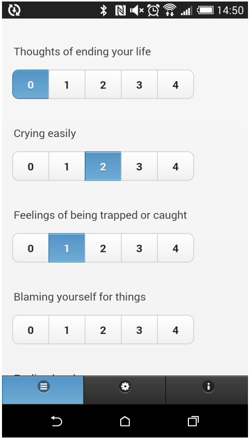
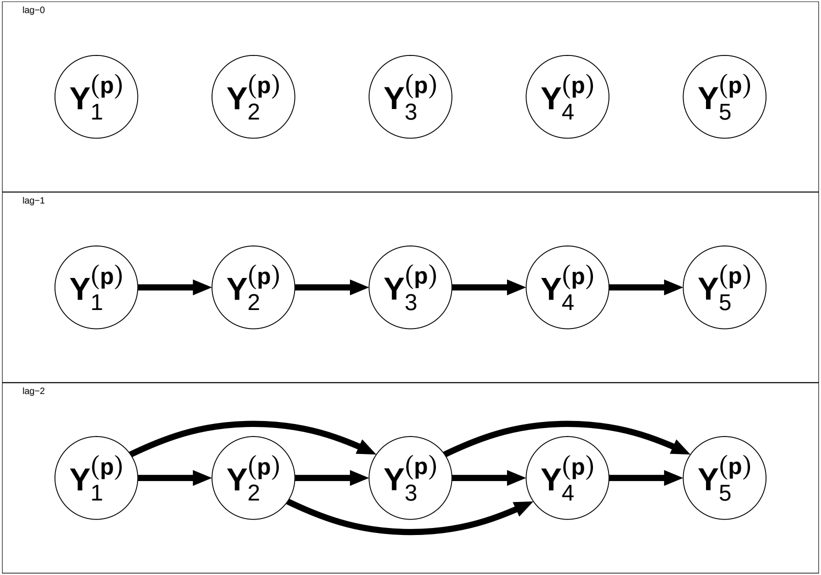
Time series

SEM 2 2020

Sacha Epskamp



- One person measured several times in a short period
- Cases can **not** reasonably be assumed to be *independent*
  - Knowing someone's level of fatigue at a time point helps predict his or her level of fatigue at the next time point.
- Likelihood not easy to compute without two assumptions:
  - The time-series factorize according to a graph
  - The model does not change over time


lag−0
$Y_1^{(p)}$ $Y_2^{(p)}$ $Y_3^{(p)}$ $Y_4^{(p)}$ $Y_5^{(p)}$
lag−1
$Y_1^{(p)}$ $Y_2^{(p)}$ $Y_3^{(p)}$ $Y_4^{(p)}$ $Y_5^{(p)}$
lag−2
$Y_1^{(p)}$ $Y_2^{(p)}$ $Y_3^{(p)}$ $Y_4^{(p)}$ $Y_5^{(p)}$

Suppose the first five measurements of a time-series dataset with three variables are:

$$\boldsymbol{Y} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{bmatrix} 6 & 6 & 7 \\ 5 & 5 & 5 \\ 5 & 5 & 2 \\ 2 & 2 & 2 \\ 1 & 1 & 3 \end{bmatrix} \end{array}$$

Suppose the first five measurements of a time-series dataset with three variables are:

$$\boldsymbol{Y} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{bmatrix} 6 & 6 & 7 \\ 5 & 5 & 5 \\ 5 & 5 & 2 \\ 2 & 2 & 2 \\ 1 & 1 & 3 \end{bmatrix} \end{array}$$

If we want to properly model the likelihood, we need to take into account that *all* cases are correlated. This can be done by modeling $\mathrm{vec}\left(\boldsymbol{Y}^\top\right)$, of which we have one case:

$$\mathrm{vec}\left(\boldsymbol{Y}^\top\right)^\top = \left[\begin{array}{ccccccccccccccc} y_{11} & y_{12} & y_{13} & y_{21} & y_{22} & y_{23} & y_{31} & y_{32} & y_{33} & y_{41} & y_{42} & y_{43} & y_{51} & y_{52} & y_{53} \\ 6 & 6 & 7 & 5 & 5 & 5 & 5 & 5 & 2 & 2 & 2 & 2 & 1 & 1 & 3 \end{array} \right]$$

Suppose the first five measurements of a time-series dataset with three variables are:

$$
\boldsymbol{Y} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{bmatrix} 6 & 6 & 7 \\ 5 & 5 & 5 \\ 5 & 5 & 2 \\ 2 & 2 & 2 \\ 1 & 1 & 3 \end{bmatrix} \end{array}
$$

If we want to properly model the likelihood, we need to take into account that *all* cases are correlated. This can be done by modeling $\operatorname{vec}\left(\boldsymbol{Y}^\top\right)$, of which we have one case:

$$
\operatorname{vec}\left(\boldsymbol{Y}^\top\right)^\top = \begin{bmatrix} \overset{y_{11}}{6} & \overset{y_{12}}{6} & \overset{y_{13}}{7} & \overset{y_{21}}{5} & \overset{y_{22}}{5} & \overset{y_{23}}{5} & \overset{y_{31}}{5} & \overset{y_{32}}{5} & \overset{y_{33}}{2} & \overset{y_{41}}{2} & \overset{y_{42}}{2} & \overset{y_{43}}{2} & \overset{y_{51}}{1} & \overset{y_{52}}{1} & \overset{y_{53}}{3} \end{bmatrix}
$$

The variance–covariance structure of this vector then is highly structured, and should include all implied covariances (lag-1, lag-2, lag-3...). Computing such a model is typically too complex!

Trick: rather than modeling $\boldsymbol{y}_t$, model the vector $\boldsymbol{z}^\top = \begin{bmatrix} \boldsymbol{y}_{t-1} & \boldsymbol{y}_t \end{bmatrix}$, and treat these as independent cases:

$$\boldsymbol{Z} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} y_{t-1,1} & y_{t-1,2} & y_{t-1,3} & y_{t,1} & y_{t,2} & y_{t,3} \end{array} \\ \begin{bmatrix} \cdot & \cdot & \cdot & 6 & 6 & 7 \\ 6 & 6 & 7 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 & 5 & 2 \\ 5 & 5 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 1 & 1 & 3 \end{bmatrix} \end{array}$$

Trick: rather than modeling $\boldsymbol{y}_t$, model the vector $\boldsymbol{z}^\top = \begin{bmatrix}\boldsymbol{y}_{t-1} & \boldsymbol{y}_t\end{bmatrix}$, and treat these as independent cases:

$$\boldsymbol{Z} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} y_{t-1,1} & y_{t-1,2} & y_{t-1,3} & y_{t,1} & y_{t,2} & y_{t,3} \end{array} \\ \begin{bmatrix} \cdot & \cdot & \cdot & 6 & 6 & 7 \\ 6 & 6 & 7 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 & 5 & 2 \\ 5 & 5 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 1 & 1 & 3 \end{bmatrix} \end{array}$$

The variance–covariance structure then is a *Toeplitz matrix*:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{t-1,t-1} & \boldsymbol{\Sigma}_{t-1,t} \\ \boldsymbol{\Sigma}_{t,t-1} & \boldsymbol{\Sigma}_{t,t} \end{bmatrix}$$

Which is typically used to fit SEM models. Two downsides of this method:

- Not the *true* likelihood
- Lag factorization now is assumption, not tested!

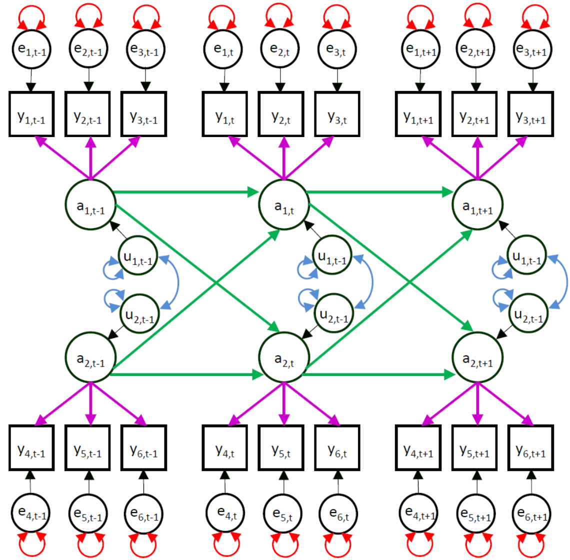
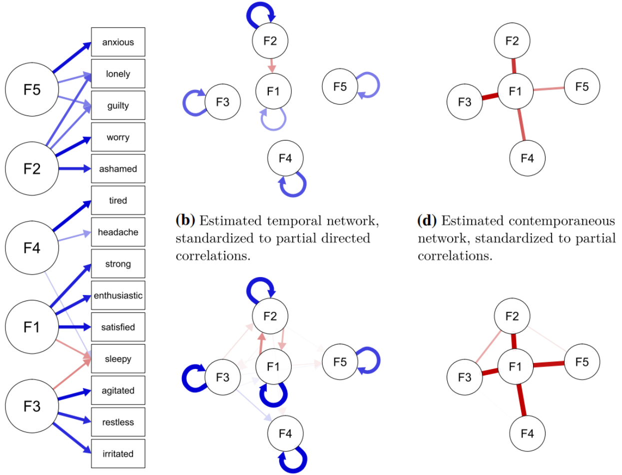


Figure by Ellen Hamaker



(a) Estimated factor loadings.

(b) Estimated temporal network, standardized to partial directed correlations.

(d) Estimated contemporaneous network, standardized to partial correlations.

(c) Case-drop bootstrap (25%) inclusion proportions of edges in the temporal network.

(e) Case-drop bootstrap (25%) inclusion proportions of edges in the contemporaneous network.

Epskamp, S. (2020). Psychometric Network Models from Time-series and Panel Data. *Psychometrika*. https://doi.org/10.1007/s11336-020-09697-3