The ‘Structural After Measurement’ (SAM) approach to SEM

Yves Rosseel
Department of Data Analysis
Ghent University – Belgium

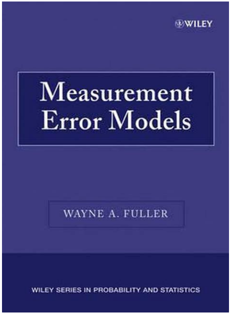
The Joint Quantitative Brownbag (JQBB) Speaker Series
25 March, 2024

acknowledgements

- Dr. Wen Wei Loh (Maastricht University)
- Dr. Sara Dhaene (UGent)

inspiration

Fuller, W.A. (1987). *Measurement Error Models*. Wiley.



the SAM paper

Rosseel, Y., & Loh, W.W. (2022). A structural after measurement approach to structural equation modeling. *Psychological Methods*
https://doi.org/10.1037/met0000503
https://osf.io/pekbm/ (includes original version)

related literature

Levy, R. (2023). Precluding interpretational confounding in factor analysis with a covariate or outcome via measurement and uncertainty preserving parametric modeling. *Structural Equation Modeling: A Multidisciplinary Journal.*
https://doi.org/10.1080/10705511.2022.2154214

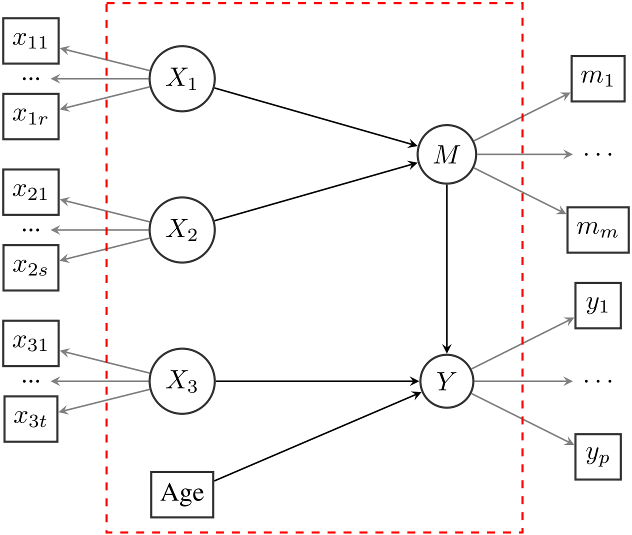
Bakk, Z., & Kuha, J. (2021). Relating latent class membership to external variables: An overview. *British Journal of Mathematical and Statistical Psychology*, 74(2), 340–362.
https://doi.org/10.1111/bmsp.12227 (Open Access)

Kuha, J., & Bakk, Z. (arxiv.org). Two-step estimation of latent trait models.
https://arxiv.org/pdf/2303.16101.pdf

the setting

- structural equation models
- we focus on 'large' models with many (say, > 100) parameters:
  - many constructs (motivation, ability, personality traits, ...)
  - each construct is measured by a set of (observed) indicators
  - many 'background' variables (age, gender, ...)
  - multilevel data, missing data, ...
- we are mostly interested in the structural part of the model:
  - regression model: variables are either dependent or independent
  - path analysis model: includes mediating effects, perhaps non-recursive
- assumption: the measurement instruments for the latent variables are well established, and usually fit (reasonably) well
- BUT: the sample size is not large (say, $N = 150$)

the setting



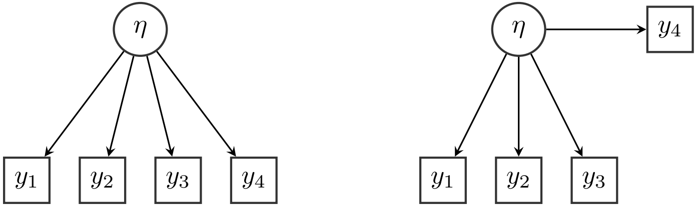
model parameters

- the majority of the model parameters are related to the measurement part of the model:
  - factor loadings
  - residual variances for the indicators
  - factor (co)variances
- a small portion of the model parameters are related to the structural part of the model:
  - regression coefficients
  - (residual) (co)variances

the standard estimation approach in SEM

- all parameters (measurement and stuctural) are estimated jointly: "system-wide" estimation
- frequentist: typically using an iterative optimization approach (e.g., ML)
- Bayesian: typically using MCMC
- advantages:
  - one-step
  - efficient (in terms of sampling variability)
  - inference is straightforward (standard errors, hypothesis testing)
  - (relatively) easy to handle constraints, missing data, ...
- works very well if the following conditions are met:
  - correctly specified model
  - large sample size
  - (normally distributed data)

example (stolen from Roy Levy, 2023)



- left panel: only measurement
- right panel: measurement + structural
- mathematically identical (in standard SEM)
- conceptually very different

generate some data

```
> library(lavaan)
> Sigma <- matrix(c(2.0, 1.0, 1.0, 1.0, 1.5,
                    1.0, 2.0, 1.0, 1.0, 0.1,
                    1.0, 1.0, 2.0, 1.0, 0.1,
                    1.0, 1.0, 1.0, 2.0, 0.1,
                    1.5, 0.1, 0.1, 0.1, 2.0), nrow = 5, ncol = 5)
> rownames(Sigma) <- colnames(Sigma) <- c("y1", "y2", "y3", "y4", "z")
> Sigma
```

```
    y1  y2  y3  y4   z
y1 2.0 1.0 1.0 1.0 1.5
y2 1.0 2.0 1.0 1.0 0.1
y3 1.0 1.0 2.0 1.0 0.1
y4 1.0 1.0 1.0 2.0 0.1
z  1.5 0.1 0.1 0.1 2.0
```

```
> set.seed(3)
> Data <- MASS::mvrnorm(n = 200L, mu = rep(0, 5), Sigma = Sigma)
```

R code left panel (model1)

```
> model1 <- '
      f =~ y1 + y2 + y3 + y4
 '
> fit1 <- sem(model1, data = Data)
> summary(fit1)

lavaan 0.6.18.2004 ended normally after 25 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                         8

  Number of observations                           200

Model Test User Model:

  Test statistic                                 0.140
  Degrees of freedom                                 2
  P-value (Chi-square)                           0.932

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured
```

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  f =~
    y1                1.000
    y2                0.996    0.123    8.122    0.000
    y3                0.894    0.116    7.689    0.000
    y4                0.911    0.120    7.608    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .y1                1.096    0.148    7.388    0.000
   .y2                0.837    0.127    6.601    0.000
   .y3                0.987    0.128    7.694    0.000
   .y4                1.081    0.138    7.820    0.000
    f                 1.031    0.208    4.959    0.000
```

R code right panel (model2)

```
> model2 <- '
     f =~ y1 + y2 + y3
     y4 ~ f
 '
> fit2 <- sem(model2, data = Data)
> summary(fit2)

lavaan 0.6.18.2004 ended normally after 26 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                         8

  Number of observations                           200

Model Test User Model:

  Test statistic                                 0.140
  Degrees of freedom                                 2
  P-value (Chi-square)                           0.932

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured
```

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  f =~
    y1                1.000
    y2                0.996    0.123    8.122    0.000
    y3                0.894    0.116    7.689    0.000

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  y4 ~
    f                 0.911    0.120    7.608    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .y1                1.096    0.148    7.388    0.000
   .y2                0.837    0.127    6.601    0.000
   .y3                0.987    0.128    7.694    0.000
   .y4                1.081    0.138    7.820    0.000
    f                 1.031    0.208    4.959    0.000
```

change outcome variable (y4 becomes z) (model3)

```
> model3 <- '
     f =~ y1 + y2 + y3
     z ~ f
 '
> fit3 <- sem(model3, data = Data)
> summary(fit3)

lavaan 0.6.18.2004 ended normally after 58 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                         8

  Number of observations                           200

Model Test User Model:

  Test statistic                                50.659
  Degrees of freedom                                 2
  P-value (Chi-square)                           0.000

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured
```

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  f =~
    y1                1.000
    y2                0.138    0.053    2.601    0.009
    y3                0.133    0.051    2.591    0.010

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  z ~
    f                 0.276    0.087    3.156    0.002

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .y1               -4.169    1.801   -2.315    0.021
   .y2                1.740    0.173   10.071    0.000
   .y3                1.699    0.168   10.086    0.000
   .z                 1.677    0.211    7.953    0.000
    f                 6.296    1.715    3.671    0.000
```

interpretational confounding

- replacing y_4 by z (= changing the structural part) also changes the parameters of the measurement model
- if the resulting parameters of the measurement model imply a different 'meaning' of the latent variable than was intended by the researcher, we have a problem
- this problem was coined "interpretational confounding" by Burt (1976)

Burt, R.S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods & Research, 5*(1), 3–52.

- Burt (1976) already suggested the solution: first fit the measurement part of the model, and then fit the structural part of the model

solution: replace sem() by sam()

```
> fit3.sam <- sam(model3, data = Data)
> parameterEstimates(fit3.sam, remove.step1 = FALSE, ci = FALSE,
                     output = "text")
```

```
Latent Variables:
                      Step Estimate  Std.Err  z-value  P(>|z|)
  f =~
    y1                   1    1.000
    y2                   1    0.969    0.139    6.977    0.000
    y3                   1    0.884    0.127    6.979    0.000

Regressions:
                      Step Estimate  Std.Err  z-value  P(>|z|)
  z ~
    f                    2    0.540    0.121    4.471    0.000

Variances:
                      Step Estimate  Std.Err  z-value  P(>|z|)
   .y1                   1    1.067    0.170    6.287    0.000
   .y2                   1    0.865    0.151    5.723    0.000
   .y3                   1    0.983    0.142    6.900    0.000
   .z                    2    1.846    0.195    9.462    0.000
    f                    2    1.059    0.226    4.678    0.000
```

the structural-after-measurement (SAM) approach

- SAM is an umbrella term to describe many different (estimation) approaches that have the following in common:
  - in the first step: we estimate the parameters related to the measurement part (factor loadings, residual variance of the indicators)
  - in the second step: we estimate the parameters related to the structural part (regression coefficients, residual (co)variances)
- the term SAM was used by Rosseel & Loh (2022), to avoid the overloaded terms ‘two-step’, ‘two-stage’, ...
- various SAM approaches have been suggested in the literature:
  - early references: Burt (1976), Hunter & Gerbing (1982), Lance, Cornwell & Mulaik (1988)
  - (uncorrected and bias-corrected) factor score regression (FSR)
  - SAM is the default approach in many other fields
- ...but they never received much attention in the SEM literature/community

critique on the SAM approach

- the (naive) standard errors in the second step are wrong (because they ignore the uncertainty that stems from the first step)
- in general: inference is (more) complicated
- multiple step methods are less efficient (more sampling variability)
- Fornell and Yi (1992) gave an example where a misspecified (but well-fitting) measurement model was embedded in a correctly specified structural model; but the model fit of the full model suggested that the model did not fit well, thus incorrectly implying a misspecified structural model
- software packages only allow for joint estimation
- ...
- to address (most of) these issues, we developed the sam() function in lavaan
- per default, sam() uses a method called 'local' SAM

local SAM: rationale

- the measurement model:

$$\boldsymbol{y} = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}$$

- to solve this for $\boldsymbol{\eta}$, we proceed as follows:

$$\begin{aligned}
\boldsymbol{\nu} + \boldsymbol{\Lambda}\,\boldsymbol{\eta} + \boldsymbol{\epsilon} &= \boldsymbol{y} \\
\boldsymbol{\Lambda}\,\boldsymbol{\eta} &= \boldsymbol{y} - \boldsymbol{\nu} - \boldsymbol{\epsilon} \\
\mathbf{M}\boldsymbol{\Lambda}\,\boldsymbol{\eta} &= \mathbf{M}\left[\boldsymbol{y} - \boldsymbol{\nu} - \boldsymbol{\epsilon}\right] \\
\boldsymbol{\eta} &= \mathbf{M}\left[\boldsymbol{y} - \boldsymbol{\nu} - \boldsymbol{\epsilon}\right]
\end{aligned}$$

  where $\mathbf{M}$ is $M \times P$ mapping matrix such that $\mathbf{M}\boldsymbol{\Lambda} = \mathbf{I}_M$

- we assume $\mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ and write $\mathrm{Var}(\boldsymbol{\epsilon}) = \boldsymbol{\Theta}$; it follows that

$$\begin{aligned}
\mathrm{E}(\boldsymbol{\eta}) &= \mathbf{M}\left[\mathrm{E}(\boldsymbol{y}) - \boldsymbol{\nu}\right] \\
\mathrm{Var}(\boldsymbol{\eta}) &= \mathbf{M}\left[\mathrm{Var}(\boldsymbol{y}) - \boldsymbol{\Theta}\right]\mathbf{M}^T
\end{aligned}$$

local SAM: first stage

- first stage: estimation of the measurement part of the model (only)
- this results in estimates for $\boldsymbol{\nu}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Theta}$
- M is the number of latent variables; B is the number of measurement 'blocks'
- three options:
  1. $B = 1$: single CFA
  2. $B = M$: as many 'blocks' as we have latent variables
  3. $B < M$: if some blocks are 'linked' together
- measurement models that are 'linked' (due to cross-loadings, correlated residuals, or equality constraints) should be fitted together
- measurement models that are 'weak' (low construct reliability) should also be fitted together in order to 'borrow strength' from each other
- for each block, we can use ML, GLS, $\ldots$, or we can we use noniterative estimators

local SAM: creating the mapping matrix M

- recall, the mapping matrix must chosen such that $\mathbf{M}\boldsymbol{\Lambda} = \mathbf{I}_M$
- three possible solutions for the mapping matrix $\mathbf{M}$:

$$\mathbf{M} = (\boldsymbol{\Lambda}^T\boldsymbol{\Theta}^{-1}\boldsymbol{\Lambda})^{-1}\,\boldsymbol{\Lambda}^T\boldsymbol{\Theta}^{-1} \qquad (ML)$$

$$\mathbf{M} = (\boldsymbol{\Lambda}^T\mathbf{S}^{-1}\boldsymbol{\Lambda})^{-1}\,\boldsymbol{\Lambda}^T\mathbf{S}^{-1} \qquad (GLS)$$

$$\mathbf{M} = (\boldsymbol{\Lambda}^T\boldsymbol{\Lambda})^{-1}\,\boldsymbol{\Lambda}^T \qquad (ULS)$$

- we then estimate $\mathrm{E}(\boldsymbol{\eta})$ and $\mathrm{Var}(\boldsymbol{\eta})$ as follows:

$$\widehat{\mathrm{E}(\boldsymbol{\eta})} = \hat{\mathbf{M}}\,[\bar{\boldsymbol{y}} - \hat{\boldsymbol{\nu}}]$$

$$\widehat{\mathrm{Var}(\boldsymbol{\eta})} = \hat{\mathbf{M}}\,[\mathbf{S} - \hat{\boldsymbol{\Theta}}]\,\hat{\mathbf{M}}^T$$

- in local SAM, we proceed in the second stage with these ‘sufficient statistics’ only

local SAM: second stage

- second stage: $\widehat{\mathrm{E}(\boldsymbol{\eta})}$ and $\widehat{\mathrm{Var}(\boldsymbol{\eta})}$ are used to estimate structural parameters of the model
- this can be done using 'path analysis', where we treat everything as observed, and the data is presented via summary statistics
- we can use ML, GLS, ...
- or we can use noniterative estimators: OLS (if the model is recursive) or TSLS (if the model is not recursive)

local SAM: further comments

- two-step corrected standard errors are available (seen Appendix C in the SAM paper)
- local fit measures only (for each measurement block, for the structural part)
- the (co)variance matrix of the latent variables is always positive definite
- we can handle missing data (fiml or two-stage), categorical indicators, twolevel data (random intercepts only), . . .
- but still some limitations (we plan to address these limitations in future work):
  - all indicators (of latent constructs) must be observed (i.e., no second-order measurement models)
  - the factor loading matrix ($\mathbf{\Lambda}$) must have full column rank
  - no support for (e.g., variance components) models where zeroes in the variance-covariance matrix of the latent variables are needed in order to identify the model

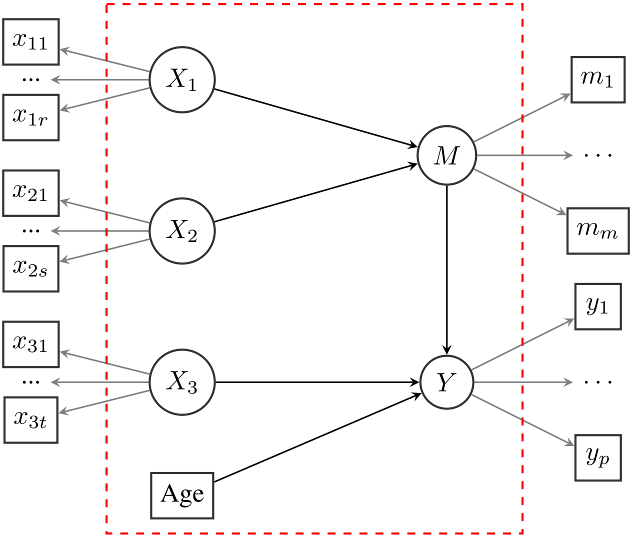
example: generate population data mediation model

```
> library(lavaan)
> pop.model <- '
     # factor loadings
     Y  =~ 1*y1 + 1.2*y2  + 0.8*y3  + 0.5*y4
     M  =~ 1*m1 + 0.5*m2  + 0.5*m3  + 0.7*m4
     X1 =~ 1*x1 + 0.7*x2  + 0.6*x3  + 1.1*x4
     X2 =~ 1*x5 + 0.7*x6  + 0.6*x7  + 0.9*x8
     X3 =~ 1*x9 + 0.7*x10 + 0.6*x11 + 1.1*x12

     # covariances among exogenous X1-X3
     X1 ~~ 0.4*X2; X1 ~~ -0.2*X3; X2 ~~ 0.4*X3

     # regression part
     Y ~  0.25*X3 + 0.4*M + (-0.1)*Age
     M ~ -0.30*X1 + 1.1*X2
 '
> set.seed(1234)
> Data <- simulateData(pop.model, sample.nobs = 200L, empirical = TRUE)
```

example: diagram



example: fitting the model using traditional SEM

```
> model <- '
     # measurement part
     Y  =~ y1 + y2  + y3  + y4
     M  =~ m1 + m2  + m3  + m4
     X1 =~ x1 + x2  + x3  + x4
     X2 =~ x5 + x6  + x7  + x8
     X3 =~ x9 + x10 + x11 + x12

     # structural part
     Y ~ X3 + M + Age
     M ~ X1 + X2
 '
> fit.sem <- sem(model, data = Data, estimator = "ML")
> parameterEstimates(fit.sem, ci = FALSE, output = "text")[21:25,]
```

```
Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  Y ~
    X3                0.250    0.108    2.308    0.021
    M                 0.400    0.079    5.078    0.000
    Age              -0.100    0.083   -1.203    0.229
  M ~
    X1               -0.300    0.133   -2.258    0.024
    X2                1.100    0.165    6.670    0.000
```

example: model-implied variance-covariance matrix latent variables

```
> lavInspect(fit.sem, "cov.lv")

        Y      M      X1     X2     X3
Y   1.498
M   0.939  2.036
X1  0.006  0.140  1.000
X2  0.492  0.980  0.400  1.000
X3  0.450  0.500 -0.200  0.400  1.000
```

example: fit measurement blocks, using $B = M$

```
> fit.Y  <- sem('Y  =~ y1 + y2  + y3  + y4',  data = Data)
> fit.M  <- sem('M  =~ m1 + m2  + m3  + m4',  data = Data)
> fit.X1 <- sem('X1 =~ x1 + x2  + x3  + x4',  data = Data)
> fit.X2 <- sem('X2 =~ x5 + x6  + x7  + x8',  data = Data)
> fit.X3 <- sem('X3 =~ x9 + x10 + x11 + x12', data = Data)

> # assemble Lambda and Theta
> Lambda <- matrix(0, 20, 5)
> Lambda[ 1:4,  1] <- lavInspect(fit.Y,  "est")$lambda
> Lambda[ 5:8,  2] <- lavInspect(fit.M,  "est")$lambda
> Lambda[ 9:12, 3] <- lavInspect(fit.X1, "est")$lambda
> Lambda[13:16, 4] <- lavInspect(fit.X2, "est")$lambda
> Lambda[17:20, 5] <- lavInspect(fit.X3, "est")$lambda

> Theta <- lav_matrix_bdiag(lavInspect(fit.Y,  "est")$theta,
                            lavInspect(fit.M,  "est")$theta,
                            lavInspect(fit.X1, "est")$theta,
                            lavInspect(fit.X2, "est")$theta,
                            lavInspect(fit.X3, "est")$theta)
```

example: compute ML version of the mapping matrix M

```
> Theta.inv <- solve(Theta)
> M <- solve(t(Lambda) %*% Theta.inv %*% Lambda) %*% t(Lambda) %*% Theta.inv

> # add age
> M     <- lav_matrix_bdiag(M,     matrix(1, nrow = 1L, ncol = 1L))
> Theta <- lav_matrix_bdiag(Theta, matrix(0, nrow = 1L, ncol = 1L))
> rownames(M) <- c("Y", "M", "X1", "X2", "X3", "Age")

> # compute (biased) sample covariance matrix 'S'
> N <- nrow(Data)
> S <- cov(Data) * (N - 1L)/N

> # compute Var(Eta)
> Var.eta <- M %*% (S - Theta) %*% t(M)
> round(Var.eta, 3)

         Y     M     X1    X2    X3  Age
Y    1.498 0.939  0.006 0.492  0.45 -0.1
M    0.939 2.036  0.140 0.980  0.50  0.0
X1   0.006 0.140  1.000 0.400 -0.20  0.0
X2   0.492 0.980  0.400 1.000  0.40  0.0
X3   0.450 0.500 -0.200 0.400  1.00  0.0
Age -0.100 0.000  0.000 0.000  0.00  1.0
```

example: second stage – using OLS

```
> # compute regression coefficients for M
> beta.M <- ( solve(Var.eta[c("X1", "X2"), c("X1", "X2")]) %*%
              Var.eta[c("X1", "X2"), "M", drop = FALSE] )
> round(beta.M, 3)

      M
X1 -0.3
X2  1.1
```

```
> # compute regression coefficients for Y
> beta.Y <- ( solve(Var.eta[c("X3", "M", "Age"), c("X3", "M", "Age")]) %*%
              Var.eta[c("X3", "M", "Age"), "Y", drop = FALSE] )
> round(beta.Y, 3)

        Y
X3   0.25
M    0.40
Age -0.10
```

example: using the sam() function

```
> fit.lsam <- sam(model = model, data = Data)
> parameterEstimates(fit.lsam, ci = FALSE, output = "text")[1:5,]

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)
  Y ~
    X3                0.250    0.109    2.301    0.021
    M                 0.400    0.080    4.971    0.000
    Age              -0.100    0.083   -1.203    0.229
  M ~
    X1               -0.300    0.133   -2.251    0.024
    X2                1.100    0.176    6.235    0.000
```

application 1: adding latent quadratic and interaction terms

- in the joint setting, adding latent quadratic/interaction terms is not trivial
- two popular methods are the product-indicator (PI) approach, and the so-called 'Latent Moderated Structural Equations' (LMS) approach
- none of these scale well: they cannot handle many quadratic and latent interaction terms simultaneously
- but if you can decouple the measurement and structural part, this becomes feasible
- a very general SAM solution (allowing for polynomial relations between latent variables) was already described in Wall & Amemiya (2000)
- the local SAM approach: find an explicit expression for

$$\mathrm{E}(\boldsymbol{\eta} \otimes \boldsymbol{\eta}) \qquad \text{and} \qquad \mathrm{Var}(\boldsymbol{\eta} \otimes \boldsymbol{\eta})$$

where $\otimes$ denotes the tensor (or Kronecker) product

implementation in lavaan

```
> model <- '
     # measurement part
     f1 =~ y1 + y2 + y3
     f2 =~ y4 + y5 + y6
     f3 =~ y7 + y8 + y9

     # structural part
     f3 ~ f1 + f2 + f1:f1 + f2:f2 + f1:f2
 '
> fit <- sam(model, data = Data, se = "none") # or se = "naive"
```

- no two-step analytic standard errors yet; but bootstrapping is possible
- forthcoming paper:

Rosseel, Y., Burghgraeve, E., Loh, W.W., Schermelleh-Engel, K. (accepted). Structural after Measurement (SAM) approaches for accommodating latent quadratic and interaction effects. *Behavior Research Methods*.

application 2: noniterative SEM

- for CFA, many noniterative estimators are available; some (i.e., the multiple group method) perform better than ML in terms of mean squared error

  Dhaene, S. & Rosseel, Y. (2023). An Evaluation of Non-Iterative Estimators in Confirmatory Factor Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*.

- we can use these noniterative estimators for the measurement part in SAM

  Dhaene, S., & Rosseel, Y. (2023). An Evaluation of Non-Iterative Estimators in the Structural after Measurement (SAM) Approach to Structural Equation Modeling (SEM). *Structural Equation Modeling: A Multidisciplinary Journal*, 30(6), 926–940

- “[the] local SAM approach outperforms traditional SEM in small to moderate samples (both in terms of convergence and MSE values), especially when reliability drops. ”

- BUT: no analytic standard errors yet (pseudo-ML or bootstrapping)

application 3: comparing structural relations across many groups

- reference:

  Perez Alonso, A.F., Rosseel, Y., Vermunt, J.K., & De Roover, K. (in press). Mixture Multigroup Structural Equation Modeling: A Novel Method for Comparing Structural Relations Across Many Groups. *Psychological Methods*.

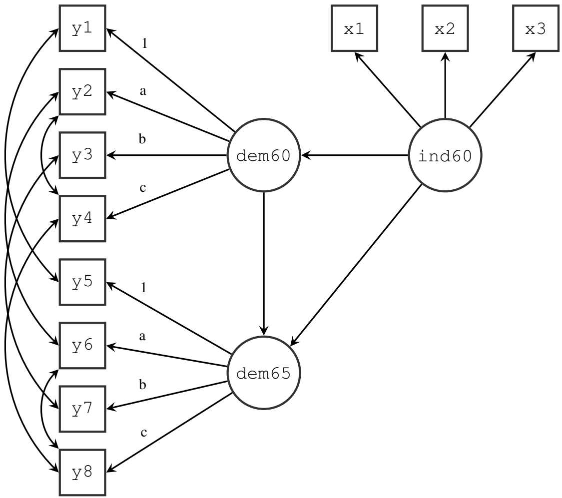
- relationships between latent variables are often different across groups (e.g., countries); but some groups may be similar in the sense that they have similar values for the regression coefficients
- we like to 'discover' these hidden clusters of similar groups
- in a first step, we estimated the measurement part across all groups (fixing the factor loadings to be the same across groups); this resulted in (model-implied) latent (co)variance matrices for all the groups
- in a second step, a mixture modeling approach is used to find homogeneous clusters that share similar regression coefficients

SAM: software implementation

- the SAM approach has been implemented in the sam() function in the R package lavaan
- available methods:
  - `sam.method = "local"` (default)
  - `sam.method = "global"`
  - `sam.method = "fsr"` (using Bartlett factor scores)
- typical call:

```
> fit.sam <- sam(model, data = PoliticalDemocracy,
                 sam.method = "local",
                 # link measurement blocks
                 mm.list = list(ind = "ind60", dem = c("dem60", "dem65")),
                 # measurement options
                 mm.args = list(estimator = "ML"),
                 # structural options
                 struc.args = list(estimator = "GLS"),
                 # global options
                 meanstructure = FALSE)
```

diagram of the Political Democracy model



sam() output

```
> summary(fit.sam, remove.step1 = FALSE)

This is lavaan 0.6.18.2004 -- using the SAM approach to SEM

  SAM method                                          LOCAL
  Mapping matrix M method                                ML
  Number of measurement blocks                            2
  Estimator measurement part                             ML
  Estimator  structural part                            GLS

  Number of observations                                 75

Summary Information Measurement + Structural:

  Block       Latent Nind Chisq Df
      1        ind60    3  0.00  0
      2 dem60,dem65    8 15.32 16

  Model-based reliability latent variables:

  ind60 dem60 dem65
  0.966 0.868  0.87

  Summary Information Structural part:

  chisq df cfi rmsea srmr
```

```
     0  0   1     0    0

Parameter Estimates:

  Standard errors                             Twostep
  Information                                 Expected
  Information saturated (h1) model          Structured

Latent Variables:
                       Step Estimate  Std.Err  z-value  P(>|z|)
  ind60 =~
    x1                    1    1.000
    x2                    1    2.193    0.142   15.403    0.000
    x3                    1    1.824    0.153   11.883    0.000
  dem60 =~
    y1                    1    1.000
    y2          (a)       1    1.213    0.143    8.483    0.000
    y3          (b)       1    1.210    0.125    9.690    0.000
    y4          (c)       1    1.273    0.122   10.453    0.000
  dem65 =~
    y5                    1    1.000
    y6          (a)       1    1.213    0.143    8.483    0.000
    y7          (b)       1    1.210    0.125    9.690    0.000
    y8          (c)       1    1.273    0.122   10.453    0.000

Regressions:
                       Step Estimate  Std.Err  z-value  P(>|z|)
  dem60 ~
```

```
    ind60             2    1.454    0.389    3.741    0.000
  dem65 ~
    ind60             2    0.558    0.225    2.480    0.013
    dem60             2    0.871    0.076   11.497    0.000

Covariances:
                   Step Estimate  Std.Err  z-value  P(>|z|)
 .y1 ~~
   .y5                1    0.577    0.364    1.585    0.113
 .y2 ~~
   .y4                1    1.390    0.685    2.030    0.042
   .y6                1    2.068    0.733    2.822    0.005
 .y3 ~~
   .y7                1    0.727    0.611    1.190    0.234
 .y4 ~~
   .y8                1    0.476    0.453    1.049    0.294
 .y6 ~~
   .y8                1    1.257    0.583    2.156    0.031

Variances:
                   Step Estimate  Std.Err  z-value  P(>|z|)
   .x1                1    0.084    0.020    4.140    0.000
   .x2                1    0.108    0.074    1.455    0.146
   .x3                1    0.468    0.091    5.124    0.000
   .y1                1    1.879    0.431    4.355    0.000
   .y2                1    7.530    1.363    5.523    0.000
   .y3                1    4.966    0.966    5.141    0.000
   .y4                1    3.214    0.722    4.449    0.000
```

```
  .y5                    1     2.499     0.518     4.824     0.000
  .y6                    1     4.809     0.924     5.202     0.000
  .y7                    1     3.302     0.699     4.722     0.000
  .y8                    1     3.227     0.720     4.482     0.000
   ind60                 2     0.446     0.087     5.135     0.000
  .dem60                 2     3.766     0.848     4.439     0.000
  .dem65                 2     0.189     0.224     0.843     0.399
```

discussion

- why should we decouple the measurement and structural part?
  - because we should (avoid interpretational confounding)
  - because we can (we can still do SEM)
  - this is what is done in most fields outside SEM
  - good performance in simulation studies (small/moderate sample sizes)
  - opens up modeling possibilities that were (computationally) difficult (if not impossible) in a joint estimation approach
- but still some obstacles:
  - analytic standard errors not always available (yet)
  - limitations (e.g., no higher-order measurement models)
  - more study is needed to discover potential weaknesses
- the SAM approach deserves the (renewed) interest of the SEM community

Thank you!

(questions?)

https://lavaan.org

https://lavaan.ugent.be/about/donate.html

https://jobs.ugent.be/job/
Ghent-Post-doctoral-researcher-9000/786525702/

references

Bakk, Z., & Kuha, J. (2021). Relating latent class membership to external variables: An overview. *British Journal of Mathematical and Statistical Psychology*, 74(2), 340–362.
https://doi.org/10.1111/bmsp.12227 (Open Access)

Burt, R.S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods & Research*, 5(1), 3–52.

Dhaene, S., & Rosseel, Y. (2023). An Evaluation of Non-Iterative Estimators in the Structural after Measurement (SAM) Approach to Structural Equation Modeling (SEM). *Structural Equation Modeling: A Multidisciplinary Journal*, 30(6), 926–940
https://doi.org/10.1080/10705511.2023.2220135 (Open Access)

Dhaene, S. & Rosseel, Y. (2023). An Evaluation of Non-Iterative Estimators in Confirmatory Factor Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*.
https://doi.org/10.1080/10705511.2023.2187285. (Open Access)

Fornell, C., & Yi, Y. (1992). Assumptions of the two-step approach to latent variable modeling. *Sociological Methods & Research*, 20(3), 291–320.

Fuller, W A. (1987). *Measurement error models*. John Wiley & Sons.

Hunter, J.E., & Gerbing, D. W. (1982). Unidimensional measurement, second-order factor analysis, and causal models. *Research in Organizational Behavior*, 4, 267–320.

Lance, C.E., Cornwell, J.M., & Mulaik, S.A. (1988). Limited information parameter estimates for latent or mixed manifest and latent variable models. *Multivariate Behavioral Research*, 23 (2), 171–187.

Levy, R. (2023). Precluding interpretational confounding in factor analysis with a covariate or outcome via measurement and uncertainty preserving parametric modeling. *Structural Equation Modeling: A Multidisciplinary Journal.* https://doi.org/10.1080/10705511.2022.2154214

Perez Alonso, A.F., Rosseel, Y., Vermunt, J.K., & De Roover, K. (in press). Mixture Multigroup Structural Equation Modeling: A Novel Method for Comparing Structural Relations Across Many Groups. *Psychological Methods*.

Rosseel, Y., & Loh, W.W. (2022). A structural after measurement approach to structural equation modeling. *Psychological Methods*
https://doi.org/10.1037/met0000503
https://osf.io/pekbm/ (includes original version)

Rosseel, Y., Burghgraeve, E., Loh, W.W., Schermelleh-Engel, K. (accepted). Structural after Measurement (SAM) approaches for accommodating latent quadratic and interaction effects. *Behavior Research Methods*.

Wall, M.M., & Amemiya, Y. (2000). Estimation for polynomial structural equation models. *Journal of the American Statistical Association*, 95(451), 929–940.