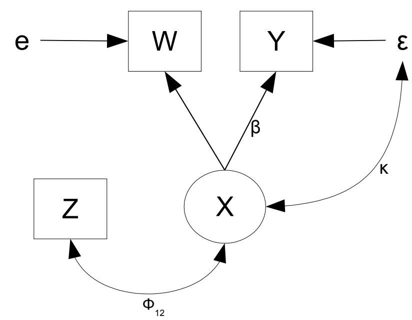
STA 2101f19 Assignment Eight[1]

1. In the model pictured below, the explanatory variable X is measured with error as well as being correlated with omitted variables. Z is an instrumental variable.



   (a) Give the model equations without intercepts. Don't mention the expected values.

   (b) Guided by the symbols on the path diagram, provide notation for the variances and covariances of the error terms and exogenous variables.

   (c) Let $\boldsymbol{\theta}$ denote the vector of parameters you have witten down so far. These are the parameters that will appear in the covariance matrix of the observable data. What is $\boldsymbol{\theta}$?

   (d) Does this model pass the test of the Parameter Count Rule? Answer Yes or No and give the numbers. (Notice that we are only trying to identify the parameters in $\boldsymbol{\theta}$, which is a function of the full parameter vector. The full parameter vector has intercepts and unknown probability distributions.)

   (e) Calculate the covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{D_i}$, a single observable data vector.

   (f) Is the parameter β identifiable provided $\phi_{12} \neq 0$? Answer Yes or No. If the answer is Yes, prove it. If the answer is No, give a simple numerical example of two parameter vectors with different β values, yielding the same covariance matrix $\boldsymbol{\Sigma}$.

   (g) Why is is reasonable to assume $\phi_{12} \neq 0$?

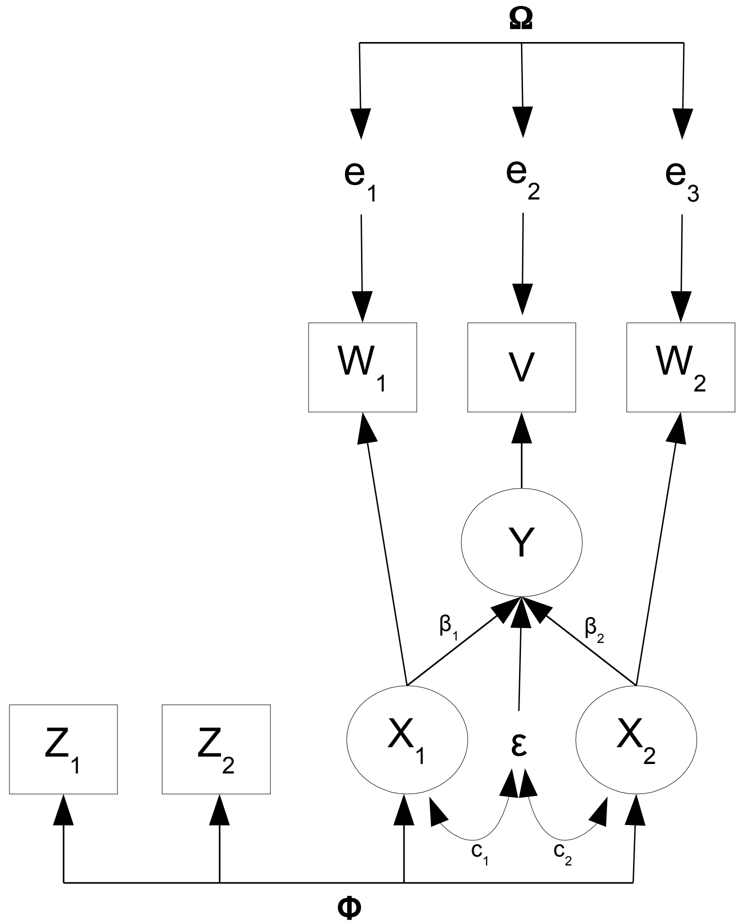
   (h) Now let's make the model more realistic and scary. The response variable is measured with error, so $V = Y + e_2$. Furthermore, because of omitted variables, all the error terms might be correlated with one another and with X.

      i. Do your best to make a path diagram of the new model. You need not write symbols on the curved double-headed arrows you have added.

      ii. Show that β is still identifiable.

[1] 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. Use any part of it as you like and share the result freely. The LaTeX source code is available from the course website: http://www.utstat.toronto.edu/~brunner/oldclass/2101f19

2. Here is a model with two explanatory variables and two instrumental variables. The path diagram looks busy, but it has features that make sense once you think about them. The instrumental and explanatory variables have covariance matrix $\boldsymbol{\Phi} = [\phi_{ij}]$, so that for example $Var(X_1) = \phi_{33}$. No doubt there are omitted explanatory variables that are correlated with X_1 and X_2, and affect Y. That is the source of $c_1 = Cov(X_1, \epsilon)$ and $c_2 = Cov(X_2, \epsilon)$. The variables in the latent regression model are all measured (once) with error. Because of omitted variables in the measurement process, the measurement errors are correlated, with 3×3 covariance matrix $\boldsymbol{\Omega} = [\omega_{ij}]$.



(a) Give the model equations without intercepts. Don't mention the expected values.

(b) How many parameters appear in the covariance matrix of the observable data? Scanning from the bottom, I get 10+2+1+2+6=21.

(c) Does this model pass the test of the Parameter Count Rule? Answer Yes or No and give the numbers.

(d) The next step would be to calculate the covariance matrix $\boldsymbol{\Sigma}$ of the observable data vector, but that's a big job. To save work and also to reveal the essential features of the problem, please just calculate $cov\left((Z_1, Z_2)^\top, (W_1, W_2, V)^\top\right)$.

(e) Are the parameters β_1 and β_2 identifiable? Answer Yes or No. If the answer is Yes, prove it. You don't have to finish solving for β_1 and β_2. You can stop once you have two linear equations in two unknowns, where the coefficients are σ_{ij} quantities. Presumably it's possible to solve two linear equations in two unknowns. To prove identifiability, you don't have to actually recover the parameters from the covariance matrix. All you have to do is show it can be done. In $\boldsymbol{\Sigma}$, please maintain the order Z_1, Z_2, W_1, W_2, V so we will have the same answer.

3. Here is a matrix version of instrumental variables. Independently for $i = 1, \ldots, n$, the model equations are

$$
\begin{aligned}
\mathbf{Y}_i &= \boldsymbol{\beta}\mathbf{X}_i + \boldsymbol{\epsilon}_i \\
\mathbf{W}_i &= \mathbf{X}_i + \mathbf{e}_{i,1} \\
\mathbf{V}_i &= \mathbf{Y}_i + \mathbf{e}_{i,2}.
\end{aligned}
$$

The random vectors $\mathbf{X}_i$ and $\mathbf{Y}_i$ are latent, while $\mathbf{W}_i$ and $\mathbf{V}_i$ are observable. In addition, there is a vector of observable instrumental variables $\mathbf{Z}_i$. The random vectors $\mathbf{X}_i$ and $\mathbf{Z}_i$ are $p \times 1$, while $\mathbf{Y}_i$ is $q \times 1$. This determines the sizes of all the matrices. The variances and covariances are as follows: $cov(\mathbf{X}_i) = \boldsymbol{\Phi}_x$, $cov(\mathbf{Z}_i) = \boldsymbol{\Phi}_z$, $cov(\mathbf{Z}_i, \mathbf{X}_i) = \boldsymbol{\Phi}_{zx}$, $cov(\boldsymbol{\epsilon}_i) = \boldsymbol{\Psi}$, $cov(\mathbf{X}_i, \boldsymbol{\epsilon}_i) = \mathbf{C}$, and $cov\left(\begin{array}{c}\mathbf{e}_{i,1} \\ \mathbf{e}_{i,2}\end{array}\right) = \boldsymbol{\Omega}$. All variance-covariance matrices are positive definite (why not), and in addition, the $p \times p$ matrix of covariances $\boldsymbol{\Phi}_{zx}$ has an inverse. Covariances that are not specified are zero; in particular, the instrumental variables have zero covariance with the error terms.

Collecting $\mathbf{Z}_i$, $\mathbf{W}_i$, $\mathbf{V}_i$ into a single long data vector $\mathbf{D}_i$, we write its variance-covariance matrix as a partitioned matrix:

$$
\boldsymbol{\Sigma} = \left(\begin{array}{c|c|c}
\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} \\ \hline
 & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23} \\ \hline
 & & \boldsymbol{\Sigma}_{33}
\end{array}\right),
$$

where $cov(\mathbf{Z}_i, \mathbf{W}_i) = \boldsymbol{\Sigma}_{12}$, and so on.

(a) Give the dimensions (number of rows and number of columns) of the following matrices: $\boldsymbol{\beta}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Omega}$, $\boldsymbol{\Sigma}_{23}$.

(b) This problem fails the test of the Parameter Count Rule, though you are not required to show it. Fortunately, all we care about is $\boldsymbol{\beta}$. Doing as little work as possible, prove that $\boldsymbol{\beta}$ is identifiable by showing how it can be recovered from the $\boldsymbol{\Sigma}_{ij}$ matrices.

(c) Give the formula for an estimator of $\boldsymbol{\beta}$ and show that it is consistent.

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 $\mathbf{v}^\top(\mathbf{I} - \boldsymbol{\beta}) = \mathbf{0}$. Under this assumption, show $\mathbf{v}^\top \boldsymbol{\Psi} \mathbf{v} = 0$, contradicting $\boldsymbol{\Psi}$ positive definite.

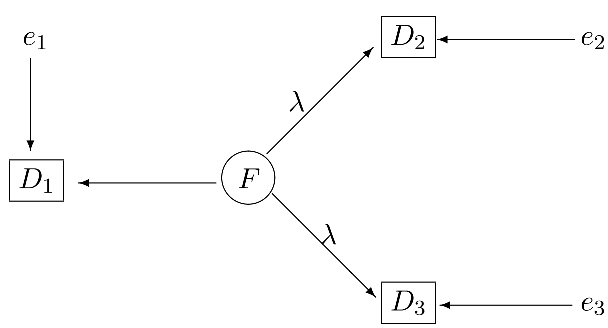
6. The following model has zero covariance between all pairs of exogenous variables, including error terms.

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

   (a) 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 from the covariance matrix? Respond Yes or No and justify your answer.

7. Consider the following model.



   (a) Write the model equations without intercepts. Don't mention the expected values. Please start by writing "Independently for $i = 1, \ldots, n$, ..." and put a subscript i on all the random variables.

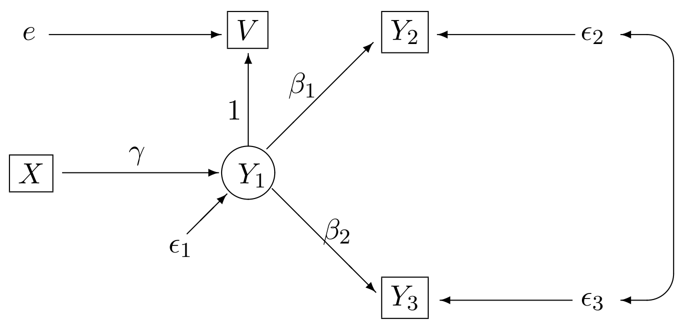
   (b) Let $\boldsymbol{\theta}$ denote the vector of parameters that appear in the covariance matrix of the observable data. What is $\boldsymbol{\theta}$?

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

(d) Are the elements of $\boldsymbol{\theta}$ identifiable from the covariance matrix? 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 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.

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



(a) Write the model equations in scalar form.

(b) What is the parameter vector $\boldsymbol{\theta}$ for this model? Use standard notation. Include unknown parameters in the covariance matrix only.

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

(d) It's a bit time-consuming to write $\boldsymbol{\Sigma} = cov(X, V, Y_2, Y_3)^\top$, but it's worth it. Please do so.

(e) Verify that all the parameters are identifiable at points in the parameter space where $\gamma \neq 0$.

(f) Even where $\gamma = 0$, you can tell whether β_1 and β_2 are zero, and if they are non-zero, you can identify the sign (a function of $\boldsymbol{\theta}$). Do you agree?

(g) Using the parameter count rule, there should be one model-induced equality constraint on the $\sigma_{i,j}$ quantities. Provided that γ, β_1 and β_2 are all non-zero, I can see what it is. What is the equality constraint?