Structural Equation Modelling & Causal Inference

Set-3: Repeated and Longitudinal Data

Ozan Aksoy

UCL Social Research Institute & NCRM
University College London

November 6, 2023

Topics

1. Notes
2. Multi-trait multi-method models (MTMM)
3. Measurement and structural invariance for matched samples
4. Dynamic panel models
5. Latent curve models (LCM)
6. Example
7. Extensions of LCM
8. FIML for missing data

Notes on repeated/longitudinal data

- Wide data: fewer rows, more columns: different variables for different time points
- Long data: fewer columns, more rows: same variable for all time points
- Multilevel Modeling: long data
- Longitudinal SEM: mostly wide data
- One may want to account that observations (or residuals) are correlated over time ...
- ... even more correlated the more proximate in time they are
- Note: comparisons across time not possible if only correlation matrices available

Multi-trait multi-method (MTMM)

Measurements xij crossed by *traits* (Ts) and by *methods* (Ms)

vars	M1	M2	M3	M4	...
T1	x11	x12	x13	x14	...
T2	x21	x22	x23	x24	...
...	...	...	...	...	...

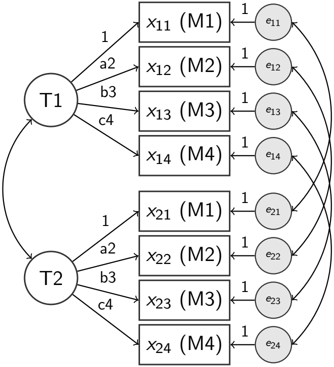
Examples

- classical: two traits (e.g., extraversion, political participation) each measured by 4 methods (e.g., interview, questionaire, expert judgement, peer judgement)
- 4 sexrole items (methods) measured at 2 time points (2 traits within individual)
- 4 sexrole items (methods) for husbands and wives (2 traits within household)

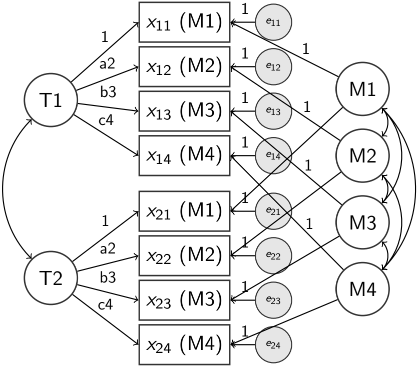
Two approaches to MTMM

- Common methods: introduce latent variables for methods Ms, uncorrelated with latent variables for traits Ts
  - item is measurement of a T and of a M
  - theoretically appealing
  - identification may be complicated
  - often inadmissible solutions
- Correlated uniqueness: correlate errors associated with common measurements

Correlated uniqueness vs common methods



correlated uniqueness



common method

Multi-trait multi-method (MTMM) in R

common methods

```
cm <-   'T1 =~ x11 + x12 + x13 + x14       # trait  1
         T2 =~ x21 + x22 + x23 + x24       # trait  2
         T3 =~ x31 + x32 + x33 + x34       # trait  3
         M1 =~ x11 + x21 + x31             # method 1
         M2 =~ x12 + x22 + x32             # method 2
         M3 =~ x13 + x23 + x33             # method 3
         M4 =~ x14 + x24 + x34             # method 4
         T1+T2+T3 ~~ p*M1+p*M2+p*M3+p*M4 # Ts,Ms uncorrelated
         p == 0'
```

correlated uniqueness

```
cu <-   'T1 =~ x11 + x12 + x13 + x14    # trait  1
         T2 =~ x21 + x22 + x23 + x24    # trait  2
         T3 =~ x31 + x32 + x33 + x34    # trait  3
         x11+x11+x21 ~~ x21+x31+x31     # Corr.s
         x12+x12+x22 ~~ x22+x32+x32
         x13+x13+x23 ~~ x23+x33+x33'
```

Measurement invariance for matched samples

- Previously: measurement invariance for independent samples (e.g., gender, ethnicity, country)
- Now: two (or more) dependent (paired) samples
- Examples of matched samples:
  - sexrole attitudes (or political participation) measured on husbands and wives (measurements nested in households)
  - sexrole attitudes (or political participation) measured on same people at two time-points (measurements nested in people)
- Failure to account for dependence similar to ignoring multilevel aspects of data

Measurement invariance for paired samples in R

For matches samples, we should impose/test these constraints:

- Strong invariance: Same loadings and same intercepts for latent variables
- If strong invariance does not hold, find partial invariance
- MTMM by correlated uniqueness
- Set the mean of the first latent variable mean to zero, free the mean(s) of other latent variable(s)

Measurement invariance matched pairs

```
mi <- '
hSR =~ hsr1 + l2*hsr2 + l3*hsr3 + l4*hsr4 # 4 items for h/w
wSR =~ wsr1 + l2*wsr2 + l3*wsr3 + l4*wsr4 # loadings equal
hsr1 + wsr1 ~ I1*1      # intercepts equal
hsr2 + wsr2 ~ I2*1
hsr3 + wsr3 ~ I3*1
hsr4 + wsr4 ~ I4*1
wSR         ~ 1         # mean of SR for wife's set free

hsr1 ~~ wsr1            # correlated uniqueness
hsr2 ~~ wsr2
hsr3 ~~ wsr3
hsr4 ~~ wsr4'
```

Testing for measurement invariance for paired samples

LRtesting for measurement invariance (MI) for paired samples:

- fit the constrained model C: impose MI (see previous slide)
  `output: chi2=193.190 df=21`
- fit the unconstrained model U: do not impose MI (see below)
  `output: chi2=106.371 df=15`
- compute LR test for C vs U:
  `manual: LR=193.190-106.371=86.9, df=21-15=6, p=.001`
  `non-manual: anova(f, g)`

no invariance

```
mv <- '
hSR =~ hsr1 + hsr2 + hsr3 + hsr4 # loadings free
wSR =~ wsr1 + wsr2 + wsr3 + wsr4 # to vary on h/w
hsr1 + wsr1 + hsr2 + wsr2 ~ 1    # intercepts free
hsr3 + wsr3 + hsr4 + wsr4 ~ 1
wSR ~ 0*1     # mean of SR at 0 for identification (default)

hsr1 ~~ wsr1  # still correlated uniqueness
hsr2 ~~ wsr2
hsr3 ~~ wsr3
hsr4 ~~ wsr4'
summary(g <- sem(mv, data=drole, missing = "ML"))
```

Partial measurement invariance for paired samples

- Strong invariance rejected, what should we do?
- Partial invariance: relax equality constraints on (some) intercepts, and (maybe) some loadings
- I made the following modifications
  - relax equality intercept hsr1/wsr1
  - relax equality intercept hsr2/wsr2
  - free covariance cov(e.hsr3,e.hsr4) and cov(e.wsr3,e.wsr4)
- The resulting model has reasonable fit (next slide)
- We may interpret/compare the latent variables in final model

```
   E(HSR)    = 0.000    E(WSR)    = -0.276
   VAR(HSR) = 0.734    VAR(WSR) =  0.674
   CORR(HSR,WSR) = 0.586
```

Partial strong measurement invariance in matched samples

Input

```
mp <- ' hSR =~ hsr1 + l2*hsr2 + l3*hsr3 + l4*hsr4 #equal
        wSR =~ wsr1 + l2*wsr2 + l3*wsr3 + l4*wsr4 #loadings
hsr1 + wsr1 ~ 1  # relax intercepts
hsr2 + wsr2 ~ 1  # relax intercepts
hsr3 + wsr3 ~ I3*1
hsr4 + wsr4 ~ I4*1
wSR         ~ 1
hsr3 ~~ hsr4     # add error cov
wsr3 ~~ wsr4     # add error cov
hsr1 ~~ wsr1     # still correlated uniqueness
hsr2 ~~ wsr2
hsr3 ~~ wsr3
hsr4 ~~ wsr4'
h <- sem(mp, data=drole, missing = "ML")
summary(h, fit.measures = TRUE, standardized = TRUE)
```

Selected and edited output

```
  Minimum Function Test Stat (df) [p]     41.171 (17) [0.001]
  Comparative Fit Index (CFI)                     0.989
  RMSEA (90% CI)                          0.031 (0.019  0.043)
  P-value RMSEA <= 0.05                           0.997
  SRMR                                            0.024
Covariances:
                Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  hSR ~~ wSR       0.412    0.040   10.429    0.000    0.586    0.586
Intercepts:
    wSR           -0.276    0.048   -5.801    0.000   -0.336   -0.336
    hSR            0.000                               0.000    0.000
Variances:
    hSR            0.734    0.061   12.001    0.000    1.000    1.000
    wSR            0.674    0.054   12.460    0.000    1.000    1.000
```

Explaining division-of-labour

Stylized causal diagram

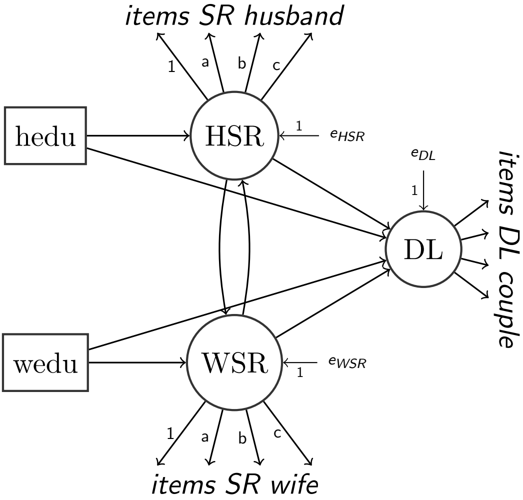
education of husband and wife (hedu, wedu) affect (latent) division of labour in household

(latent) sex role attitudes HSR WSR are (partial) mediators

interpersonal influence of attitudes

details about measurement (free cov's, MTMM) suppressed for ease of presentation

reasonable fit . . .



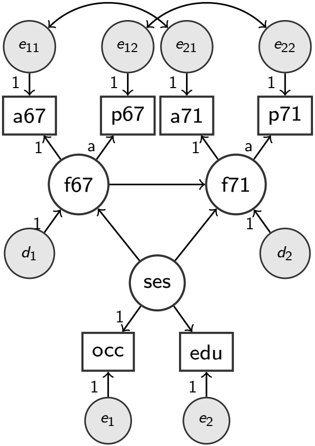
Explaining division-of-labour—R


```
mdl<-  '
# Measurement of sex roles with partial strong invariance
HSR =~ hsr1 + l2*hsr2 + l3*hsr3 + l4*hsr4
WSR =~ wsr1 + l2*wsr2 + l3*wsr3 + l4*wsr4
hsr1 + wsr1 ~ 1
hsr2 + wsr2 ~ 1
hsr3 + wsr3 ~ I3*1
hsr4 + wsr4 ~ I4*1
WSR         ~ 1
hsr3 ~~ hsr4
wsr3 ~~ wsr4

hsr1 ~~ wsr1     # still correlated uniqueness
hsr2 ~~ wsr2
hsr3 ~~ wsr3
hsr4 ~~ wsr4

#measurement of DL
DL   =~ dl1 + dl2 + dl3 + dl4
dl2  ~~ dl4

#structural part (non-recursive)
DL   ~ HSR + WSR + hedu + wedu
HSR  ~ WSR + hedu
WSR  ~ HSR + wedu
WSR ~~ HSR'

i <- sem(mdl, data=dlab, missing = "ML")
summary(i, fit.measures = TRUE, standardized = TRUE)
```

Explaining division-of-labour—R


```
  Number of observations                          1523
  Estimator                                         ML
  Minimum Function Test Statistic              169.216
  Degrees of freedom                                66
  P-value (Chi-square)                           0.000
  Comparative Fit Index (CFI)                    0.977
  RMSEA                                          0.032
  90 Percent Confidence Interval          0.026  0.038
  P-value RMSEA <= 0.05                          1.000
  SRMR                                           0.030
Regressions:
             Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  DL ~
    HSR        -0.247    0.054   -4.601    0.000   -0.243   -0.243
    WSR        -0.255    0.055   -4.643    0.000   -0.241   -0.241
    hedu       -0.030    0.013   -2.255    0.024   -0.034   -0.078
    wedu        0.082    0.015    5.344    0.000    0.094    0.203
  HSR ~
    WSR         0.987    0.002  409.451    0.000    0.946    0.946
    hedu       -0.025    0.012   -2.100    0.036   -0.029   -0.067
  WSR ~
    HSR         0.516    0.200    2.579    0.010    0.538    0.538
    wedu       -0.063    0.027   -2.310    0.021   -0.075   -0.163
```

Dynamic panel models–Wheaton on anomie and powerlessness as indicators of psychological disorder



Wheaton in R

wheaton input

```
mw <- '
F67 =~ anomie67 + p*pwless67    # same loadings
F71 =~ anomie71 + p*pwless71
anomie67 +     anomie71 ~ ma*1  # same intercepts
anomie67 ~~ anomie67
anomie71 ~~ anomie71

pwless67 +     pwless71 ~ mp*1  # same intercepts
pwless67 ~~ pwless67
pwless71 ~~ pwless71

F71 ~ 1                # mean of latent var free in 71

anomie67 ~~ anomie71   # MTMM--correlated uniqueness
pwless67 ~~ pwless71   # MTMM--correlated uniqueness

SES =~ occstat + educ  # reflexive measurement
F71 ~  F67 + SES       # structural part
F67 ~  SES'

fit.w <- sem(mw, sample.cov = w, sample.mean = m,
              sample.nobs = 932)
summary(fit.w, fit.measures = TRUE, standardized = TRUE)
```

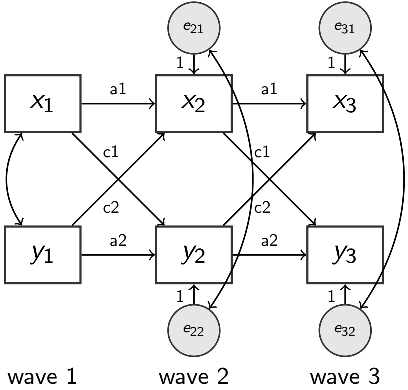
Wheaton in R

Wheaton selected/edited output

```
  Minimum Function Test Statistic               12.931
  Degrees of freedom                                 6
  P-value (Chi-square)                           0.044
  Comparative Fit Index (CFI)                    0.996
  RMSEA                                          0.035
  90 Percent Confidence Interval          0.005  0.062
  P-value RMSEA <= 0.05                          0.797
  SRMR                                           0.013
Latent Variables:
             Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  F67 =~
    anomie67     1.000                              2.707    0.782
    pwless67(p) 0.947    0.053   17.838    0.000    2.564    0.843
  F71 =~
    anomie71     1.000                              2.815    0.796
    pwless71(p) 0.947    0.053   17.838    0.000    2.666    0.842
  SES =~
    occstat      1.000                              1.361    0.203
    educ         1.915    0.555    3.447    0.001    2.607    0.841
Regressions:
  F71 ~
    F67          0.597    0.065    9.217    0.000    0.574    0.574
    SES         -0.413    0.139   -2.979    0.003   -0.200   -0.200
  F67 ~
    SES         -1.125    0.228   -4.939    0.000   -0.566   -0.566
Covariances:
 .anomie67 ~~
   .anomie71     1.637    0.322    5.076    0.000    1.637    0.354
 .pwless67 ~~
   .pwless71     0.289    0.267    1.082    0.279    0.289    0.103
```

Cross-lagged panel models

The causal feedback loop $x \Longleftrightarrow y$ assumes influence is stabilized (equilibrium). With panel-data, we may explain each var from the var at the previous wave. (Potential problem: var changes between the waves...). Preferably at least 3 waves



Cross-lagged design with time constant effects

Latent curve models (LCM)

- Panel data Y_{it} for subject i at time points $t = 0, 1, 2, \ldots$
- Research questions
  - income over the life course
  - development of cogitive ability, identity, attitudes . . . over time
  - increase in price over time
  - . . .
- Such data can be modelled in various ways
- Multilevel Modelling is one approach
- Latent Curve Modelling is an alternative

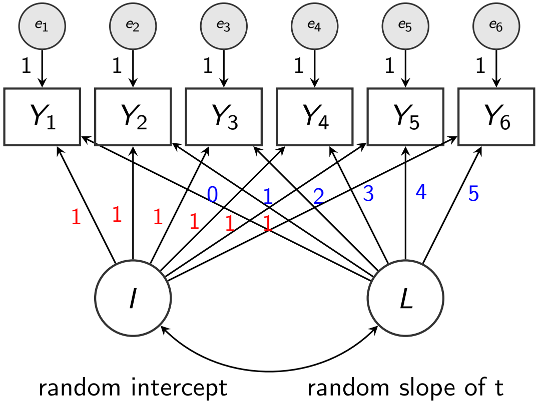
Multilevel modelling vs. LCM

- Multilevel modelling:

$$
\begin{aligned}
Y_{ti} &= \pi_{0i} + \pi_{1i} T_{ti} + \varepsilon_{ti} \\
\pi_{0i} &= \beta_{00} + \beta_{01} Z_i + u_{0i} \\
\pi_{1i} &= \beta_{10} + \beta_{11} Z_i + u_{1i}
\end{aligned}
$$

  - coefficients at the lowest (measurement/occasion) level: π
  - person level (second-level) coefficients: β
  - Y_{ti}: dependent variable for i measured at time t
  - T_{ti}: time variable, that indicates the time point (0,1,2...)
  - Z_i: time invariant (second-level/person-level) covariates
  - $\pi_\bullet \sim$ MVN(m,v)
- LCM: specify two correlated factors measured by Y's:
  - intercept I
  - linear slope L
  - loadings of all items Y_t on I are 1
  - loadings of items Y_t on L are 0, 1, 2, 3, 4, ...
  - intercepts of all items are 0, means of I,L free
  - regress I,L on Z

LCM—a path diagram



Multilevel modelling vs. LCM

- LCM requires data in wide format—in multilevel long
- LCM allows latent DVs and (time constant or time varying) latent IVs
- Multilevel allows only observed variables
- LCM allows flexible specicifation of level 1 error distribution: heteoskedasticity, autocorrelation, . . .
- Multilevel typically assumes homoskedasticity, no-autocorrelation, though relaxing some of these are possible

Example

- Example taken from Hox (2010)
- 200 students
- GPA: Grade Point Average for six successive semesters
- job: time-varying (semester-level) covariate: hours of work per week
- gender
- High-school GPA

Wide versus long data: Multilevel (`lmer`) wants long data LCM (`sem`) wants wide data

	student	sex	highgpa	GPA1	GPA2	GPA3	GPA4	GPA5	GPA6	JOB1	JOB2	JOB3	JOB4	JOB5	JOB6
1	1	2	2.8	2.3	2.1	3.0	3.0	3.0	3.3	2	2	2	2	2	2
2	2	1	2.5	2.2	2.5	2.6	2.6	3.0	2.8	2	3	2	2	2	2
3	3	2	2.5	2.4	2.9	3.0	2.8	3.3	3.4	2	2	2	3	2	2
4	4	1	3.8	2.5	2.7	2.4	2.7	2.9	2.7	3	2	2	2	2	2
5	5	1	3.1	2.8	2.8	2.8	3.0	2.9	3.1	2	2	2	2	2	2
6	6	2	2.9	2.5	2.4	2.4	2.3	2.7	2.8	2	3	3	2	3	3
7	7	1	2.3	2.4	2.4	2.8	2.6	3.0	3.0	3	2	3	2	2	2
8	8	2	3.9	2.8	2.8	3.1	3.3	3.3	3.4	2	2	2	2	2	2
9	9	1	2.0	2.8	2.7	2.7	3.1	3.1	3.5	2	2	3	2	2	2
10	10	1	2.8	2.8	2.8	3.0	2.7	3.0	3.0	2	2	2	3	2	2
11	11	2	3.9	2.6	2.9	3.2	3.6	3.6	3.8	2	3	2	2	2	2
12	12	2	2.9	2.6	3.0	2.3	2.9	3.1	3.3	3	2	2	2	2	2

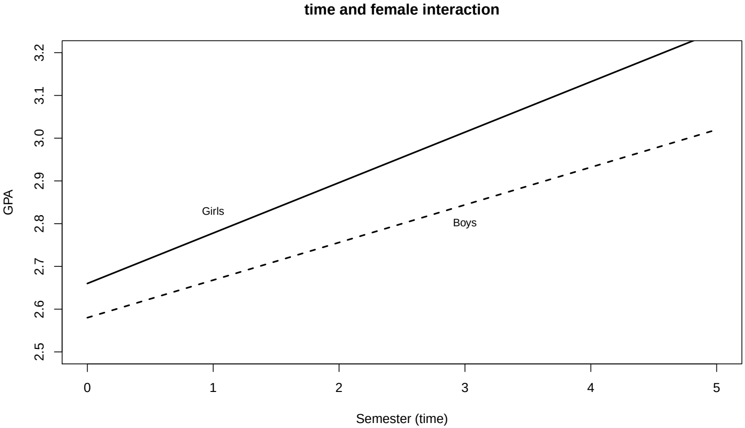
```
dlong <- reshape(data=dwide, varying = list(4:9, 10:15), timevar="time", times
                 v.names = c("gpa", "job"), direction="long", idvar="sid")
```

	row.names	student	sex	highgpa	time	gpa	job	sid
1	1.0	1	2	2.8	0	2.3	2	1
2	1.1	1	2	2.8	1	2.1	2	1
3	1.2	1	2	2.8	2	3.0	2	1
4	1.3	1	2	2.8	3	3.0	2	1
5	1.4	1	2	2.8	4	3.0	2	1
6	1.5	1	2	2.8	5	3.3	2	1
7	2.0	2	1	2.5	0	2.2	2	2
8	2.1	2	1	2.5	1	2.5	3	2
9	2.2	2	1	2.5	2	2.6	2	2
10	2.3	2	1	2.5	3	2.6	2	2
11	2.4	2	1	2.5	4	3.0	2	2

Multilevel model with cross-level interactions

```
m5 <- lmer(gpa ~ time+female + job + highgpa + (1+time|sid), data=dlong, REML = FALSE)
m8 <- lmer(gpa ~ time*female + job + highgpa + (1+time|sid), data=dlong, REML = FALSE)

screenreg(list(m5,m8), digits = 6)
============================================================
                            Model 1          Model 2
------------------------------------------------------------
(Intercept)                    2.557656 ***     2.581084 ***
                              (0.092100)       (0.092388)
time                           0.103373 ***     0.087829 ***
                              (0.005586)       (0.007951)
job                           -0.131119 ***    -0.132150 ***
                              (0.017264)       (0.017229)
highgpa                        0.088541 ***     0.088504 ***
                              (0.026280)       (0.026271)
femaleTRUE                     0.115670 ***     0.075506 *
                              (0.031300)       (0.034652)
time:femaleTRUE                                 0.029564 **
                                               (0.010958)
------------------------------------------------------------
AIC                          188.117446       182.971012
Log Likelihood               -85.058723       -81.485506
Var: sid (Intercept)           0.038234         0.037811
Var: sid time                  0.003837         0.003614
Cov: sid (Intercept) time     -0.002491        -0.002197
Var: Residual                  0.041542         0.041555
============================================================

(0.003837 - 0.003614)/0.003837
[1] 0.05811832 # prop. explained variance in slope of time by gender
```

Time and female interaction



LCM—basic setup in R


```
m <- 'I =~ 1*GPA1 + 1*GPA2 + 1*GPA3 + 1*GPA4 + 1*GPA5 + 1*GPA6
      L =~ 0*GPA1 + 1*GPA2 + 2*GPA3 + 3*GPA4 + 4*GPA5 + 5*GPA6
      # intercepts fixed at zero:
      GPA1 + GPA2 + GPA3 + GPA4 + GPA5 + GPA6 ~ 0*1
      L ~ NA*1  #means of L and I free
      I ~ NA*1
      I ~~ L    #cov between L,I free
    # GPA1 ~~ a*GPA1  #note: no homoskedasticity assumed
    # GPA2 ~~ a*GPA2  #to impose h.dicity add these
    # ...             # MLM assumes h.dicity
'
f <- sem(m, data=dwide)
```

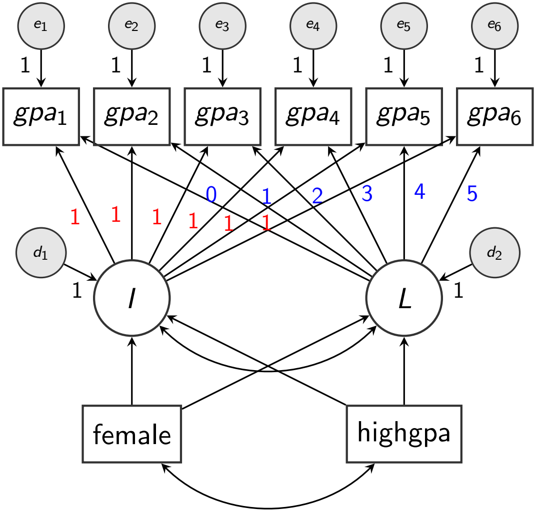
Equivalent compact syntax

```
c <- 'I =~ 1*GPA1 + 1*GPA2 + 1*GPA3 + 1*GPA4 + 1*GPA5 + 1*GPA6
      L =~ 0*GPA1 + 1*GPA2 + 2*GPA3 + 3*GPA4 + 4*GPA5 + 5*GPA6'
fc <- growth(mc, data=dwide)
```

LCM output in R


```
  Minimum Function Test Statistic                43.945
  Degrees of freedom                                 16
  P-value (Chi-square)                            0.000
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  I =~
    GPA1              1.000
    GPA2              1.000
    ...                ...
  L =~
    GPA1              0.000
    GPA2              1.000
    ...                ...
Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  I ~~
    L                 0.002    0.002    1.629    0.103
Intercepts:
                   Estimate  Std.Err  z-value  P(>|z|)
   .GPA1              0.000
   .GPA2              0.000
    ...                ...
    I                 2.598    0.018  141.956    0.000
    L                 0.106    0.005   20.338    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .GPA1              0.080    0.010    8.136    0.000
   .GPA2              0.071    0.008    8.799    0.000
   .GPA3              0.054    0.006    9.039    0.000
   .GPA4              0.029    0.003    8.523    0.000
   .GPA5              0.015    0.002    5.986    0.000
   .GPA6              0.016    0.003    4.617    0.000
    I                 0.035    0.007    4.947    0.000
    L                 0.003    0.001    5.645    0.000
```

LCM with time constant predictors—a path diagram



LCM—with time-constant predictors

Explaining differences in curves: regress latent variables I and L on subject level predictors male and highgpa, with correlated residuals of the 2 regressions.

```
L + I ~ female + highgpa
```

```
m2 <-'I =~ 1*GPA1 + 1*GPA2 + 1*GPA3 + 1*GPA4 + 1*GPA5 + 1*GPA6
      L =~ 0*GPA1 + 1*GPA2 + 2*GPA3 + 3*GPA4 + 4*GPA5 + 5*GPA6
      #intercepts zero:
      GPA1 + GPA2 + GPA3 + GPA4 + GPA5 + GPA6 ~ 0*1
      L ~ NA*1  # means of L and I free
      I ~ NA*1
      I ~~ L #cov between e(L),e(I) free
      L + I ~ female + highgpa # regressions
'
f2 <- sem(m2, data=dwide)
```

LCM—with time-varying predictors

GPA likely affected by time spent on paid work, measured by time-varying predictors (`job1-job6`).

- Add lines `gpa1 ∼ a*job1, ...`
- The `*a` spec ensures regressions with equal coefficients — the effect of working hours is time-constant.

```
m3 <-'I =~ 1*GPA1 + 1*GPA2 + 1*GPA3 + 1*GPA4 + 1*GPA5 + 1*GPA6
      L =~ 0*GPA1 + 1*GPA2 + 2*GPA3 + 3*GPA4 + 4*GPA5 + 5*GPA6
      GPA1 + GPA2 + GPA3 + GPA4 + GPA5 + GPA6 ~ 0*1
      L + I ~ NA*1  # intercepts of L and I free
      I ~~ L        # cov between e(L),e(I) free
      L + I ~ female + highgpa  # regressions
      GPA1 ~ a*JOB1             # time-varying predictors
      GPA2 ~ a*JOB2
      GPA3 ~ a*JOB3
      GPA4 ~ a*JOB4
      GPA5 ~ a*JOB5
      GPA6 ~ a*JOB6'
```

Selected oputput


```
 Estimator                                          ML
 Minimum Function Test Statistic               201.731
 Degrees of freedom                                 59
 P-value (Chi-square)                            0.000
Latent Variables:
                  Estimate  Std.Err  z-value  P(>|z|)
 I =~
   GPA1              1.000
   ...                ...
 L =~
   GPA1              0.000
   GPA2              1.000
   ...                ...
Regressions:
 L ~
   female            0.026    0.010    2.669    0.008
   highgpa          -0.003    0.008   -0.365    0.715
 I ~
   female            0.087    0.034    2.521    0.012
   highgpa           0.095    0.029    3.280    0.001
 GPA1 ~
   JOB1       (a)   -0.103    0.015   -7.073    0.000
 GPA2 ~
   JOB2       (a)   -0.103    0.015   -7.073    0.000
 GPA3 ~
   JOB3       (a)   -0.103    0.015   -7.073    0.000
   ...
Intercepts:
  .GPA1              0.000
   ...                ...
  .L                 0.099    0.026    3.805    0.000
  .I                 2.492    0.097   25.640    0.000
Variances:
  .GPA1              0.075    0.009    8.121    0.000
  .GPA2              0.069    0.008    8.855    0.000
  .GPA3              0.052    0.006    9.061    0.000
  .GPA4              0.030    0.003    8.645    0.000
  .GPA5              0.015    0.002    6.211    0.000
  .GPA6              0.014    0.003    4.425    0.000
  .I                 0.028    0.006    4.472    0.000
  .L                 0.003    0.001    5.275    0.000
```

LCM—with autocorrelated errors

Correlation between errors at *near* time points:

$$
\begin{array}{lcll}
\mathrm{cov}(e_{it}, e_{is}) & = & \sigma_1^2 & \text{if } \mid s-t \mid = 1 \\
\mathrm{cov}(e_{it}, e_{is}) & = & \sigma_2^2 < \sigma_1^2 & \text{if } \mid s-t \mid = 2 \\
\mathrm{cov}(e_{it}, e_{is}) & = & 0 & \text{if } \mid s-t \mid > 2
\end{array}
$$

```
m4 <-'I =~ 1*GPA1 + 1*GPA2 + 1*GPA3 + 1*GPA4 + 1*GPA5 + 1*GPA6
      L =~ 0*GPA1 + 1*GPA2 + 2*GPA3 + 3*GPA4 + 4*GPA5 + 5*GPA6
      GPA1 + GPA2 + GPA3 + GPA4 + GPA5 + GPA6 ~ 0*1
      L + I ~ NA*1
      I ~~ L
# Autocorrelations with 1-lag
GPA1 ~~ a*GPA2
GPA2 ~~ a*GPA3
GPA3 ~~ a*GPA4
GPA4 ~~ a*GPA5
GPA5 ~~ a*GPA6
# Autocorrelations with 2-lags
GPA1 ~~ b*GPA3
GPA2 ~~ b*GPA4
GPA3 ~~ b*GPA5
GPA4 ~~ b*GPA6'
```

LCM—quadratic time effect

Model: $y_{it} = \beta_{i0} + \beta_{i1} t + \beta_{i2} t^2 + e_{it} \qquad e_{it} \sim N(0, \sigma^2)$

```
m6 <-'I =~ 1*GPA1 + 1*GPA2 + 1*GPA3 + 1*GPA4 + 1*GPA5 + 1*GPA6
      L =~ 0*GPA1 + 1*GPA2 + 2*GPA3 + 3*GPA4 + 4*GPA5 + 5*GPA6
      Q =~ 0*GPA1 + 1*GPA2 + 4*GPA3 + 9*GPA4 +16*GPA5 +25*GPA6
      GPA1 + GPA2 + GPA3 + GPA4 + GPA5 + GPA6 ~ 0*1
      L + I + Q ~ NA*1
      I ~~ L + Q
      L ~~ Q'
f6 <- sem(m6, data=dwide)
```

Alternative test of linearity

- free loadings on L of gpa3 ... gpa6,
- test b(gpa3.L)=2, b(gpa4.L)=3, b(gpa5.L)=4, b(gpa6.L)=5

LCM—nonlinear time effect

input

```
m <- 'I =~ 1*GPA1 + 1*GPA2 + 1*GPA3 + 1*GPA4 + 1*GPA5 + 1*GPA6
      L =~ 0*GPA1 + 1*GPA2 + 2*GPA3 + 3*GPA4 + 4*GPA5 + 5*GPA6
      GPA1 + GPA2 + GPA3 + GPA4 + GPA5 + GPA6 ~ 0*1
      L + I ~ NA*1
      I ~~ L'
f <- sem(m, data=dwide)
x <- 'I =~ 1*GPA1 + 1*GPA2 + 1*GPA3 + 1*GPA4 + 1*GPA5 + 1*GPA6
      L =~ 0*GPA1 + 1*GPA2 +   GPA3 +   GPA4 +   GPA5 +   GPA6
      GPA1 + GPA2 + GPA3 + GPA4 + GPA5 + GPA6 ~ 0*1
      L + I ~ NA*1
      I ~~ L'
g <- sem(x, data=dwide)
anova(f, g)
```

output

```
  Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
g 12 123.91 173.38 39.084
f 16 120.77 157.05 43.945     4.8611       4     0.3019
```

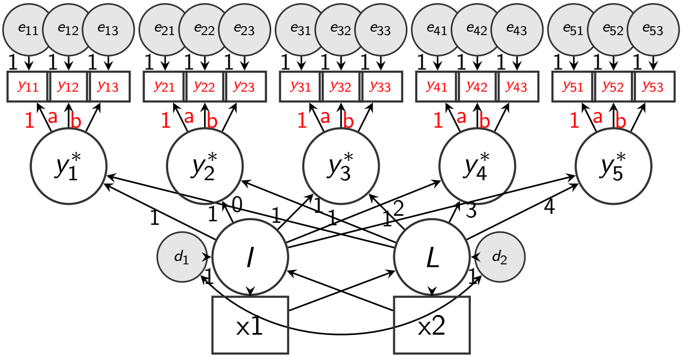
LCM—nonlinear time effect

Quadratic LCV continued

```
Latent Variables:
                     Estimate  Std.Err  z-value  P(>|z|)
  I =~
    GPA1                1.000
    GPA2                1.000
    ....                ....
  L =~
    GPA1                0.000
    GPA2                1.000
    GPA3                1.884    0.294    6.419    0.000
    GPA4                2.735    0.431    6.338    0.000
    GPA5                3.411    0.551    6.194    0.000
    GPA6                4.185    0.694    6.034    0.000
Intercepts:
   .GPA1                0.000
    ...                  ...
    L                   0.131    0.025    5.230    0.000
    I                   2.575    0.023  113.811    0.000
```

LCM with multiple indicators

- So far: growth analysis of observed variable
- Also possible: growth analysis of latent variable
- Need measurement invariance of latent variable over time
- MTMM correction can be added (thouhg I rarely see it)



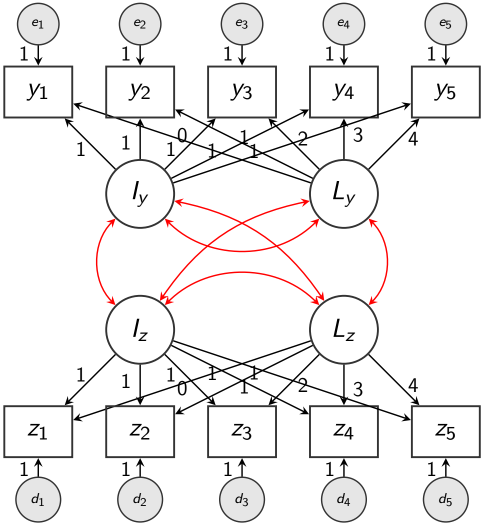
LCM with multiple indicators—Example

See Bollen & Curran 2006: Section 8.2

```
ly1 =~ l1*y11+l2*y12+l3*y13 # LY at time T=1
ly2 =~ l1*y21+l2*y22+l3*y23 # LY at time T=2
ly3 =~ l1*y31+l2*y32+l3*y33 # ...
ly4 =~ l1*y41+l2*y42+l3*y43 # same loadings at all T
ly5 =~ l1*y51+l2*y52+l3*y53

y11+y21+y31+y41+y51 ~ i1*1 ! same intercepts at all T
y12+y22+y32+y42+y52 ~ i2*1
y13+y23+y33+y43+y53 ~ i3*1


I =~ 1*ly1+1*ly2+1*ly3+1*ly4+1*ly5 # linear growth model
S =~ 0*ly1+1*ly2+2*ly3+3*ly4+4*ly5
ly1+ly2+ly3+ly4+ly5 ~ 0*1 # intercepts fixed at zero
S + I ~ NA*1               # means of I and S free
I ~~ S
I+L ~ x1+x2       # regress I L on exogenous vars
```

Parallel Latent Curve Models

See Bollen & Curran 2006: Chapter 7.4:



Missing data types

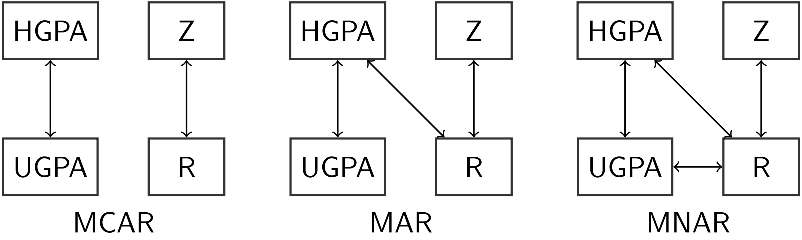
- Missing Completely at Random (MCAR) – ignorable
  - Missing data are distributed randomly
  - No variable in the dataset is correlated with an indicator of missingness
- Missing at Random (MAR) – ignorable
  - Missing data are not completely random
  - Indicator of missingness in Y is unrelated to Y but related to some X
- Missing Not at Random (MNAR) – non-ignorable
  - Missing data are not random
  - Indicator of missingness in Y is related to Y

Example: missing data types

	1st Year University GPA			
Highschool GPA	Complete	MCAR	MAR	MNAR
2.01	1.31	1.31	NA	NA
2.20	2.55	2.55	NA	2.55
2.30	1.90	NA	NA	NA
2.41	2.89	NA	NA	2.89
2.42	2.15	2.15	NA	NA
2.42	2.87	2.87	NA	2.87
2.47	2.33	2.33	2.33	2.33
2.64	1.56	1.56	1.56	NA
2.64	2.19	NA	2.19	NA
2.69	3.21	NA	3.21	3.21
2.77	2.39	2.39	2.39	2.39
2.83	2.40	2.40	2.40	2.40
2.85	2.53	2.53	2.53	2.53
2.87	2.23	2.23	2.23	2.23
3.02	2.86	NA	2.86	2.86
3.04	3.05	3.05	3.05	3.05
3.10	2.21	2.21	2.21	NA
3.10	3.36	NA	3.36	3.36
3.15	3.19	3.19	3.19	3.19
3.30	3.37	3.37	3.37	3.37

Missing data types: graphical representation

(HGPA: Highschool GPA, UGPA: 1st year Uni GPA, Z: some unmeasured variable, R: missing indicator)



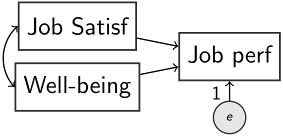
Traditional missing data methods

- Listwise deletion (default in most packages including R)
  - Analyze only cases which have complete data
- Pairwise deletion
  - Analyze pairs of cases that have complete data
- Mean imputation
  - Impute the arithmetic mean of X for all missings in X
- Regression imputation
  - Regress X on complete Z, predict missings
- Stochastic regression imputation
  - Regress X on complete Z, predict missings add some noise
- Hot-deck imputation
  - Impute missings using similar complete observations
- Last observation carried forward
  - Impute missings in wave t from the same respondent in wave t-1
- ...

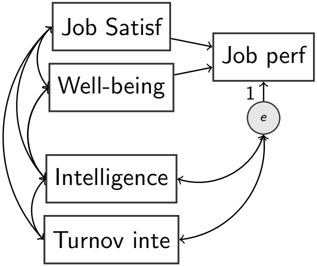
Summary FIML

- Listwise deletion maximizes: $logL_{complete}$
- FIML maximizes: $logL = logL_{complete} + logL_{incomplete}$
- FIML approach can be generalized to any other modelling (regressions etc.)
- ... though more complex maximization techniques may be needed for more complex models (e.g., EM)
- Accuracy of FIML can be improved further by using auxiliary variables

Example (data source: Enders (2010))



Model (FIML)



Saturated correlates with auxiliary vars (FIML)

input

```
mm <- 'jobperf ~ jobsat + wbeing
       jobperf ~1
       jobsat ~~ wbeing'
mma <- 'jobperf ~ jobsat + wbeing
        jobperf ~1
        jobsat ~~ wbeing
        iq + turnover ~~ jobperf + jobsat + wbeing
        iq ~~ turnover'
summary(f1 <- sem(mm, data = emp))    #listwise deletion
summary(f2 <- sem(mm, data = emp, missing = "ML")) #FIML
summary(f3 <- sem(mma, data = emp, missing = "ML")) #FIML with AV
```

edited output

```
# FIML wo Auxiliary ====================================
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  jobperf ~
    jobsat            0.027    0.060    0.444    0.657
    wbeing            0.476    0.055    8.665    0.000
Covariances:
  jobsat ~~
    wbeing            0.467    0.098    4.781    0.000
# FIML wi Auxiliary =====================================
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  jobperf ~
    jobsat            0.035    0.058    0.607    0.544
    wbeing            0.475    0.054    8.797    0.000
Covariances:
  jobsat ~~
    wbeing            0.442    0.096    4.613    0.000
 .jobperf ~~
    iq                3.010    0.439    6.862    0.000
  # ... [rest deleted for brevity]
```