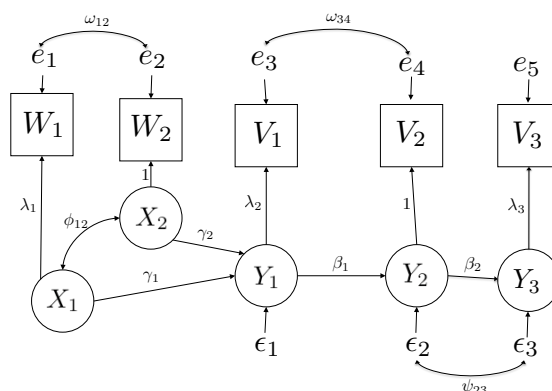


## STA 431s11 Assignment 12

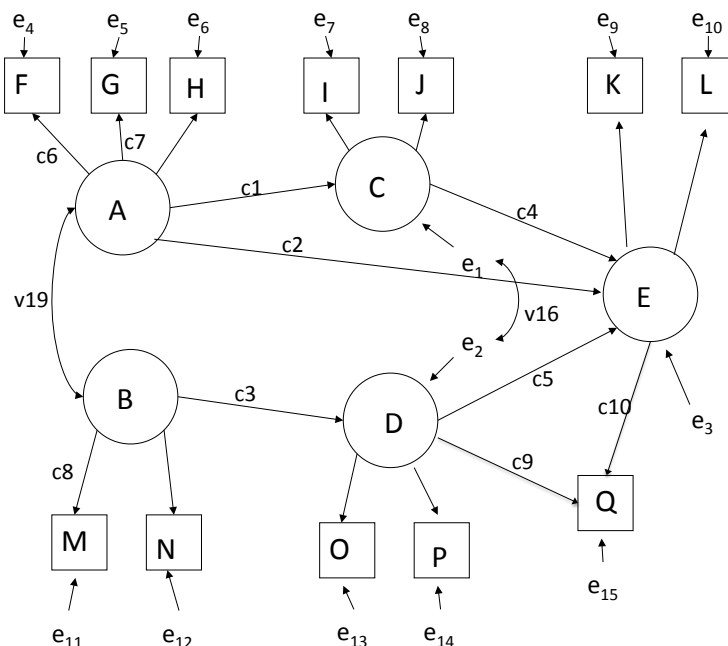
In addition to reviewing past assignments, please do this assignment in preparation for the final examination. You will *not* bring any printouts from the SAS part to the final exam. Instead, you will use my printouts.

1. Consider the following path diagram.



- (a) Write down the model equations in scalar form.
  - (b) Give the matrix  $\beta$ , using symbols from the path diagram. Make sure it has the correct dimensions.
  - (c) Give the matrix  $\Gamma$ , using symbols from the path diagram. Make sure it has the correct dimensions.
  - (d) Give the matrix  $\Lambda$ , using symbols from the path diagram. Make sure it has the correct dimensions.
  - (e) Give the matrix  $\Phi_{11}$ , using symbols from the path diagram. Make sure it has the correct dimensions.
  - (f) Give the matrix  $\Psi$ , using symbols from the path diagram. Make sure it has the correct dimensions.
  - (g) Give the matrix  $\Omega$ , using symbols from the path diagram. Make sure it has the correct dimensions.
  - (h) In what specific way does the latent variable part of this model *not* fit the acyclic rule?
  - (i) How do you know for sure (and fast) that the parameters of this model are not identifiable at all points in the parameter space?
2. For the measurement part of the general two-stage structural equation model  $\mathbf{Y} = \beta\mathbf{Y} + \Gamma\mathbf{X} + \epsilon$ , calculate the following in matrix form:
    - (a)  $V(\mathbf{Y})$
    - (b)  $C(\mathbf{X}, \mathbf{Y})$

3. 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.



- Write down the model equations in scalar form.
- Give the matrix  $\beta$ , using symbols from the path diagram. Make sure it has the correct dimensions.
- Give the matrix  $\Gamma$ , using symbols from the path diagram. Make sure it has the correct dimensions.
- Give the matrix  $\Lambda$ , using symbols from the path diagram. Make sure it has the correct dimensions.
- Give the matrix  $\Phi_{11}$ , using symbols from the path diagram. Make sure it has the correct dimensions.
- Give the matrix  $\Psi$ , using symbols from the path diagram. Make sure it has the correct dimensions.
- Give the matrix  $\Omega$ , using symbols from the path diagram. Make sure it has the correct dimensions.
- Explain why the model parameters are all identifiable, making specific reference to the identifiability rules on the reference sheet.

4. Let

$$\begin{aligned}D_1 &= \lambda_1 F_1 + e_1 \\D_2 &= \lambda_2 F_1 + e_2 \\D_3 &= \lambda_3 F_1 + e_3 \\D_4 &= \lambda_4 F_2 + e_4 \\D_5 &= \lambda_5 F_2 + e_5 \\D_6 &= \lambda_6 F_2 + e_6 \\D_7 &= \lambda_7 F_3 + e_7 \\D_8 &= \lambda_8 F_3 + e_8 \\D_9 &= \lambda_9 F_3 + e_9,\end{aligned}$$

where all expected values are zero,  $V(e_i) = \omega_i$  for  $i = 1, \dots, 9$ , the error terms are independent of each other and of the factors,

$$V \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{12} & \phi_{22} & \phi_{23} \\ \phi_{13} & \phi_{23} & \phi_{33} \end{bmatrix},$$

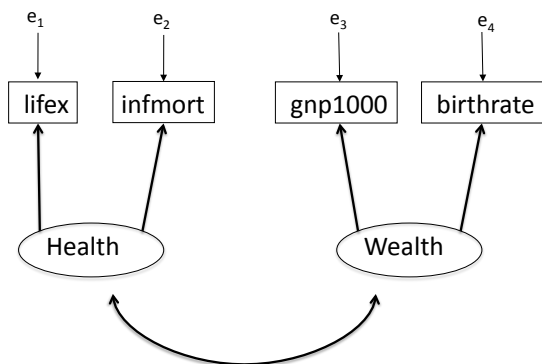
and all the factor loadings are non-zero with  $\lambda_1 > 0$ ,  $\lambda_4 > 0$  and  $\lambda_7 > 0$ . This is the true model. Do not re-parameterize until you are asked to do so.

- (a) Draw the path diagram. Put coefficients on the arrows where necessary. For arrows unmarked by a coefficient, it will be assumed that the coefficient equals one.
- (b) Calculate  $\Sigma$ . This is a scalar calculation.
- (c) Show that for any set of parameter values, there are infinitely many other sets of parameter values that yield the same covariance matrix (hence the parameters are not identifiable).
- (d) Show that the parameter  $\omega_1$  is identifiable.
- (e) Show that the ratio of factor loadings  $\lambda_5/\lambda_4$  is identifiable.
- (f) Show that the function  $g(\boldsymbol{\theta}) = \text{Corr}(F_2, F_3)$  is identifiable. Hint: Temporarily re-parameterize by standardizing the factors. Solve for  $\phi_{23}$  in terms of  $\sigma_{ij}$  quantities, and then try that same function with the original (true) model. If the answer is the correlation, you've got it.

5. The SAS part of the final exam will be based upon the Poverty data: [poverty.data](#). This data set contains information from a sample of 97 countries. The variables include Live birth rate per 1,000 of population, Death rate per 1,000 of population, Infant deaths per 1,000 of population under 1 year old, Life expectancy at birth for males, Life expectancy at birth for females, and Gross National Product per capita in U.S. dollars. Here is my SAS data step. Please use it in your program so we have the same variable names and you can avoid some potential problems.

```
data misery;
  infile 'poverty.data';
  input birthrate deathrate infmort lifexM lifexF gnp group country $;
  gnp1000 = gnp/1000; /* In thousands of dollars */
  lifex = (lifexM+lifexF)/2;
```

This can be a very challenging and frustrating data set to work with, because correlated measurement errors produce negative variance estimates almost everywhere you turn. To make your job easier for the final, I did the initial exploration and came up with a few reasonable models with just four observable variables: `lifex`, `infmort`, `gnp1000` and `birthrate`. Here is a picture of a factor analysis model with 2 factors.



The reason for making birth rate an indicator of wealth is that birth control costs money.

- Fit the model with SAS. You will have to re-parameterize. Which of the two standard re-parameterizations should you choose? Suppose we are interested in the correlation between Health and Wealth.
- What is the parameter  $\theta$  for this model? Give your answer in the form of a list of names from your SAS job.

- (c) Prove that the re-parameterization you have chosen results in a parameter vector that is identifiable – at least, it's identifiable in most of the parameter space. This is similar to a calculation done in lecture.
- (d) Does this model fit the data adequately? Answer Yes or No, and back up your answer with two numbers from the printout: The value of a test statistic, and a  $p$ -value.
- (e) What is the maximum likelihood estimate of the correlation between factors? The answer is a single number from the printout.
- (f) Now try a latent regression model. Which latent variable should be  $X$ , and which should be  $Y$ ? I used a different re-parameterization for this part.
- (g) Compare the test for goodness of fit to the one for the factor analysis model.
- (h) Be able to interpret the  $Z$  tests for the  $\gamma$  and  $\lambda$  values. Do the *signs* of the estimates (positive and negative) make sense?
- (i) Finally, the high estimated correlation between factors from the first part of this question suggests that there might be just one underlying factor: wealth. Try a single-factor model and see if it fits. Do the estimated factor loadings make sense?