Structural Equation Modelling & Causal Inference
Set-4: Causal Inference with SEM

Ozan Aksoy

UCL Social Research Institute & NCRM
University College London

November 6, 2023

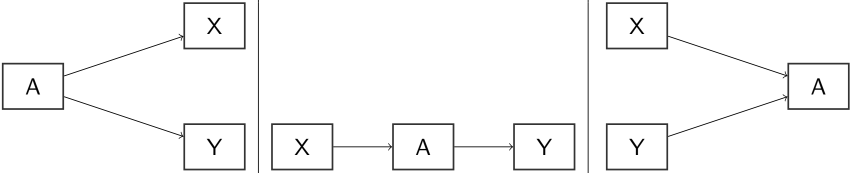
Topics

1. DAG vs SEM
2. Fixed versus random effects models
3. Cross-lagged panel models with fixed effects
4. Instrumental Variables in SEM

- Family of SEM
  - Traditional SEM (what we have been doing so far)
  - DAG a la Pearl (Structural Causal Models–SCM)
  - Composite SEM
  - All stem from Wright's (1934) path analysis
- Traditional SEM
  - Covariance-based
  - Global estimation
  - Parametric (often linear)
  - Combination of observed and latent variables
  - Factors are "reflective" (think of CFA)
- DAG
  - Nonparametric (any association, linear, quadratic etc. is allowed)
  - Model is analysed before data are collected
  - Model gives a series of testable implications (e.g. conditional independendences)
  - Analysis shows whether it is possible to obtain a "causal" effect at all given model
- Composite SEM
  - Factors are composite (linear combinations of indicators)
  - Formative measurement
  - Normally specialised software was needed to fit C-SEM
- Recent developments made the three families merge
  - Piecewise SEM (local estimation of paths): brings SEM and DAG closer
  - New ways of specifying models allow fitting C-SEM models in traditional SEM

A flavour of DAG: building blocks of DAG and SCM

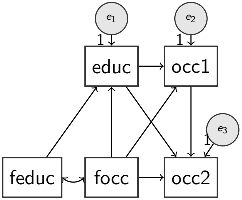
(a) fork (b) chain (c) collider (inverted fork)



(a) fork	(b) chain	(c) collider (inverted fork)
$X \not\perp Y$	$X \not\perp Y$	$X \perp Y$
$X \perp Y \mid A$	$X \perp Y \mid A$	$X \not\perp Y \mid A$

- Any given model includes some combinations of above
- A model gives many such predictions of (conditional) local (in)dependences
- Testing those predictions tests the model
- Whether the model identifies a particular total causal effect X on Y, possibly conditional on some Z can be determined (backdoor criterion)
- SCM also identifies "(conditional) instruments" for a causal effect

A flavour of DAG: example

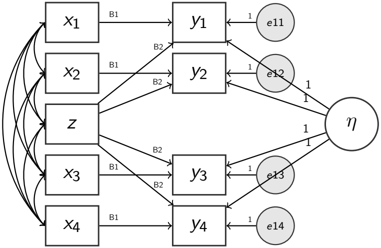



```
library(dagitty)
bddag <- dagitty("dag{feduc <-> focc; focc -> educ; focc -> occ1; focc -> occ2;
                       feduc -> educ; educ -> occ1; educ -> occ2; occ1 -> occ2}")
# minimally sufficient adjustment sets of covariates
# direct effects
adjustmentSets(bddag, "educ", "occ1", effect = "direct")
> focc
adjustmentSets(bddag, "educ", "occ2", effect = "direct")
> focc, occ1
# total effects
adjustmentSets(bddag, "educ", "occ1", effect = "total")
> focc
adjustmentSets(bddag, "educ", "occ2", effect = "total")
> focc
#implied conditional independencies
impliedConditionalIndependencies(bddag)
fedc _||_ occ1 | educ, focc
fedc _||_ occ2 | educ, focc
```

Piecewise-SEM to combine traditional SEM with DAG

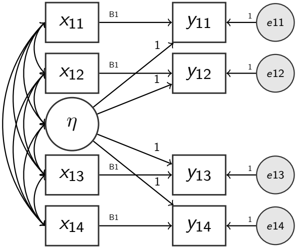

```
library(semTools)
library(lavaan)
library(piecewiseSEM)
library(psych)

#create blau & duncan raw data
bd_low <- '
1.0000
0.5160  1.0000
0.4530  0.4380  1.0000
0.3320  0.4170  0.5380  1.0000
0.3220  0.4050  0.5960  0.5410  1.0000'
bd.corr <- getCov(bd_low, names = c("feduc", "faocc", "educ", "occ1", "occ2"))
bd.data <- semTools::kd(bd.corr, 20700, type = "exact")

#piece-wise SEM
bd.model <- piecewiseSEM::psem(
  lm(occ2 ~ occ1 + educ + faocc, data = bd.data),
  lm(occ1 ~        educ + faocc, data = bd.data),
  lm(educ ~       feduc + faocc, data = bd.data),
  faed %~~% faocc,  data = bd.data)

#the same as daggity implied indeps@
piecewiseSEM::basisSet(bd.model)

#test of conditional independences:
piecewiseSEM::dSep(bd.model, conditioning = TRUE, .progressBar = FALSE)
>                        Independ.Claim Test.Type    DF Crit.Value      P.Value
> 1 occ2 ~ feduc + faocc + occ1 + educ      coef 20695  -2.181017 0.0291933989   *
> 2 occ1 ~ feduc + faocc + educ             coef 20696   3.655547 0.0002572688 ***
#test of overall model fit
piecewiseSEM::fisherC(bd.model, .progressBar = FALSE)
>   Fisher.C df P.Value
> 1   23.598  4       0

#checking observed values of those implied independences:
pr  <- psych::partial.r(bd.data, c("feduc","occ1"), c("faocc", "educ"))
print(pr, digits = 3)
> partial correlations
>       feduc  occ1
> feduc 1.000 0.025
> occ1  0.025 1.000

pr2  <- psych::partial.r(bd.data, c("feduc","occ2"), c("faocc", "educ"))
print(pr2, digits = 3)
> partial correlations
>        feduc   occ2
> feduc  1.000 -0.007
> occ2  -0.007  1.000
```

Random effects versus fixed effects models



Random effects model

$$y_{it} = B_1 x_{it} + B_2 z_i + \eta_i + e_i$$
$$e_{it} \sim N(0, \sigma^2) \; cor(\eta, x_{it}) = cor(\eta, z_i) = 0$$



Fixed effects model

$$y_{it} = B_1 x_{it} + \eta_i + e_i$$
$$e_{it} \sim N(0, \sigma^2) \; cor(\eta, x_{it}) = \theta_t$$

How to do FE in SEM?


```
> d2
# A tibble: 4,000 × 5
       id      z      t    xit     yit
 1      1 -2.41      1 -3.88  -3.68
 2      1 -2.41      2 -2.13  -3.43
 3      1 -2.41      3 -2.15  -5.13
 4      1 -2.41      4  0.455 -1.96
 5      2  0.555     1 -2.84  -1.25

summary(lm(yit ~ xit + z, data = d2))
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.02206    0.03120   0.707     0.48
xit          0.41584    0.01401  29.683   <2e-16 ***
z            0.87228    0.01637  53.290   <2e-16 ***

summary(lm(yit ~ xit , data = d2))
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.02686    0.04079  -0.658     0.51
xit          0.63551    0.01751  36.296   <2e-16 ***

summary(plm::plm(yit ~ xit, data = d2,
        index = c("id", "t"), model = "within"))
    Estimate Std. Error t-value  Pr(>|t|)
xit 0.417640   0.018468  22.614 < 2.2e-16 ***
```

How to do FE in lavaan?


```
dw <- d2 %>% pivot_wider(names_from = t, values_from = c(xit, yit))
# A tibble: 1,000 × 10
# Groups:   id [1,000]
      id       z  xit_1  xit_2   xit_3  xit_4  yit_1   yit_2  yit_3  yit_4
   <dbl>   <dbl>  <dbl>  <dbl>   <dbl>  <dbl>  <dbl>   <dbl>  <dbl>  <dbl>
 1      1 -2.41  -3.88  -2.13  -2.15    0.455 -3.68  -3.43   -5.13  -1.96
 2      2  0.555 -2.84  -1.62  -3.01    1.49  -1.25   0.174  -1.74   1.88
 m3<-'yit_1 ~ b*xit_1
      yit_2 ~ b*xit_2
      yit_3 ~ b*xit_3
      yit_4 ~ b*xit_4
      A =~ 1*yit_1 + 1*yit_2 + 1*yit_3 + 1*yit_4
      yit_1 + b*yit_2 + yit_3 + yit_4 ~ a*1
      yit_1 ~~ e*yit_1
      yit_2 ~~ e*yit_2
      yit_3 ~~ e*yit_3
      yit_4 ~~ e*yit_4
      xit_1 ~~ NA*A
      xit_2 ~~ NA*A
      xit_3 ~~ NA*A
      xit_4 ~~ NA*A'
summary(sem(m3, data=dw))
```

Results of FE in lavaan


```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  A =~
    yit_1             1.000
    yit_2             1.000
    yit_3             1.000
    yit_4             1.000
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  yit_1 ~
    xit_1      (b)    0.418    0.016   26.929    0.000
  yit_2 ~
    xit_2      (b)    0.418    0.016   26.929    0.000
...
Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  A ~~
    xit_1             0.817    0.145    5.631    0.000
    xit_2             0.559    0.136    4.098    0.000
    xit_3             0.560    0.138    4.051    0.000
    xit_4             0.574    0.143    4.023    0.000
Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)
   .yit_1      (a)   -0.024    0.061   -0.400    0.689
   .yit_2      (a)   -0.024    0.061   -0.400    0.689
...
  A                 0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .yit_1      (e)    3.916    0.101   38.730    0.000
   .yit_2      (e)    3.916    0.101   38.730    0.000
 ...
    xit_1             5.626    0.252   22.361    0.000
    xit_2             5.149    0.230   22.361    0.000
    xit_3             5.290    0.237   22.361    0.000
    xit_4             5.628    0.252   22.361    0.000
    A                 2.738    0.169   16.211    0.000
```

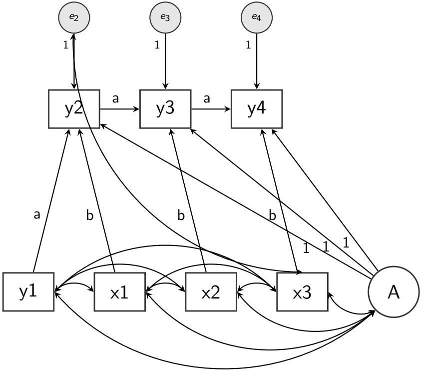
Advantages of doing FE in SEM

- Error variances can easily be allowed to vary over time
- FE can have different effects at different times
- Error auto-correlations can be added
- Missing data can be handled very easily with FIML
- Other latent variables can be easily added in the model
- Time-invariant variables can be included in the model (by constraining its correlation with FE to zero)

Disadvantages of doing FE in SEM

- Code can be tedious (not always a problem)
- Convergence issues possible (though there are often solutions)
- Does not work well for large T small N

Cross-legged panel models with fixed effects

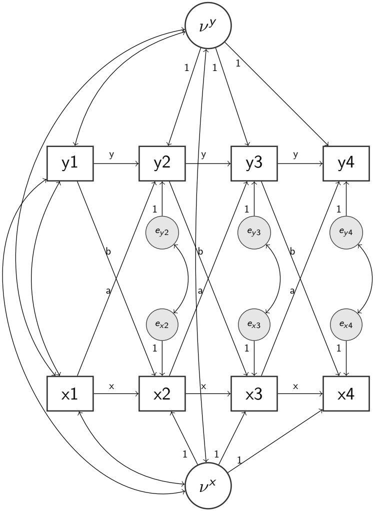


- Allison et al. (2017) Socius
  `https://journals.sagepub.com/doi/full/10.1177/2378023117710578`
- Econ lit: GMM—problematic
- ML-SEM
  - easy
  - works with all SEM packages
  - all advantages of doing FE in SEM

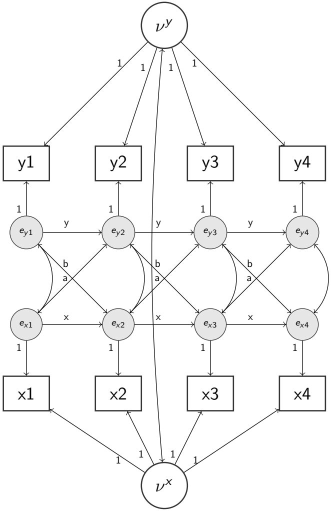
How to do it in lavaan


```
mod <- ‘
 A =~ 1*y2 + 1*y3 + 1*y4
y2 ~  b*x1 + a*y1
y3 ~  b*x2 + a*y2
y4 ~  b*x3 + a*y3
y2 ~~ x3
 A ~~ y1 + x1 + x2 + x3
y1 ~~ x1 + x2 + x3
x1 ~~ x2 + x3
x2 ~~ x3
y2 ~~ e*y2
y3 ~~ e*y3
y4 ~~ e*y4’
summary(fit <- sem(mod, data = D))
```

Variants of cross-legged panel models



Dynamic panel model with FE (FE-DPM)

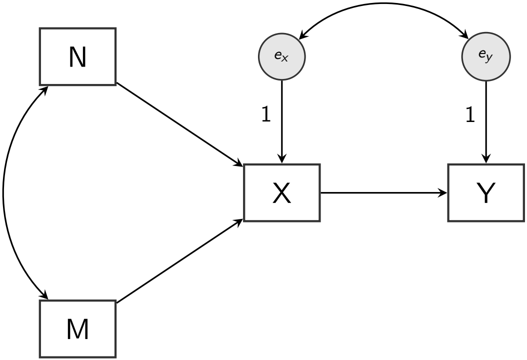


Random-intercept cross-lagged panel model (RI-CLPM)

FE-DPM vs RI-CLPM

- FE-DPM
  - Less flexible than Allison et al. (2017)
  - Otherwise very similar
  - Suitable if attention is on $x \to y$
  - But error correlations in Allison et al. makes it more flexible
- RI-CLPM
  - Suitable when x and y have stable individual component (psyc construct)
  - Focus is on temporal deviations from stable component
  - Do not block "backdoors" via FE
  - Less "causal" leverage

IV in SEM



2sls vs SEM

SEM w.o. error covariance

```
m <- 'y ~ x
      x ~ m + n'
summary(f1<-sem(m, data = d))
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  y ~ x             0.033    0.007    4.377    0.000 #biased
```

SEM with error covariance

```
m2<- 'y ~ x
      x ~ m + n
      y ~~ x'
summary(f2<-sem(m2, data = d))
                   Estimate  Std.Err  z-value  P(>|z|)
  y ~ x            -0.286    0.014  -20.403    0.000 #unbiased
```

2sls method

```
summary(f3<-ivreg(y ~ x | m+n , data = d))
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.006456   0.014085  -0.458    0.647
x           -0.285930   0.014016 -20.401   <2e-16 ***
```

Testing the exclusion restriction assumptions

Exlusion restriction

```
m3 <- 'y ~ x
      x ~ m + n
      y ~ b*m
      y ~~ x'
m4 <- 'y ~ x
      x ~ m + n
      y ~ c*n
      y ~~ x'

summary(f4<-sem(m3, data = d))
lavTestWald(f4, constraints = "b == 0")

summary(f5<-sem(m4, data = d))
lavTestWald(f5, constraints = "c == 0")
```

Results exclusion restriction assumptions

results

```
                      Estimate  Std.Err  z-value
  y ~ x                 -0.287    0.020  -14.399
  x ~ m                  0.708    0.014   50.831
    ~ n                  0.705    0.014   50.081
  y ~ m         (b)      0.001    0.020    0.056
----------------------------------------------
  y ~ x                 -0.285    0.020  -14.546
  x ~ m                  0.708    0.014   50.831
    ~ n                  0.705    0.014   50.081
  y ~ n         (c)     -0.001    0.020   -0.056

> lavTestWald(f4, constraints = "b == 0")
$stat [1] 0.003148568 $df [1] 1 $p.value [1] 0.9552525
> lavTestWald(f5, constraints = "c == 0")
$stat [1] 0.003153189 $df [1] 1 $p.value [1] 0.9552197
```

IVs in DAG framework via piecewise-SEM with 2SLS


```
> #(conditional) instrumental variables
> dagitty::instrumentalVariables(bddag, "educ", "occ1")
 feduc |  focc
> dagitty::instrumentalVariables(bddag, "educ", "occ2")
 feduc |  focc
>
> # compute conditional instruments listed next and add to data frame
> # eo, feduc |  focc
> eo <- lm(feduc ~ faocc, data = bd.data)
>
> # add instrument to data frame
> bd.data$eo <- eo$resid
>
> #check corr is zero
> cor(bd.data$eo, bd.data$faocc, method = "pearson")
[1] 1.139052e-17
>
> # educ -> occ1, instr =  feduc |  focc
> edu_oc1 <- systemfit::systemfit(occ1 ~ educ, inst = ~ eo, method = "2SLS", data = bd.data)
> coef(summary(edu_oc1))
                    Estimate  Std. Error      t value Pr(>|t|)
eq1_(Intercept) 2.126340e-17 0.005861397 3.627702e-15        1
eq1_educ        5.146789e-01 0.022118865 2.326878e+01        0
>
> # educ -> occ2, instr =  feduc |  focc
> edu_oc2 <- systemfit::systemfit(occ2 ~ educ, inst = ~ eo, method = "2SLS", data = bd.data)
> coef(summary(edu_oc2))
                    Estimate  Std. Error      t value Pr(>|t|)
eq1_(Intercept) 6.847015e-18 0.005622893 1.217703e-15        1
eq1_educ        4.979030e-01 0.021218834 2.346514e+01        0
```

Compare piecewise-SEM with traditional SEM


```
bdsem <- '
occ2 ~ occ1 + educ + faocc
occ1 ~  educ + faocc
educ ~ feduc + faocc'
summary(sem(bdsem, data = bd.data))

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  occ2 ~
    occ1              0.281    0.006   43.926    0.000
    educ              0.395    0.006   61.060    0.000
    faocc             0.115    0.006   19.211    0.000
  occ1 ~
    educ              0.440    0.006   69.488    0.000
    faocc             0.224    0.006   35.463    0.000
  educ ~
    feduc             0.309    0.007   44.383    0.000
    faocc             0.278    0.007   39.937    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .occ2              0.566    0.006  101.735    0.000
   .occ1              0.670    0.007  101.735    0.000
   .educ              0.738    0.007  101.735    0.000
```