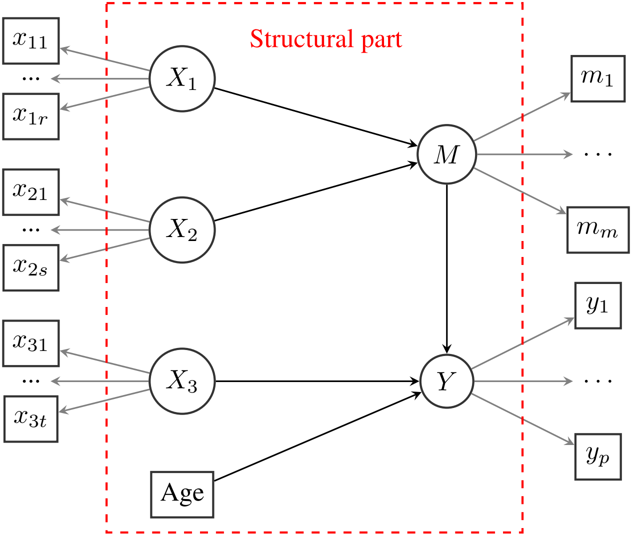
The ‘Structural After Measurement’ (SAM) approach: updates and extensions.

Yves Rosseel
Department of Data Analysis
Ghent University – Belgium

Meeting of the SEM Working Group – TU Chemnitz
20 March 2025

the setting


Structural part
x_{11}
...
x_{1r}
x_{21}
...
x_{2s}
x_{31}
...
x_{3t}
X_1
X_2
X_3
Age
M
Y
m_1
...
m_m
y_1
...
y_p

the setting (2)

- 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, binary/ordinal indicators, . . .
- we are mostly interested in the structural part of the model:
  - if not saturated: how well does the structural model fit?
  - size of direct/indirect effect, hypothesis testing
- **assumption: the measurement instruments for the latent variables are well established, and fit (reasonably) well**
- BUT: the sample size is not large (say, $N = 150$)

the standard estimation approach in SEM: ‘system-wide’ estimation

- all parameters (measurement and structural) are estimated jointly
- frequentist: typically using an iterative optimization approach (e.g., ML); Bayesian: typically using MCMC
- advantages:
  - one-step, and therefore 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, (multivariate normal data)
- but under less ideal circumstances, system-wide estimation does not (always) work well (bias, instability, nonconvergence, improper solutions, ...)
- in addition, a joint estimation approach potentially leads to interpretational confounding

the structural-after-measurement (SAM) approach

- SAM is an umbrella term to describe many different (estimation) approaches that have the following in common:
  - first step: we estimate the parameters related to the measurement part
  - second step: we estimate the parameters related to the structural part
- the term SAM was used by Rosseel & Loh (2024), to avoid the overloaded terms ‘two-step’, ‘two-stage’, …
- reference:

  Rosseel, Y., & Loh, W.W. (2024). A structural after measurement approach to structural equation modeling. *Psychological Methods, 29*(3), 561–588.
  https://doi.org/10.1037/met0000503
  https://osf.io/pekbm/ (includes original version)

- Rosseel & Loh (2024) proposed a special case: ‘local SAM’ (LSAM)

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: estimation

- first stage: estimation of the measurement part of the model (only)
- this results in estimates of:
  - $\mathrm{E}(\boldsymbol{\eta})$: the mean vector of the latent variables
  - $\mathrm{Var}(\boldsymbol{\eta})$: the variance-covariance matrix of the latent variables
  - $\Gamma(\boldsymbol{\eta})$ ('Gamma'): capturing the sampling variability of these sample statistics
- second stage: a regression or path analysis is performed, using the sample statistics of the latent variables as input
  - twostep-corrected standard errors and fit measures
  - we can use ML, GLS, ... or we can use noniterative estimators (OLS, TSLS)
- typical choice for the mapping matrix: the 'ML/Bartlett' matrix

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

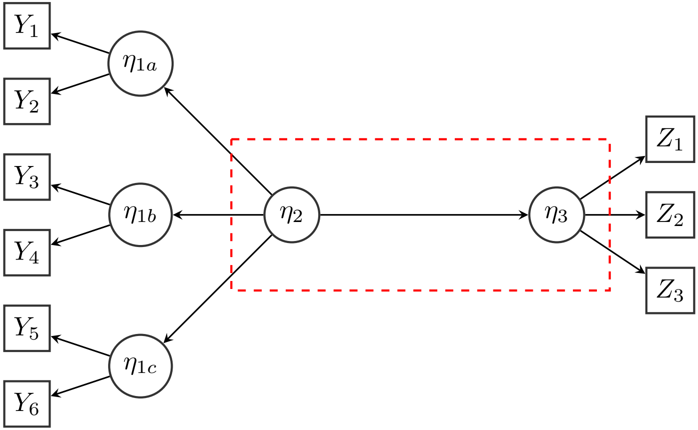
LSAM: limitations and missing features

- the goal of LSAM —and its implementation in the sam() function— is to be a plugin replacement for SEM (and therefore the sem() function)
- when we wrote the SAM paper (2021), LSAM had the following limitations:
  - all indicators of the latent variables should be observed (i.e., no higher-order factor models)
  - all indicators of the latent variables should be continuous (i.e., no binary/ordinal indicators)
  - the 'Lambda' matrix of factor loadings should not be rank deficient
  - there cannot be a priori fixed (to zero) elements in the unrestricted variance-covariance matrix of the latent variables
  - the structural model contains linear relations only (no interactions)
  - to compute two-step standard errors, we need to switch back to the global (measurement + structural) model

LSAM: limitations and missing features

- the goal of LSAM —and its implementation in the sam() function— is to be a plugin replacement for SEM (and therefore the sem() function)
- when we wrote the SAM paper (2021), LSAM had the following limitations:
  - **all indicators of the latent variables should be observed (i.e., no higher-order factor models)**
  - **all indicators of the latent variables should be continuous (i.e., no binary/ordinal indicators)**
  - the 'Lambda' matrix of factor loadings should not be rank deficient
  - there cannot be a priori fixed (to zero) elements in the unrestricted variance-covariance matrix of the latent variables
  - **the structural model contains linear relations only (no interactions)**
  - **to compute two-step standard errors, we need to switch back to the global (measurement + structural) model**

higher-order measurement models



higher-order measurement models

- the first-order factor model: $\boldsymbol{y} = \boldsymbol{\Lambda}_1 \boldsymbol{\eta}_1 + \boldsymbol{\epsilon}_1$
- the second-order factor model: $\boldsymbol{\eta}_1 = \boldsymbol{\Lambda}_2 \boldsymbol{\eta}_2 + \boldsymbol{\epsilon}_2$
- substituting the second in the first:

$$
\begin{aligned}
\boldsymbol{y} &= \boldsymbol{\Lambda}_1(\boldsymbol{\Lambda}_2 \boldsymbol{\eta}_2 + \boldsymbol{\epsilon}_2) + \boldsymbol{\epsilon}_1 \\
&= \boldsymbol{\Lambda}_1 \boldsymbol{\Lambda}_2 \boldsymbol{\eta}_2 + \boldsymbol{\Lambda}_1 \boldsymbol{\epsilon}_2 + \boldsymbol{\epsilon}_1 \\
&= \boldsymbol{\Lambda}^\star \boldsymbol{\eta}_2 + \boldsymbol{\epsilon}^\star
\end{aligned}
$$

- in the second-order case: $\boldsymbol{\Lambda}^\star = \boldsymbol{\Lambda}_1 \boldsymbol{\Lambda}_2$ and $\boldsymbol{\Theta}^\star = \boldsymbol{\Lambda}_1 \text{Var}(\boldsymbol{\epsilon}_2) \boldsymbol{\Lambda}_1^T + \boldsymbol{\Theta}_1$
- in general, we have $\boldsymbol{\Lambda}^\star = \boldsymbol{\Lambda}_1 (\mathbf{I} - \mathbf{B})^{-1}$ and

$$
\boldsymbol{\Theta}^\star = (\mathbf{I} - \mathbf{B})^{-1} \text{Var}(\boldsymbol{\epsilon}_{(-1)}) (\mathbf{I} - \mathbf{B})^{-1\prime} + \boldsymbol{\Theta}_1,
$$

  where $\mathbf{B}$ contains the ‘factor loadings’ of the higher-order factors (only)
- we now use $\boldsymbol{\Lambda}^\star$ and $\boldsymbol{\Theta}^\star$ when constructing the mapping matrix $\mathbf{M}$

binary or ordinal indicators in the measurement model

- for every measurement block, we use (D)WLS to estimate the model parameters
- but instead of using

$$\text{Var}(\boldsymbol{\eta}) = \mathbf{M}\left[\text{Var}(\boldsymbol{y}) - \boldsymbol{\Theta}\right]\mathbf{M}^T$$

  we now need to use

$$\text{Var}(\boldsymbol{\eta}) = \mathbf{M}\left[\text{Var}(\boldsymbol{y}^\star) - \boldsymbol{\Theta}\right]\mathbf{M}^T$$

  where $\text{Var}(\boldsymbol{y}^\star)$ is the matrix of polychoric (or tetrachoric, or polyserial) correlations
- warning: this only works if the *indicators* of the latent variables are binary/ordinal
- not working yet: the structural model contains (endogenous) binary/ordinal variables

adding latent quadratic and interaction terms

- basic idea: find an explicit expression for

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

```
> 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 = "bootstrap"
```

- paper:

  Rosseel, Y., Burghgraeve, E., Loh, W.W., Schermelleh-Engel, K. (2025). Structural after Measurement (SAM) approaches for accommodating latent quadratic and interaction effects. *Behavior Research Methods*. https://doi.org/10.3758/s13428-024-02532-y

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

work in progress: local two-step standard errors

- currently, to compute two-step corrected standard errors, we switch back to the global view (measurement + structural)
- in local SAM, we prefer to proceed (in the second step) with the sufficient statistics $\mathrm{E}(\boldsymbol{\eta})$ and $\mathrm{Var}(\boldsymbol{\eta})$ only
- to obtain two-step standard errors, we also need the 'Gamma' matrix $\boldsymbol{\Gamma}(\boldsymbol{\eta})$ capturing the sampling variability of these sufficient statistics
- then, to obtain local two-step standard errors for the second step parameters ($\boldsymbol{\theta}_2$), we can use the familiar (sandwich) formula for 'robust' standard errors:

$$\mathrm{Var}(\boldsymbol{\theta}_2) = \frac{1}{N} \left[(\Delta^T \mathcal{I}_1 \Delta)^{-1} (\Delta^T \mathcal{I}_1 \Gamma(\boldsymbol{\eta}) \mathcal{I}_1 \Delta) (\Delta^T \mathcal{I}_1 \Delta)^{-1} \right]$$

were $\mathcal{I}_1$ is the unit expected information matrix, and Δ is the Jacobian of the function that maps $\boldsymbol{\theta}_2$ to $\boldsymbol{\Sigma}(\boldsymbol{\theta}_2)$

how to obtain $\Gamma(\eta)$

- let $\mathbf{s}_y$ be the vector of sample statistics, containing the elements of $\bar{\boldsymbol{y}}$ and vech$[\mathbf{S}]$; the 'Gamma' matrix is then equal to N times the asymptotic variance matrix of the sample statistics: $\Gamma(\mathbf{s}_y) = N\text{Var}(\mathbf{s}_y)$
- let $\mathbf{s}_\eta$ be the vector of sample statistics that we use in the second step; that is the elements of $\text{E}(\boldsymbol{\eta})$ and vech$[\text{Var}(\boldsymbol{\eta})]$
- let f be a mapping function that takes as input the sample statistics ($\mathbf{s}_y$), and outputs $\mathbf{s}_\eta$; using to the Delta method, we have

$$\text{Var}(\mathbf{s}_\eta) = \Delta_f^T \text{Var}(\mathbf{s}_y)\Delta_f$$

  where Δ_f is the Jacobian of the mapping function f, and $\boldsymbol{\Gamma}(\boldsymbol{\eta}) = N\text{Var}(\mathbf{s}_\eta)$
- reference:

Jennrich, R.I. (2008). Nonparametric estimation of standard errors in covariance analysis using the infinitesimal jackknife. *Psychometrika, 73*(4), 579–594.

last slide

- eventually, sam() should be able to replace sem() in all circumstances
- more work is needed:
  - models where (by design) the ‘Lambda’ matrix of factor loadings is not rank deficient and/or where the variance-covariance matrix of the latent variables contains a priori fixed (to zero) values
  - categorical endogenous variables in the structural part
  - local standard errors for everything
  - more analytic work to explain why/when sam() and sem() give similar/different results
- other news:
  - lavaan now supports composites! (on GitHub only)
  - lavaan goes functional (see talk tomorrow by Marc Vidal)

Thank you!

(questions?)

https://lavaan.org

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