STA 313F2004 Final Assignment

In this assignment, all random variables and vectors have expected value zero. Scalar (that is, 1×1) random variable starting with x and y are always manifest, those starting with F are latent, and those starting with e are error terms. In path diagrams, every curved two-headed arrow corresponds to a covariance term, but the symbols for the covariances are not on the path diagram; you make them up. On the other hand, if a straight arrow has no symbol on it, the weight is one. For example, error terms never have symbols on their arrows, and their weights are always one. Unless otherwise indicated, all exogenous variables have unknown variances that are parameters in the model; you make up the symbols.

Random vectors will be represented in boldface, and we will stay close to the LISREL notation; see formula sheet.

1. Give the definition of a *recursive* model for observed variables. There are three characteristics. On the final exam, if you state that a model is identified because it's recursive (and you're right), you must mention all three characteristics for full marks.

2. Prove the null-$\boldsymbol{\beta}$ rule for observed variables. That is (considering a special case of the LISREL model),

$$\mathbf{y} = \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta},$$

   with $V(\mathbf{x}) = \boldsymbol{\Phi}$, $V(\boldsymbol{\zeta}) = \boldsymbol{\Psi}$, and $\mathbf{x}$ and $\boldsymbol{\zeta}$ independent. Of course $\mathbf{x}$ and $\mathbf{y}$ are manifest. Is this model identified? Answer Yes or No, and prove it.

   By the way, if a model expressed in terms of vectors and matrices is *not* identified in general, a good approach is often to find a small special case, usually involving scalars rather than matrices, and either (a) observe that showing identification involves trying to several equations in several unknowns, with more unknowns than equations, or (b) give a specific example of two distinct sets of parameter values that yield the same variance-covariance matrix for the manifest variables – in your small example.

3. In class, we saw that if you are doing multiple regression with errors of measurement in the independent variables and you have two independent measurements of the independent variables, the model is identified. As we did it in class, the two independent measurements were *equivalent*. That is, the measurement errors had the same variance-covariance matrix each time. It would be nice if we could relax this equivalence assumption. Then, we could deal with situations where the independent variables were measured by a different person the second time, which would help ensure independence. Anyway, let

$$\begin{aligned}
\mathbf{x}_1 &= \boldsymbol{\xi} + \boldsymbol{\delta}_1 \\
\mathbf{x}_2 &= \boldsymbol{\xi} + \boldsymbol{\delta}_2 \\
\mathbf{y} &= \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta},
\end{aligned}$$

   with $\boldsymbol{\delta}_1$, $\boldsymbol{\delta}_2$, $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ all independent, $V(\boldsymbol{\xi}) = \boldsymbol{\Phi}$, $V(\boldsymbol{\zeta}) = \boldsymbol{\Psi}$, $V(\boldsymbol{\delta}_1) = \boldsymbol{\Theta}_1$, $V(\boldsymbol{\delta}_2) = \boldsymbol{\Theta}_2$, and $\boldsymbol{\Theta}_1 \neq \boldsymbol{\Theta}_2$. Is this model identified? Answer Yes or No, and prove it.

4. Now we would like to see if we can do multiple regression entirely with latent variables, as long as we have two independent measurements of everything. You could think of this as a null-$\boldsymbol{\beta}$ rule for latent variables. Let

$$
\begin{aligned}
\mathbf{x}_1 &= \boldsymbol{\xi} + \boldsymbol{\delta}_1 \\
\mathbf{x}_2 &= \boldsymbol{\xi} + \boldsymbol{\delta}_2 \\
\mathbf{y}_1 &= \boldsymbol{\eta} + \boldsymbol{\epsilon}_1 \\
\mathbf{y}_2 &= \boldsymbol{\eta} + \boldsymbol{\epsilon}_2 \\
\boldsymbol{\eta} &= \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta},
\end{aligned}
$$

where we are following LISREL-type notation, adding $V(\boldsymbol{\epsilon}_1) = \boldsymbol{\Theta}_{\epsilon_1}$, $V(\boldsymbol{\epsilon}_2) = \boldsymbol{\Theta}_{\epsilon_2}$, $V(\boldsymbol{\delta}_1) = \boldsymbol{\Theta}_{\delta_1}$, $V(\boldsymbol{\delta}_2) = \boldsymbol{\Theta}_{\delta_2}$, and as usual all the vectors of error terms are independent of each other and of $\boldsymbol{\xi}$. Is this model identified? Answer Yes or No, and prove it.

5. The next step would be to add a vector of manifest exogenous variables, and you would think that no additional conditions on the model would be necessary; you would think that with two independent measurements on all the latent variables, the model would be identified. But instead, things start to get really complicated. The main source of the complication is the set of unknown covariances between $\boldsymbol{\xi}$ and $\mathbf{x}_3$. So we'll have to be satisfied with a scalar version of the problem. Let

$$
\begin{aligned}
x_1 &= F + e_1 \\
x_2 &= F + e_2 \\
y &= b_1 F + b_2 x_3 + e_3,
\end{aligned}
$$

where F and x_3 are independent of the error terms e_1, e_2 and e_3, the error terms are independent of each other, and $Cov(F, x_3) = \kappa$. Is the model identified? Answer Yes or No, and prove it.

6. Consider this model:

$$
\begin{aligned}
y_1 &= b_1 x_1 + b_2 y_2 + e_1 \\
y_2 &= b_3 x_2 + b_4 y_1 + e_2,
\end{aligned}
$$

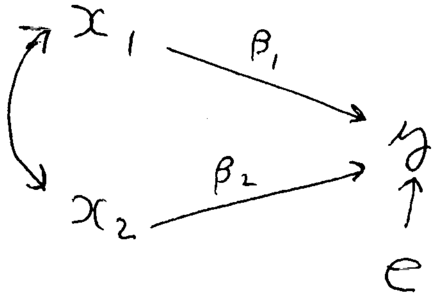
where x_1 and x_2 are correlated with each other (say $Cov(x_1, x_2) = \kappa$), and independent of the error terms e_1 and e_2, which are independent of each other. Is the model identified? Answer Yes or No, and prove it. Hint: Try to do it yourself, but if you get stuck, see Chapter 5 of our almost useless online "textbook."

7. On the following pages are some hand-drawn path models. For each one, state whether the model is identified and justify your answer. Sometimes, it's very quick. Feel free to use the results you've already proved in earlier questions, if they apply.

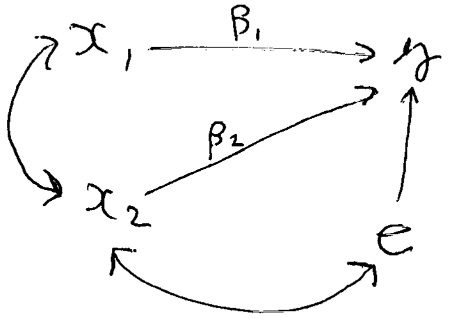
   Then, try to express each model in LISREL notation; say what all the vectors and matrices are. For many models, parts of the LISREL model are just missing. Sometimes there is more than one way to express the answer. For models containing just observed variables, you may start with

$$
\mathbf{y} = \boldsymbol{\beta}\mathbf{y} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}.
$$

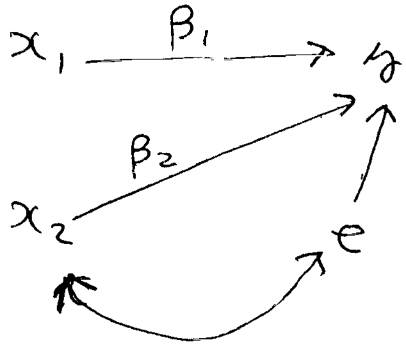
(a)



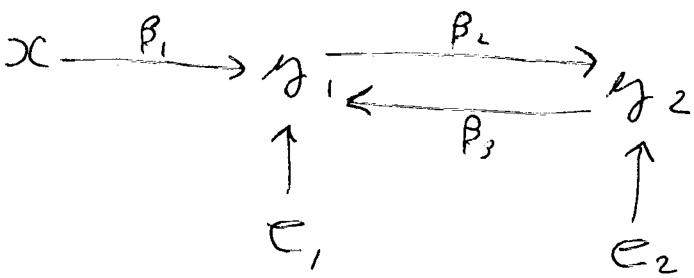
(b)



(c)

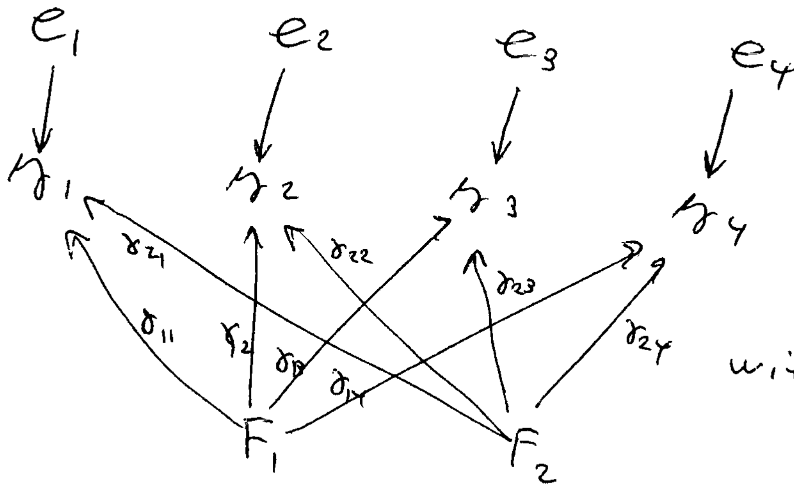


~~(d)~~



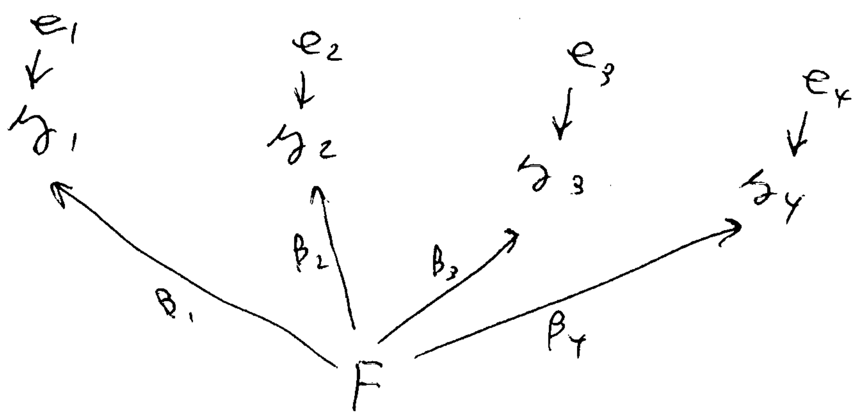
This one is scary. Can you just find Σ?
What conditions on β_s are necessary (not sufficient) for identification?

(d)



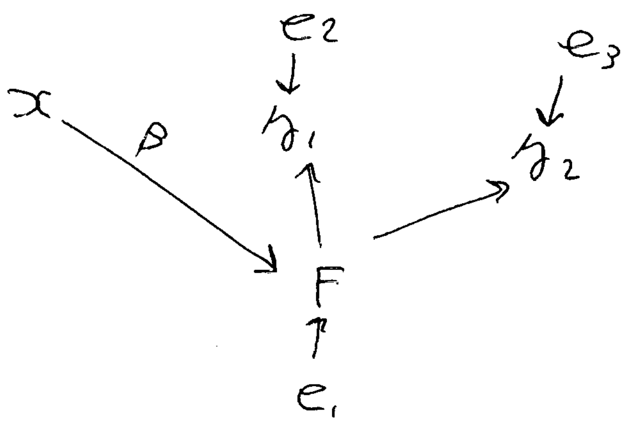
with $var(F_1) = var(F_2) = 1$

(e)

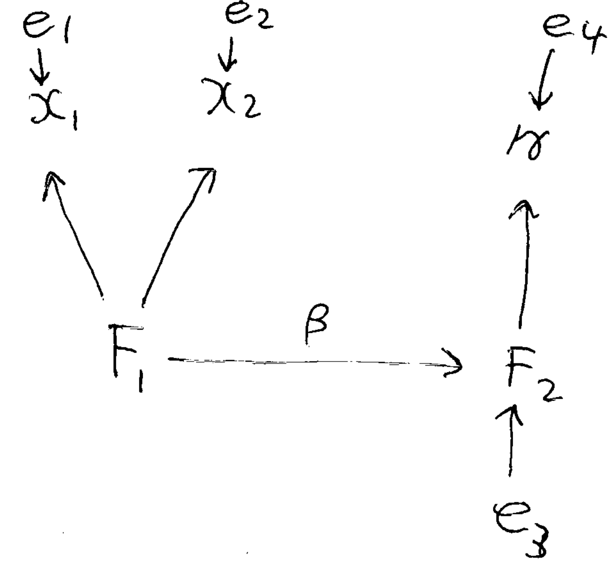


$V(F) = V(e_1) = V(e_2) = V(e_3) = V(e_4) = 1$

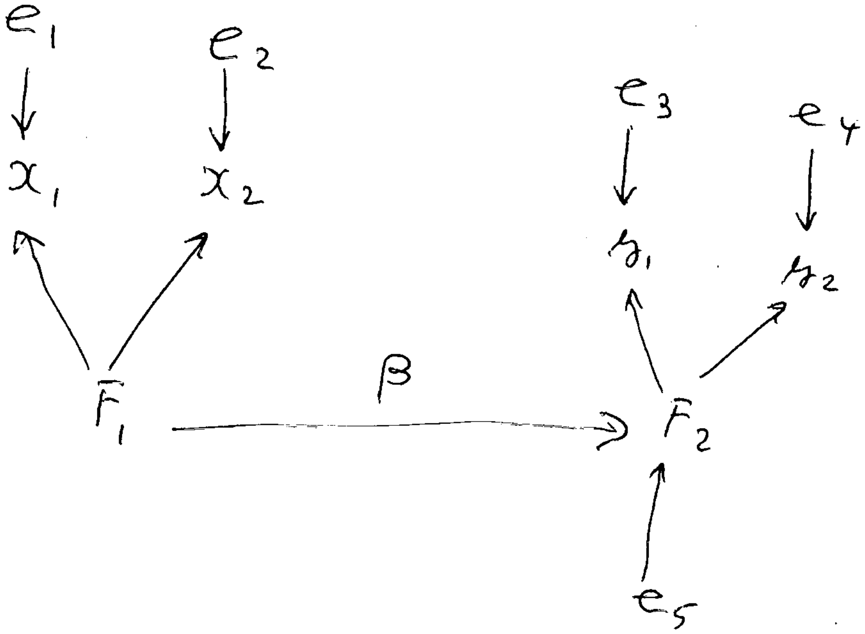
(f)



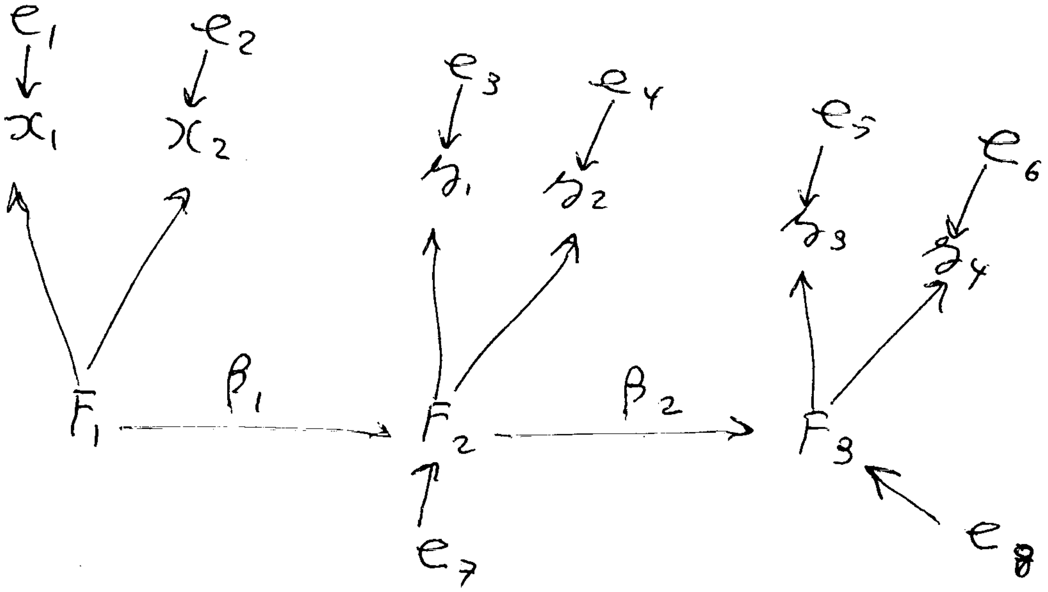
(g)



(h)

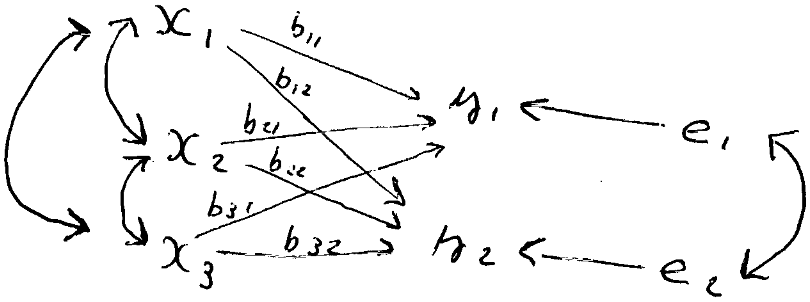


(i)

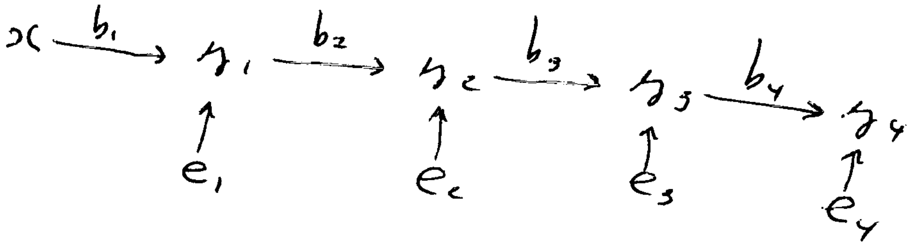


Build on the last problem. Still too big? I bet it's identified.

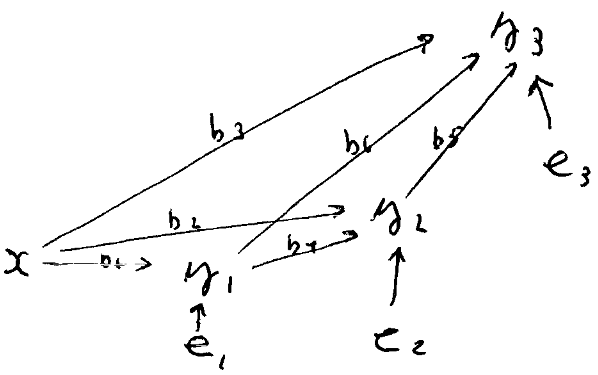
(j)


x_1
x_2
x_3
b_{11}
b_{12}
b_{21}
b_{22}
b_{31}
b_{32}
y_1
y_2
e_1
e_2

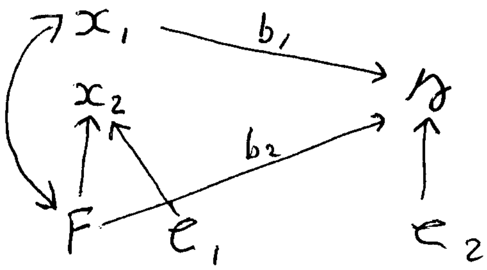
(k)


x
b_1
y_1
b_2
y_2
b_3
y_3
b_4
y_4
e_1
e_2
e_3
e_4

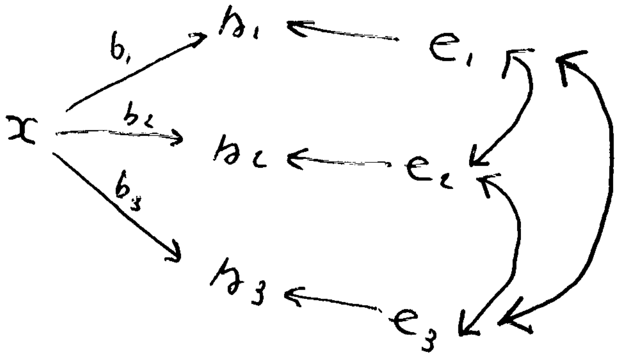
(l)


y_3
b_3
b_6
b_5
e_3
b_2
y_2
x
b_1
y_1
b_4
e_1
e_2

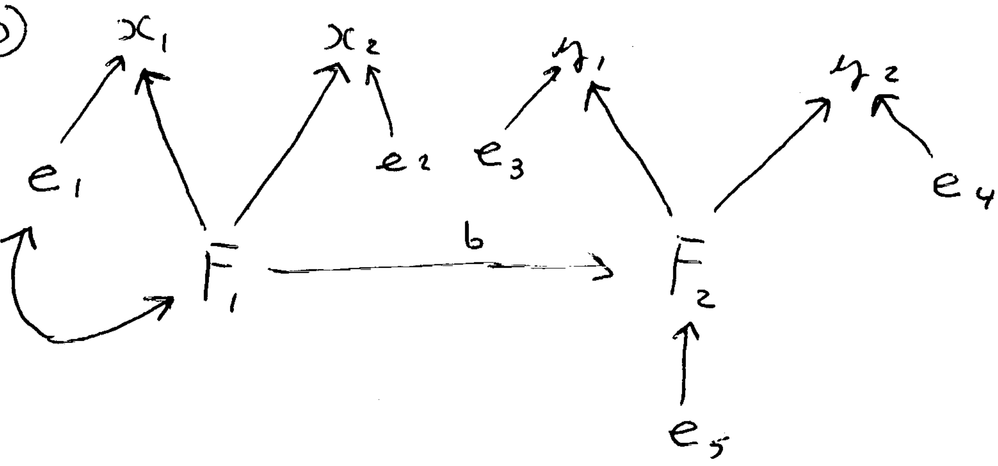
(m)


x_1
b_1
y
x_2
b_2
F
e_1
e_2

(n)

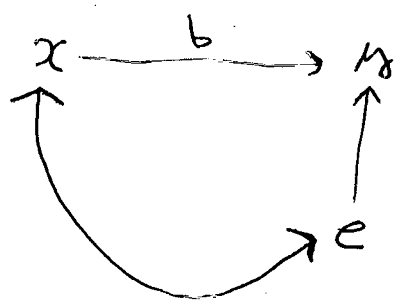


(o)

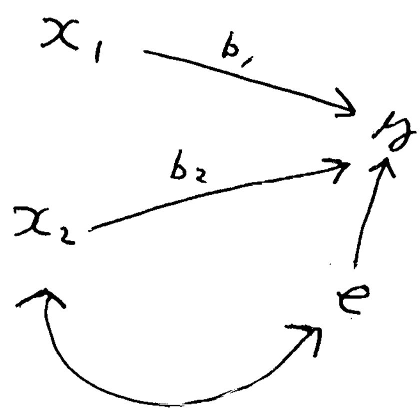


with var(F1)=1

(p)

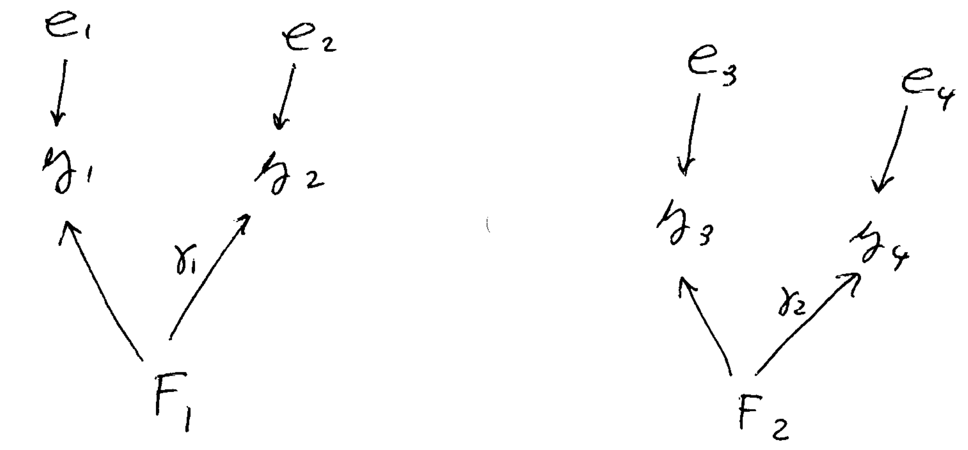


(q)



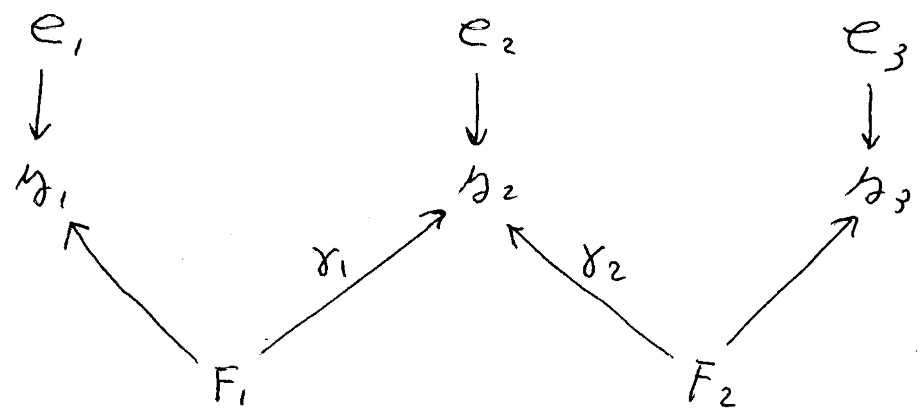
On this page, Y_1 is F_1 measured with error, and Y_2 is F_2 measured with error.

(R)



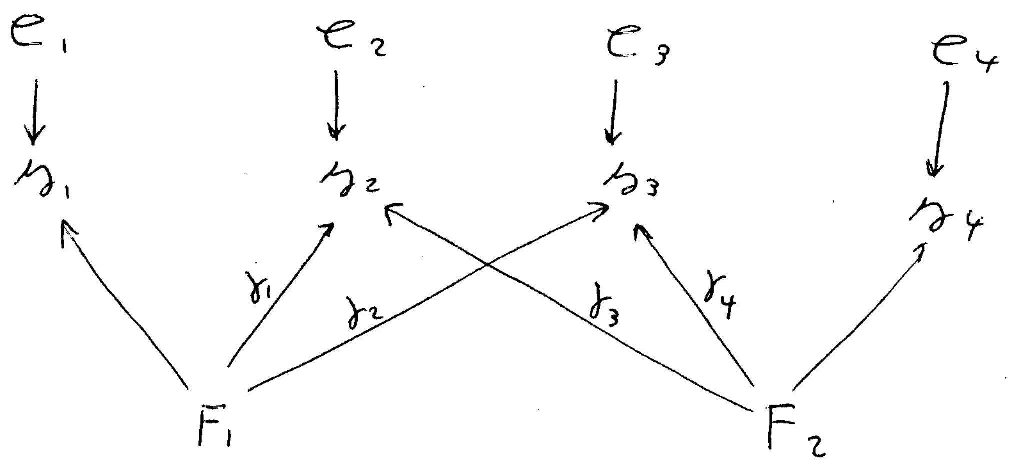
F_1 & F_2 are independent $N(0,1)$.

(s)



F_1 & F_2 are independent $N(0,1)$

(t)



F_1 & F_2 are independent $N(0,1)$