrees of freedom be? The answer is a single number.

8. In the following model, all random variables are normally distributed with expected value zero, and there are no intercepts.



- Write the model equations in scalar form.
- What is the parameter vector  $\theta$  for this model? Use standard notation. Include unknown parameters only.
- Does this model pass the test of the parameter count rule? Answer Yes or No and give both numbers.
- Show that  $\phi$  and  $\gamma$  are identifiable, and then show that  $\beta_1$  and  $\beta_2$  are identifiable provided  $\gamma \neq 0$ . The other parameters are also identifiable in most of the parameter space, but you don't have to do the calculations.

9. Simulate a data set based on the model of Question 8. You choose the true parameter values and you choose the sample size, but make it large. Fit the model with `lavaan`, and verify that you are estimating the parameters accurately. For the correlated error terms, you may find my `rmvn` function helpful. You can access it with

```
source("https://www.utstat.toronto.edu/brunner/openSEM/fun/rmvn.txt")
```

Remember, *Question 9 is not a group project*. You are required to do the work yourself. You may discuss it with your classmates, but do not look at anyone else's code, and do not show anyone else your code.

**Please bring your printout from Question 9 to the quiz. You may be asked for numbers from your printout, and you may be asked to hand it in as part of the quiz.**