

S.R. White]

2014/Sep/10

SEM

(Structural Equation Models)

Path Analysis → Exploratory/Confirmatory Factor Analysis

PCA
ICA
CCA
PLS

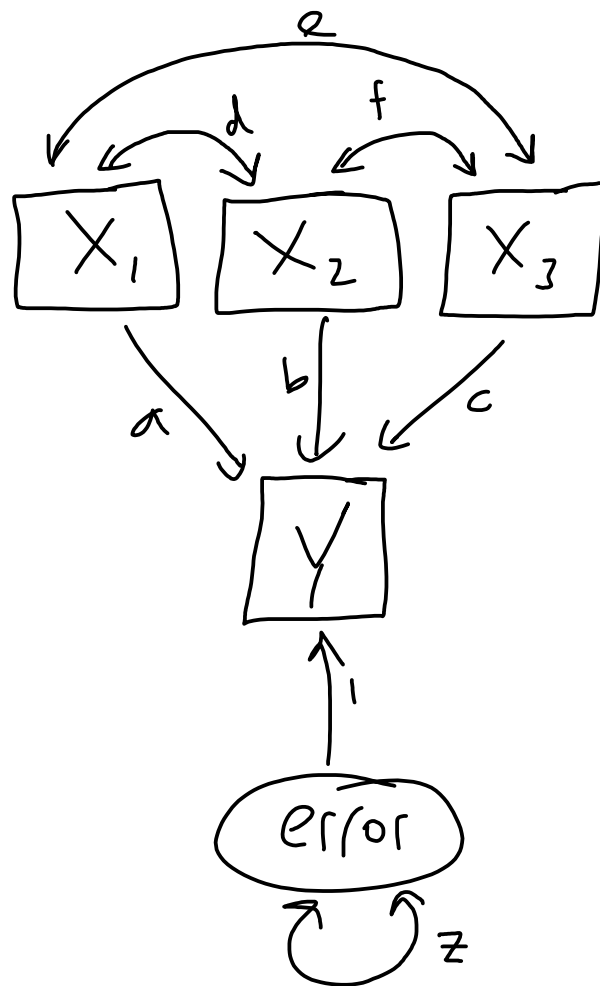
SEM

(PLS-SEM / PLS-PA / PLS-Path)

Pathway Analysis

①

Covariance/Correlation between measured
variables



$\square :=$ observed

$\circ :=$ latent

\longleftrightarrow cov/corr

\rightarrow "direct" relationship

akin to

$$y = ax_1 + bx_2 + cx_3 + \varepsilon$$

Path Analysis \rightarrow "walk along paths to derive cov/corr"
 \rightarrow partition into components

assume data are as follows:

	X_1	X_2	X_3	Y
X_1	1			
X_2	0.2	1		
X_3	0.24	0.3	1	
Y	0.7	0.8	0.3	1

Correlations



key that
basic PA/CFA/Sem
only care
about cov/corr

Can then "walk" along paths

$$d = 0.2$$

$$e = 0.24$$

$$f = 0.3$$

(direct from table)

$$r_{(X_1 Y)} = a + ec + db \Rightarrow$$

$$r_{(X_2 Y)} = b + da + fc \Rightarrow$$

$$r_{(X_3 Y)} = c + fb + ea \Rightarrow$$

$$0.7 = a + 0.24c + 0.2b$$

.....

.....

3 eqn, 3 unknowns



solve to obtain

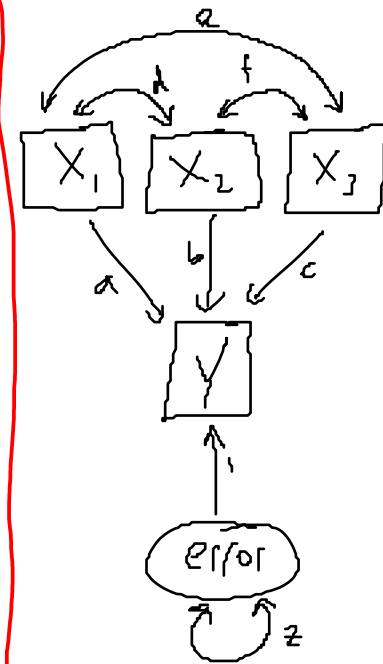
"Path Coefficients"

aka standardised
partial regression
coefficients

$$a = 0.57$$

$$b = 0.70$$

$$c = -0.05$$



Can perform PA in lavaan in R

```
CORR <- lav_matrix_lower2full(c(1.0,0.20,1,0.24,0.30,1,0.70,0.80,0.30,1))
colnames(CORR) <- rownames(CORR) <- c("X1", "X2", "X3", "Y")
CORR.model <- 'Y ~ a*X1 + b*X2 + c*X3; Y ~~ z*Y'
CORR.fit <- sem(CORR.model, sample.cov=CORR, sample.nobs = 1000)
summary(CORR.fit, rsquare = TRUE)
```

```
> summary(CORR.fit, rsquare = TRUE)
lavaan 0.6-4 ended normally after 25 iterations

  Optimization method                 NLMINB
Number of free parameters                4

Number of observations                  1000

Estimator                             ML
Model Fit Test Statistic                0.000
Degrees of freedom                      0
Minimum Function Value                  0.000000000000000

Parameter Estimates:

  Information                               Expected
Information saturated (h1) model          Structured
Standard Errors                          Standard

Regressions:
              Estimate Std.Err z-value P(>|z|)
Y ~
  X1      (a)    0.571    0.008   74.539   0.000
  X2      (b)    0.700    0.008   89.724   0.000
  X3      (c)   -0.047    0.008   -5.980   0.000

Variances:
              Estimate Std.Err z-value P(>|z|)
.Y      (z)    0.054    0.002   22.361   0.000

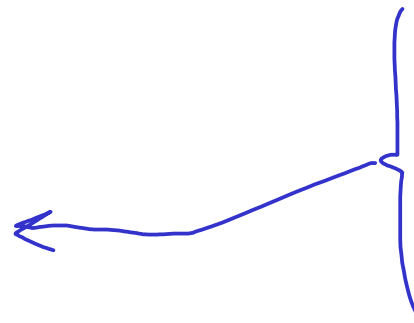
R-Square:
              Estimate
.Y          0.946
```

from PA by hand

$$a = 0.57$$

$$b = 0.70$$

$$c = -0.05$$



Adding Latent Variables

(2)

Move to models of cov/corr
(but can include means & variances)

some claim SEM
subsumes
these ??

SEM

add latent

add structural model
(ie. multiple latent
vars)

CFA/EFA

Item Response

PA

Path Analysis can be
strongly link to all
types of linear
regression

Multiple Regression & ⚡

ANOVA

Correlation

multivariate
multiple
regression

PC

Other
Multivariate

LCA PCA?
or
PCA Regression
(PCR)

Canonical
Correlation

can be part
of PLS-PA

(might be also called PLS-SEM?)

⚡ special mention for
PLS-regression / PLS-path-modelling → not to be confused with PLS-correlation [which is not this]

That was a very busy slide

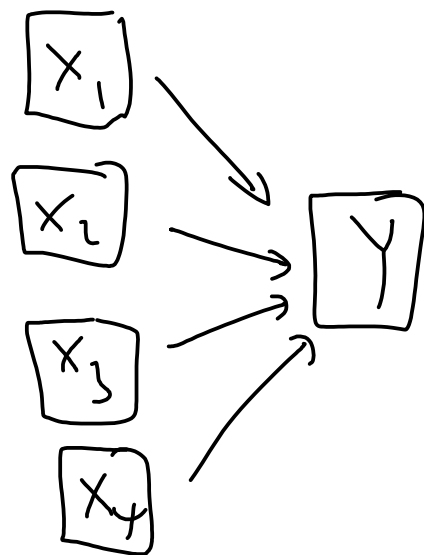
When to use what method?
 * none of them are causal *

PA > MVReg : complex patterns, issue of selecting outcome/covariates
↳ recall in PA all symmetric & bidirectional

SEM > — : multiple latent variables of interest

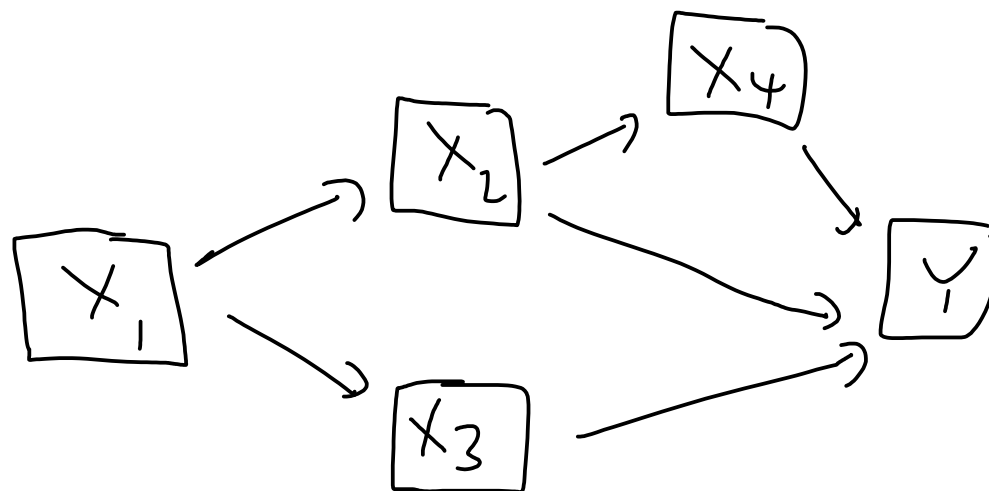
↳ PLS-correlation/PCA/ICA/etc the latent aspects are false constructs, SEM/CFA built on Theory (with a capital T)

Path Analysis Vs "std regression"



Path Analysis has been called
"causal path"

↳ but no causal interpretation



but how to define this
structure?

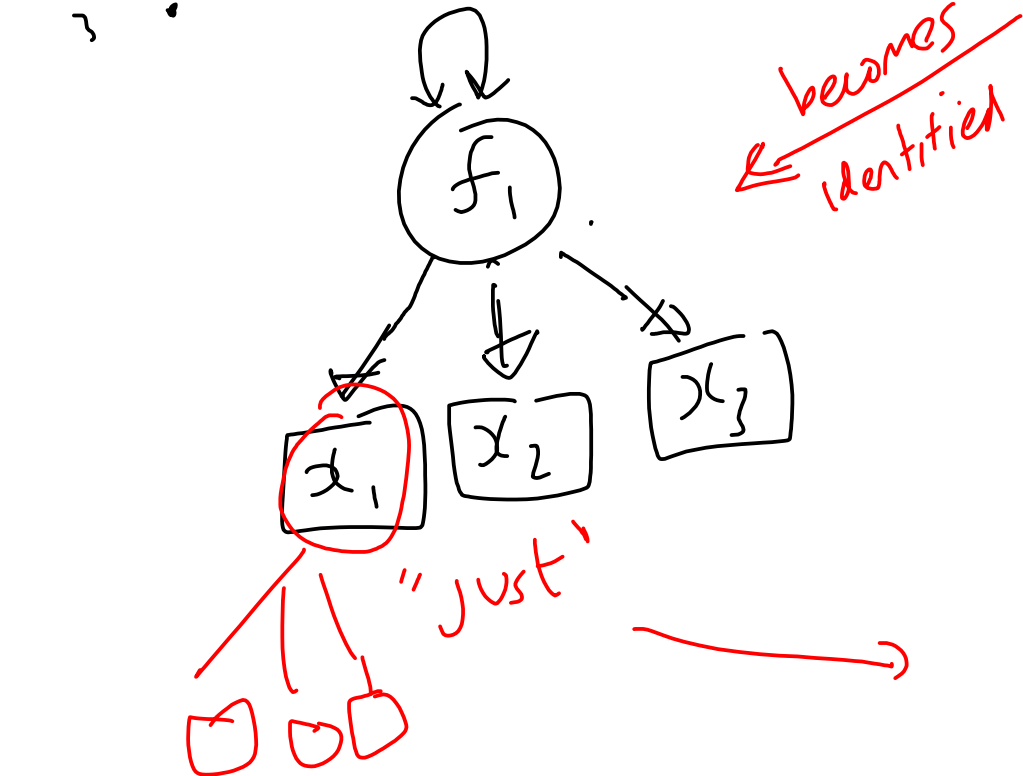
typically unstable

↳ infer by fitting
multiple regressions

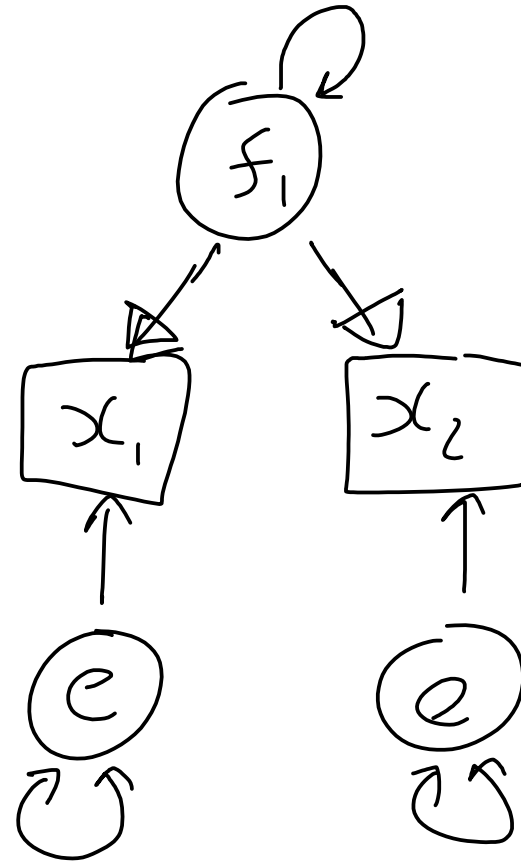
Some fun in SEM

(3)

Identifiability



"under"

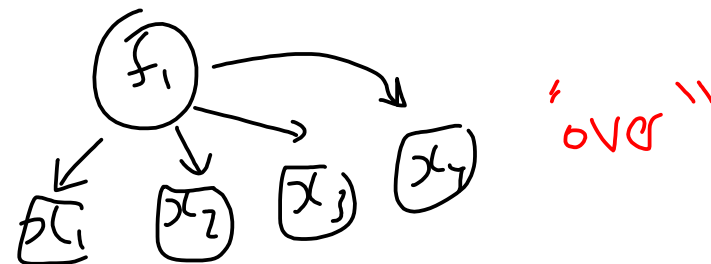


data

	x_1	x_2
x_1	1	
x_2		1

\Rightarrow 1

many unknowns in data



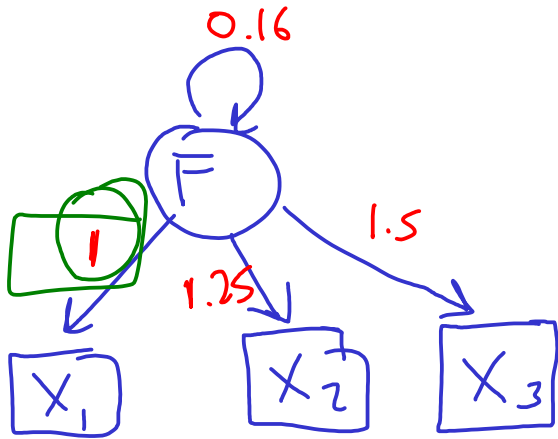
"over"

Interpretation

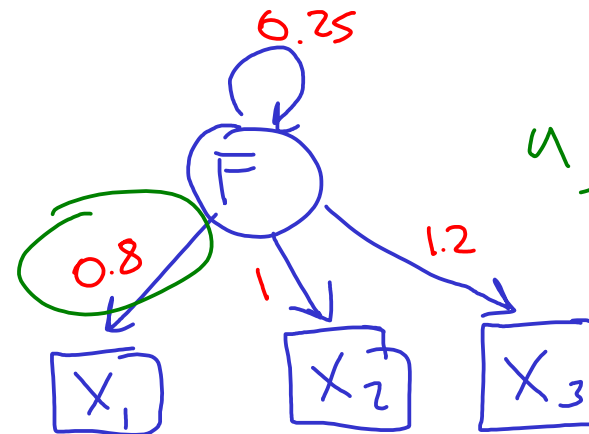
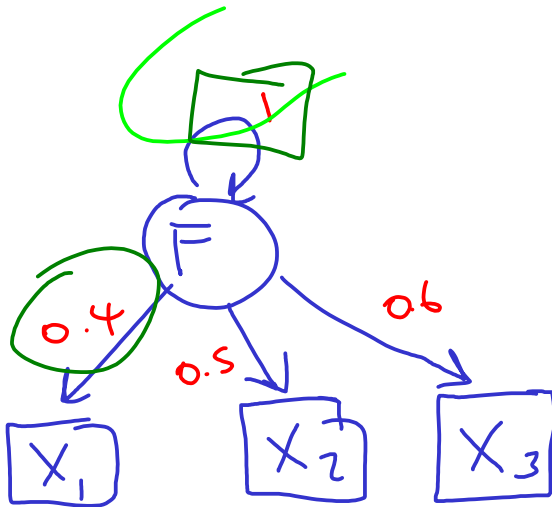
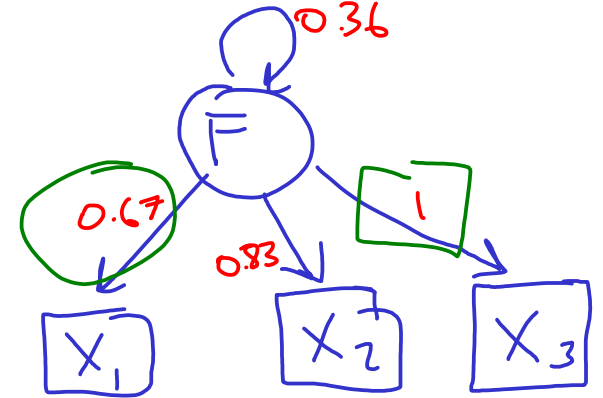
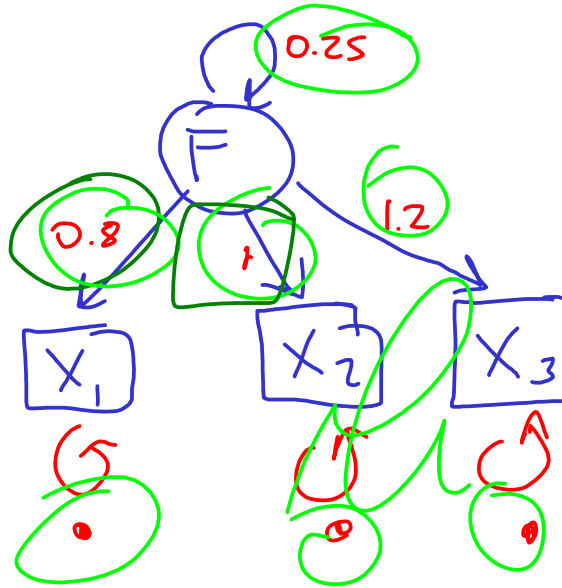
Which of these datasets shows strongest

$$X_1 \sim F \quad \underline{\underline{\text{Ink?}}}$$

Marked node



1



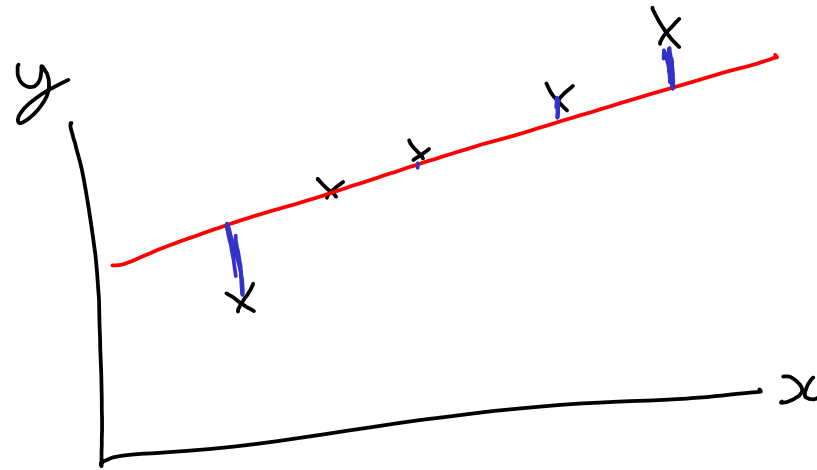
$$\frac{a + b + c}{3} = 1$$

Of course, all the same data

- standardised latent variable ($z=1$)
- marker variable (a or b or $c = 1$)
- effect coding ($a+b+c = 3$)

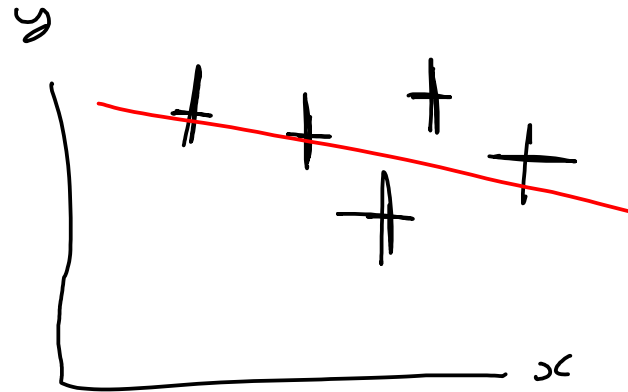
The issue of measurement error

Classic regression



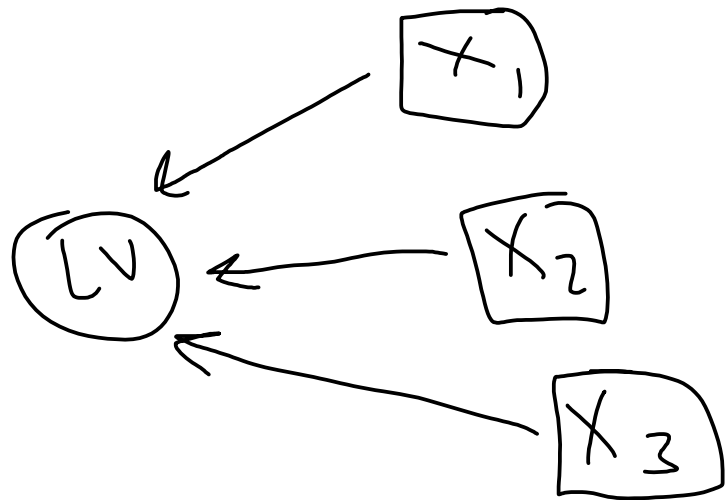
error \downarrow in y
only

error in x & y



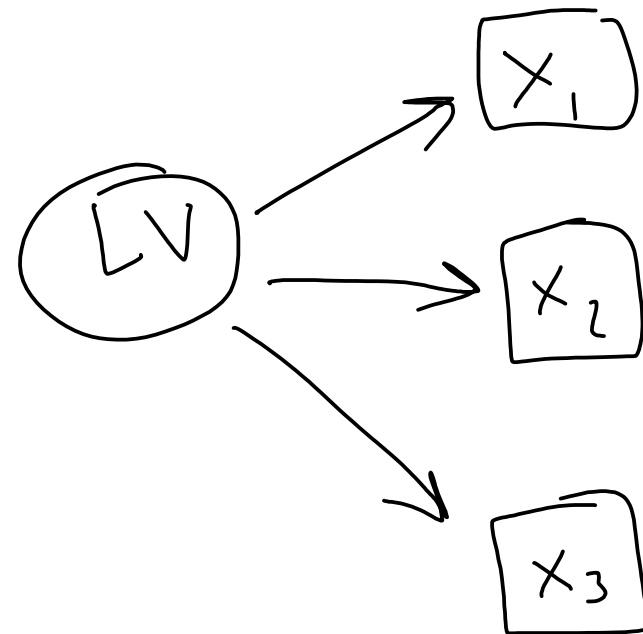
Formative Vs Reflective

Formative



composite/factor SEM
(using PLS)

Reflective



confirming SEM
(using ML)

Missing data

listwise deletion

(no)

for covariance methods — pairwise deletion

(hmm)

Full Information Maximum Likelihood (FIML)

Multiple Imputation (MI)

Does the model fit the data?

hard question for SEM \rightarrow the model is Theory driven



CFA on each construct first

\rightarrow many, many, fit criteria

\sum
too many to list

Funny stuff

Growth Curves

Hierarchical

Bifactor



SEM variants

Groups

↳ parameter constraints