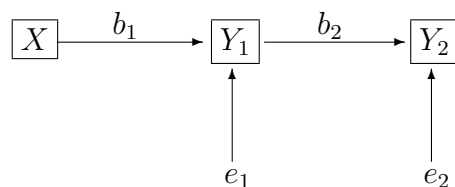


## STA 431s09 Assignment 6

Do this assignment in preparation for the quiz on Friday, Feb. 27th. Answers to the non-SAS questions are practice for the the quiz, and are not to be handed in. For Question 6, bring both your log file and your list file to the quiz; they may (or may not) be handed in.

1. Let  $X_1, \dots, X_n$  be a random sample from a normal distribution with mean  $\theta_1$  and variance  $\theta_2 + \theta_3$ , where  $-\infty < \theta_1 < \infty$ ,  $\theta_2 > 0$  and  $\theta_3 > 0$ . Is this model identified? Answer Yes or No and prove your answer.
2. Let  $X_1, \dots, X_n$  be a random sample from a normal distribution with mean  $\theta$  and variance  $\theta^2$ , where  $-\infty < \theta < \infty$ . Is this model identified? Answer Yes or No and justify your answer. You need not show the actual calculation.
3. Consider this model:



- (a) 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 everything is normal with mean zero, and give the variances and covariances of all the exogenous variables in the model (including error terms). Make up your own symbols for parameters when necessary.
- (b) What is the parameter  $\theta$ ? It has 5 elements.
- (c) Is this model saturated? Answer Yes or No.
- (d) Classify the random variables in a two-by-two table as either Exogenous or Endogenous, and either Manifest or Latent.
- (e) What is the joint distribution of the manifest variables? Express the variance-covariance matrix in terms of the parameter  $\theta$ ; show your work. Each cell in your matrix should contain a formula for the variance or covariance in terms of quantities like  $\sigma_x^2$ ,  $b_1$  and so on.
- (f) Is this model identified? Answer Yes or No and prove it.

4. Here is a multivariate regression model with no intercept (as usual, the data have been “centered”). Independently for  $i = 1, \dots, n$ ,

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

where

$\mathbf{Y}_i$  is an  $m \times 1$  random vector of observable dependent variables, so the regression can be multivariate; there are  $m$  dependent variables.

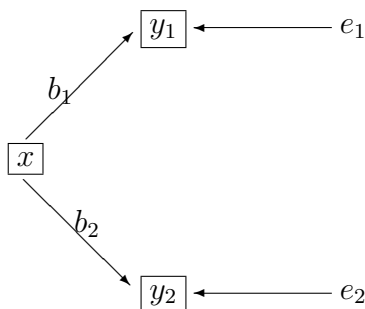
$\mathbf{X}_i$  is a  $p \times 1$  observable random vector; there are  $p$  independent variables.  $\mathbf{X}_i$  has expected value zero and variance-covariance matrix  $\boldsymbol{\Phi}$ , a  $p \times p$  symmetric and positive definite matrix of unknown constants.

$\mathbf{\Gamma}$  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  $\mathbf{X}_i$ .

Is this model identified? Answer Yes or No and show your work.

5. Consider the following model:



- Write this model as a set of simultaneous equations. Please start by writing “Independently for  $i = 1, \dots, n, \dots$ ” and put a subscript  $i$  on all the random variables.
- Make up symbols for all the non-zero variances and covariances in the model.
- What is the parameter vector  $\boldsymbol{\theta}$  for this model? It has five elements.
- Fit this model into the framework of Question 4 by giving the matrices  $\mathbf{X}_i$ ,  $\mathbf{Y}_i$ ,  $\boldsymbol{\zeta}_i$ ,  $\mathbf{\Gamma}$ ,  $\boldsymbol{\Phi}$  and  $\boldsymbol{\Psi}$  in terms of the notation of this question.
- Is the model of this question identified? Answer Yes or No. Do you have to show your work? Why or why not?

- Using the data file `wishbone.data` (there is a link from the course web page in case the one in this document does not work), fit the model of Question 5 using SAS `proc calis`. Locate the maximum likelihood estimates and the tests for whether each parameter equals zero. 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 on the quiz.