NON-ZERO COVARIANCE BETWEEN ERRORS AND INDEPENDENT VARIABLES: TRYING TO INCLUDE IT IN THE MODEL

EXAMPLE (If we can't make this work, we're in trouble)

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \, ,$$

$$\begin{pmatrix} X_i \\ \varepsilon_i \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & c \\ c & \sigma^2 \end{pmatrix} \right)$$

All we can observe is pairs (X_i, Y_i). Should be bivariate normal.

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = A \begin{pmatrix} X_i \\ \varepsilon_i \end{pmatrix} + c$$

$$= \begin{pmatrix} 1 & 0 \\ \beta_1 & 1 \end{pmatrix} \begin{pmatrix} X_i \\ \varepsilon_i \end{pmatrix} + \begin{pmatrix} 0 \\ \beta_0 \end{pmatrix}$$

save

$$\text{So } E\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \begin{pmatrix} \mu \\ \beta_0 + \beta_1 \mu \end{pmatrix}, \text{ and}$$

.2

$$\operatorname{cov}\begin{pmatrix} X_i \\ Y_i \end{pmatrix} = A \Sigma A'$$

$$= \begin{pmatrix} 1 & 0 \\ \beta_1 & 1 \end{pmatrix} \begin{pmatrix} \sigma_x^2 & c \\ c & \sigma^2 \end{pmatrix} \begin{pmatrix} 1 & \beta_1 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_x^2 & c \\ \beta_1 \sigma_x^2 + c & \beta_1 c + \sigma^2 \end{pmatrix} \begin{pmatrix} 1 & \beta_1 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{array}{c|c|c} & x & y \\ \hline x & \sigma_x^2 & \beta_1 \sigma_x^2 + c \\ \hline y & \beta_1 \sigma_x^2 + c & \begin{array}{c} \beta_1^2 \sigma_x^2 + \beta_1 c \\ + \beta_1 c + \sigma^2 \\ = \beta_1^2 \sigma_x^2 + 2\beta_1 c \\ + \sigma^2 \end{array} \end{array}$$

save

NOW, ALL DATA CAN EVER TELL YOU about IS THE PROBABILITY DISTRIBUTION FROM WHICH THE DATA COME

- If $\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N(\mu, \Sigma)$ all you can ever learn is the values $\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ & \sigma_{22} \end{pmatrix}$

with more and more accuracy for bigger n

But we have a <u>MODEL</u>, with parameters

$\theta = (\mu, \sigma_x^2, \beta_0, \beta_1, \sigma_\epsilon^2, c)$

Suppose as an extreme best-case, we had an infinite sample size.

- We would know $\mu_1\ \mu_2\ \sigma_{11}\ \sigma_{12}\ \sigma_{22}$.
- We know the connections between these & θ
- <u>Could</u> we know θ?

A mathematical problem: Can we solve for the model parameters, given <u>complete</u> knowledge of the distribution of the data?

5 equations in 6 unknowns

Preserve these
call it ✡

$$\mu = \mu_1$$

$$\beta_0 + \beta_1 \mu = \mu_2$$

$$\sigma_x^2 = \sigma_{11}$$

$$\beta_1 \sigma_x^2 + c = \sigma_{12}$$

$$\beta_1^2 \sigma_x^2 + 2\beta_1 c + \sigma^2 = \sigma_{22}$$

A system of linear equations with more unknowns than equations has either no solutions or ∞ many, but these are not linear

Start solving for the model parameters

$$\mu = \mu_{1}$$

$$\sigma_x^2 = \sigma_{11}$$

This is good news. It shows that at least we can recover some model parameters.

The important one is β_1. Can we know that? Keep working.

SUBSTITUTE for μ & σ_x^2

Obtaining 3 equations in 4 unknowns

$$\beta_0 + \beta_1 \mu_1 = \mu_2$$

$$\beta_1 \sigma_{11} + c = \sigma_{12}$$

$$\beta_1^2 + 2\beta_1 c + \sigma^2 = \sigma_{22}$$

- First equation could be solved for β_0 if we knew β_1
- Last equation could be solved for σ^2 if we knew β_1 & c
- Look at 2nd equation. Infinitely many pairs (c, β_1) will satisfy it.

Solve for β_0, c & σ^2 in terms of β_1. Obtaining ∞ many parameter vectors satisfying ☆ as $-\infty < \beta_1 < \infty$

$$\beta_0 = \mu_2 - \beta_1 \mu$$

$$c = \sigma_{12} - \beta_1 \sigma_{11}$$

$$\sigma^2 = \sigma_{22} - \beta_1^2 \sigma_x^2 - 2\beta_1 c$$

$$= \sigma_{22} - \beta_1^2 \sigma_{11} - 2\beta_1 c$$

$$= \sigma_{22} - \beta_1^2 \sigma_x^2 - 2\beta_1 (\sigma_{12} - \beta_1 \sigma_{11})$$

$$= \sigma_{22} - \beta_1^2 \sigma_{11} - 2\beta_1 \sigma_{12} + 2\beta_1^2 \sigma_{11}$$

$$= \sigma_{22} + \beta_1^2 \sigma_{11} - 2\beta_1 \hat{\sigma}_{12}$$

Instead of forming a single point, the solutions to ✱ form an infinite set, a 3-d surface in the 6-d parameter space.

In particular, ∞ many values of β_1 yield the same $(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$ distribution of $\binom{X_i}{Y_i}$. How could you use $(\hat{\mu}, \hat{\Sigma})$ to distinguish between them?

Example: X_i correlated with ε_i

```
> # MVN Mu and Sigma as a function of model parameters
> MuSig = function(mu,sigsqx,beta0,beta1,sigsq,c)
+         {
+         Mu = rbind(0,0)
+         Sigma = matrix(0,2,2)
+         Mu[1,1] = mu
+         Mu[2,1] = beta0 + beta1*mu
+         Sigma[1,1] = sigsqx;     Sigma[1,2] = beta1*sigsqx + c
+         Sigma[2,1] = Sigma[1,2]; Sigma[2,2] = beta1^2*sigsqx + 2*beta1*c + sigsq
+         out = list(Mu,Sigma)
+         return(out)
+         } # End function MuSig
>
> Example1 = MuSig(mu=7, sigsqx=5, beta0=3, beta1=0, sigsq=30, c=10)
> Example1
[[1]]
     [,1]
[1,]    7
[2,]    3

[[2]]
     [,1] [,2]
[1,]    5   10
[2,]   10   30

> Mu = Example1[[1]]; Sigma = Example1[[2]]; Mu; Sigma
     [,1]
[1,]    7
[2,]    3
     [,1] [,2]
[1,]    5   10
[2,]   10   30
>
```

```
> Mu; Sigma
     [,1]
[1,]    7
[2,]    3
     [,1] [,2]
[1,]    5   10
[2,]   10   30

> # Produce the same Mu and Sigma from different parameter sets
> b1=3 # New beta1
> b0 = Mu[2,1] - b1*Mu[1,1] # New beta0 = -18
> cc = Sigma[1,2] - b1*Sigma[1,1] # New c = -5
> s2 = Sigma[2,2] + b1^2*Sigma[1,1] - 2*b1*Sigma[1,2] # New sigsq = 15
> MuSig(mu=Mu[1,1], sigsqx=Sigma[1,1], beta0=b0, beta1=b1, sigsq=s2, c=cc)
[[1]]
     [,1]
[1,]    7
[2,]    3

[[2]]
     [,1] [,2]
[1,]    5   10
[2,]   10   30

>
> b1=-3 # New beta1
> b0 = Mu[2,1] - b1*Mu[1,1] # New beta0 = 24
> cc = Sigma[1,2] - b1*Sigma[1,1] # New c = 25
> s2 = Sigma[2,2] + b1^2*Sigma[1,1] - 2*b1*Sigma[1,2] # New sigsq = 135
> MuSig(mu=Mu[1,1], sigsqx=Sigma[1,1], beta0=b0, beta1=b1, sigsq=s2, c=cc)
[[1]]
     [,1]
[1,]    7
[2,]    3

[[2]]
     [,1] [,2]
[1,]    5   10
[2,]   10   30

>
> # Conclusion: When combined with other parameter values,
> # beta0 = 0 or +3 or -3 can produce exactly the same (Mu,Sigma),
> # and so EXACTLY the same probability distribution of the observable data.
> # There is no way to recover the value of beta1 from the distribution of
> # the data. Accurate estimation of beta1 based on the data is impossible.
> # There can be no acceptable test of H0: beta1=0 based on the data.
>
```