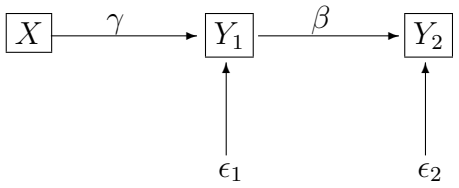
STA 431s13 Assignment Eight[1]

There is no SAS assignment this time. The questions are just practice for the quiz on Friday March 15, and are not to be handed in.

1. Make a path diagram for the pig data of Assignment 7.
2. 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 as either Exogenous or Endogenous, and as either Manifest or Latent.

   (b) Express the model as a set of equations. Please start by writing "Independently for $i = 1, \ldots, n, \ldots$" and put a subscript i on all the random variables. Assume that everything is normal, and include this in your 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 $\boldsymbol{\theta}$ for this model?

   (d) What is the joint distribution of the manifest variables? Express the mean vector and variance-covariance matrix in terms of the model parameters; show your work. Each element of the mean vector and the variance-covariance matrix should contain a formula in terms of quantities like ϕ, β 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.

[1] Copyright information is at the end of the last page.

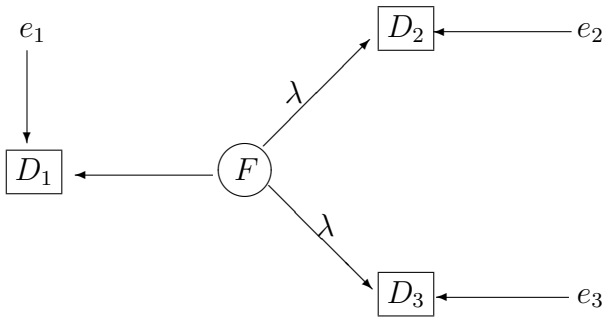
3. The following model has zero covariance between all pairs of exogenous variables.

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

   (a) 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) As the notation suggests, the observable variables are W, V_1 and V_2. Are the parameters of this model identifiable at every point in the parameter space? Respond Yes or No and justify your answer. You may assume $E(X) = 0$.

4. In the following model, all expected values are zero.



   (a) Write this model as a set of simultaneous equations. Please start by writing "Independently for $i = 1, \ldots, n, \ldots$" and put a subscript i on all the random variables. You may assume that everything is normal

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

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

   (d) 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.

   (e) In a test of model fit, what would the degrees of freedom be? The answer is a single number.

5. Make a path diagram for the model indicated below.

   A farm co-operative (co-op) is an association of farmers. The co-op can buy fertilizer and other suppies in large quantities for a lower price, it often provides a common storage location for harvested crops, and it arranges sale of farm products in large quantities to grocery store chains and other food suppliers. Farm co-ops usually have professional managers, and some do a better job than others.

   We have data from a study of farm co-op managers. The response variable of interest is job performance, a latent variable. The variables in the "latent variable" part of the model are the following, but note that one of them is assumed observable.

   X_1: Knowledge of business principles and products (economics, fertilizers and chemicals). This is a latent variable measured by W_{11} and W_{12}.

   X_2: Profit-loss orientation ("Tendency to rationally evaluate means to an economic end"). This is a latent variable measured by W_{21} and W_{22}.

   X_3: Job satisfaction. This is a latent variable measured by W_{31} and W_{32}.

   X_4: Formal Education = Number of years of formal schooling divided by 6. This is an observable variable, assumed to be measured without error.

   Y: Job performance. This is a latent variable measured by V_1 and V_2.

   The data file has these observable variables in addition to an identification code for the managers.

   W_{11}: Knowledge measurement 1

   W_{12}: Knowledge measurement 2

   W_{21}: Profit-Loss Orientation 1

   W_{22}: Profit-Loss Orientation 2

   W_{31}: Job Satisfaction 1

   W_{32}: Job Satisfaction 2

   X_4: Formal education, assumed measured without error

   V_1: Job Performance 1

   V_2: Job Performance 2

   In this study, the double measurements are obtained by just splitting questionnaires in two, as in split half reliability. Furthermore, all the measurement errors are assumed independent of one another. This is consistent with mainstream psychometric theory, though maybe not with common sense. For this assignment, please assume that the errors are independent of one another, and independent of the explanatory variables. The explanatory variables, of course, should *not* be assumed independent of one another.
