STA 431s09 Assignment 7

Do this assignment in preparation for the quiz on Friday, March 6th. Except for the last part of Question 5, the questions are practice for the quiz and are not to be handed in.

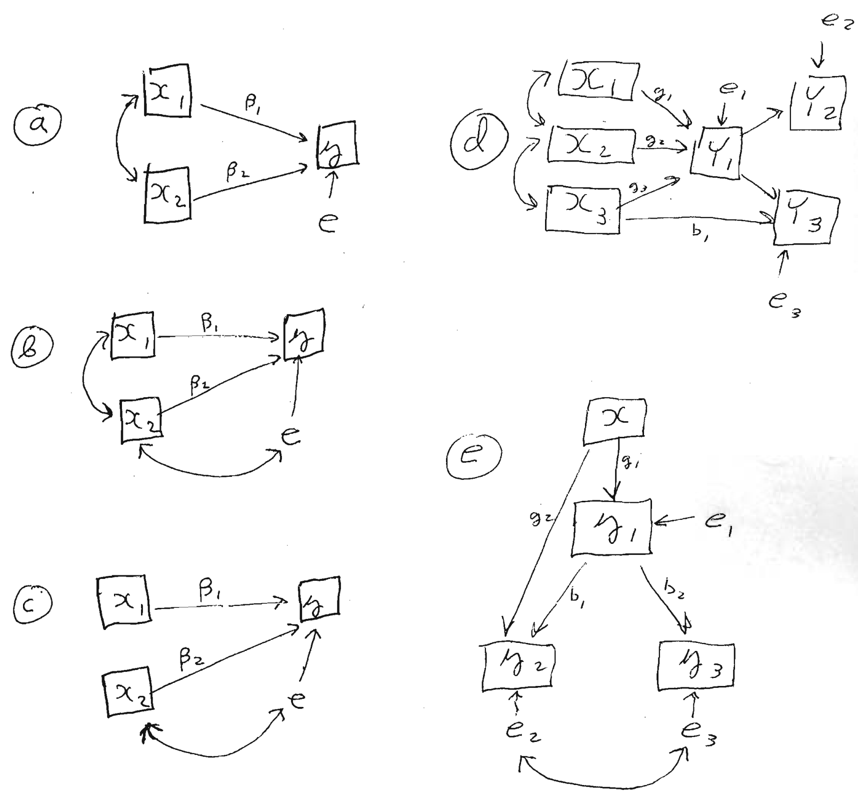
1. Here are path diagrams for some observed variable models; all expected values are zero. For each one, state whether the model parameters can be identified from the covariance matrix, and prove it. In some cases, the "proof" consists of just citing the Regression rule or the Recursion rule. If these rules do *not* apply, you will need to

   (a) Write the model equations, specifying some notation for the variances and covariances of the error terms and exogenous variables.

   (b) Calculate the covariance matrix $\mathbf{\Sigma}$ of the observed variables in terms of the model parameters.

   (c) Either

   - Show it is possible to solve for the model parameters in terms of the $\sigma_{i,j}$ elements of the covariance matrix (proving the model is identified), or
   - Show that a unique solution is impossible by giving an example of two different parameter vectors that yield the same $\mathbf{\Sigma}$; a simple numerical example is best.



2. Here is a recursive model for observed variables with non-zero expected values and intercepts. Independently for $i = 1, \ldots, n$, let

$$\begin{aligned} Y_{i,1} &= \alpha_1 + \gamma_1 X_i + \zeta_{i,1} \\ Y_{i,2} &= \alpha_2 + \beta Y_{i,1} + \gamma_2 X_i + \zeta_{i,1}, \end{aligned}$$

where $X_i \sim N(\kappa, \phi)$, $\zeta_{i,1} \sim N(0, \psi_1)$, $\zeta_{i,2} \sim N(0, \psi_2)$. The parameter $\boldsymbol{\theta}$ consists of $\alpha_1, \alpha_2, \gamma_1, \gamma_2, \beta, \phi, \psi_1$ and ψ_2.

(a) Write

$$\boldsymbol{\mu} = E \left[\begin{array}{c} X_i \\ Y_{i,1} \\ Y_{i,2} \end{array} \right] \text{ and } \boldsymbol{\Sigma} = V \left[\begin{array}{c} X_i \\ Y_{i,1} \\ Y_{i,2} \end{array} \right].$$

in terms of the elements of $\boldsymbol{\theta}$.

(b) Prove that the model is identified by solving for the elements of $\boldsymbol{\theta}$ in terms of the elements of

$$\boldsymbol{\mu} = \left[\begin{array}{c} \mu_1 \\ \mu_2 \\ \mu_3 \end{array} \right] \text{ and } \boldsymbol{\Sigma} = \left[\begin{array}{ccc} \sigma_{1,1} & \sigma_{1,2} & \sigma_{1,3} \\ & \sigma_{2,2} & \sigma_{2,3} \\ & & \sigma_{3,3} \end{array} \right],$$

or else show that the model is not identified by giving an example of two different sets of parameter values that yield the same pair $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. A simple numerical example is best for proving non-identification.

3. Here is a multivariate regression model with intercepts and non-zero expected values. Independently for $i = 1, \ldots, n$,

$$\mathbf{Y}_i = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{X}_i + \boldsymbol{\zeta}_i,$$

where

$\mathbf{Y}_i$ is an $m \times 1$ random vector of observed dependent variables; there are m dependent variables.

$\boldsymbol{\alpha}$ is an $m \times 1$ matrix of unknown constants. These are the intercepts.

$\mathbf{X}_i$ is a $p \times 1$ observable random vector; there are p independent variables. $E(\mathbf{X}_i) = \boldsymbol{\kappa}$, an $m \times 1$ matrix of unknown constants. $V(\mathbf{X}_i) = \boldsymbol{\Phi}$, a $p \times p$ symmetric and positive definite matrix of unknown constants.

$\boldsymbol{\Gamma}$ is an $m \times p$ matrix of unknown constants. These are the regression coefficients, with one row for each dependent variable and one column for each independent variable.

$\boldsymbol{\zeta}_i$ is the error term of the latent regression. It is an $m \times 1$ random vector with expected value zero and variance-covariance matrix $\boldsymbol{\Psi}$, an $m \times m$ symmetric and positive definite matrix of unknown constants. $\boldsymbol{\zeta}_i$ is independent of $\mathbf{X}_i$.

To make things simple, we will assume that all the random variables are multivariate normal.

(a) What is the parameter $\boldsymbol{\theta}$ for this model?

(b) Write

$$\boldsymbol{\mu} = E\left[\begin{array}{c} \mathbf{X}_i \\ \mathbf{Y}_i \end{array}\right] \text{ and } \boldsymbol{\Sigma} = V\left[\begin{array}{c} \mathbf{X}_i \\ \mathbf{Y}_i \end{array}\right].$$

in terms of the elements of $\boldsymbol{\theta}$. Your answers are partitioned matrices.

(c) Prove that the model is identified by solving for the elements of $\boldsymbol{\theta}$ in terms of the elements of

$$\boldsymbol{\mu} = \left[\begin{array}{c} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{array}\right] \text{ and } \boldsymbol{\Sigma} = \left[\begin{array}{cc} \boldsymbol{\Sigma}_{1,1} & \boldsymbol{\Sigma}_{1,2} \\ \boldsymbol{\Sigma}_{1,2}' & \boldsymbol{\Sigma}_{2,2} \end{array}\right],$$

or else show that the model is not identified by giving an example of two different sets of parameter values that yield the same pair $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. A simple numerical example is best for proving non-identification.

4. Given the answer to the previous question, do you think that recursive models with intercepts are identified in general (provided the exogenous variables are independent of the error terms)? You are no asked for a proof. Just answer Yes or No and briefly state why you think so.

5. The file `org1.data` contains data from a study of workers' commitment to the companies they work for (there is a link from the course web page in case the one in this document does not work). Except for the last variable, the variables are mean response to several questions on a seven-point scale.

   `OCommit`: Commitment to the organization

   `RelColl`: Good relationships with colleagues at work

   `RelMan`: Good relationship with immediate supervisor

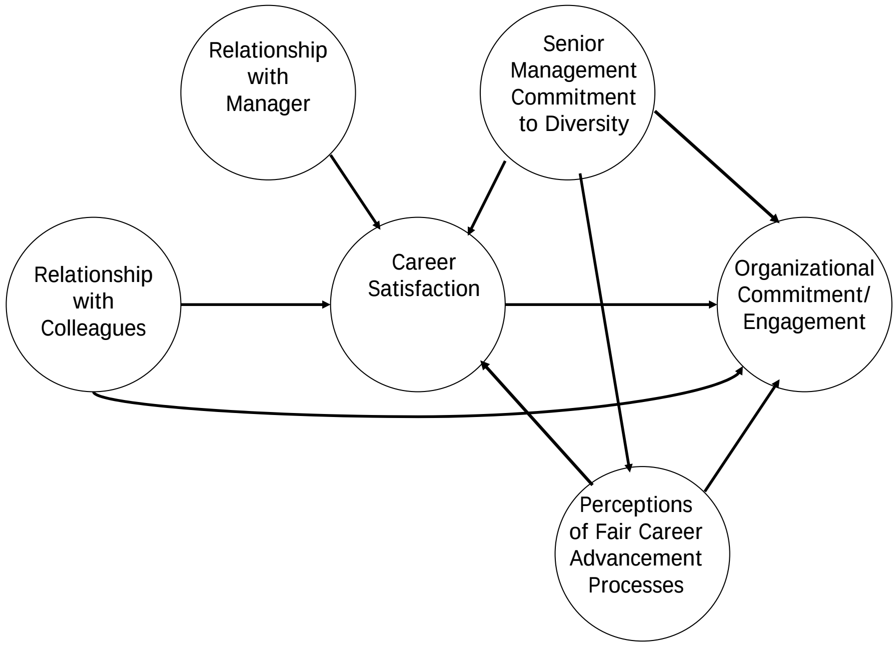
   `FairCar`: Perceived fair chance for career advancement

   `CarSat`: Career satisfaction

   `SMComDiv`: Perceived commitment to diversity on the part of senior management

   `Gender`: 1=Female, 0=Male. We will not be using this variable yet.

On the next page is a path diagram proposed by the client (this is from a statistical consulting job). All variables are observed, even though they are enclosed by circles. The error terms are not shown, but of course every endogenous variable has one; assume they are independent. Also, the client forgot to put in covariances among the exogenous variables, but readily agreed they should be in the model.



(a) Is this model identified? Answer Yes or No and in one sentence, say how you know.

(b) Fit the model using SAS `proc calis`. Locate the maximum likelihood estimates and the tests for whether each parameter equals zero. We are mostly interested in the parameters associated with straight arrows. Also look at the chisquare test for goodness of fit and the independence model chisquare. For each test, what is the numerical value of the test statistic? The answer is a number from the printout. What (if anything) do you conclude? You might be asked questions like these on the quiz, and you might have to hand in your log and list files. Oh yes, what is the sample size?

You will notice that the data file has a useful header with the variable names. As far as I know, SAS cannot use this information; variable names are specified in the `input` statement of your program. So you can delete the first line of the data file, or you can instruct SAS to skip it. I skipped it with

```
infile 'org1.data' firstobs=2;
```