Introduction to Continuous-Time VAR(1) models

Oisín Ryan

Department of Methodology and Statistics, Utrecht University

Summary

1. Recap of the (discrete-time) VAR(1) model
2. The problem of time-interval dependency
3. Continuous-Time VAR(1) models
4. Mediation, causal mechanisms, and Continuous-Time
5. Estimation options

The VAR(1) Model

$$\boldsymbol{X}_\tau = \boldsymbol{c} + \boldsymbol{\Phi}\boldsymbol{X}_{\tau-1} + \epsilon_\tau$$

where

- $\boldsymbol{c}$ is the $p \times 1$ column vector of intercepts
- $\boldsymbol{\Phi}$ is the $p \times p$ matrix of auto-regressive and cross-lagged effects
- $\epsilon_\tau \sim MV\mathcal{N}(0, \boldsymbol{\sigma}^2)$
- $\boldsymbol{\sigma}^2$ is a $p \times p$ diagonal matrix
- Subscript τ denotes measurement occassion

The VAR(1) Model

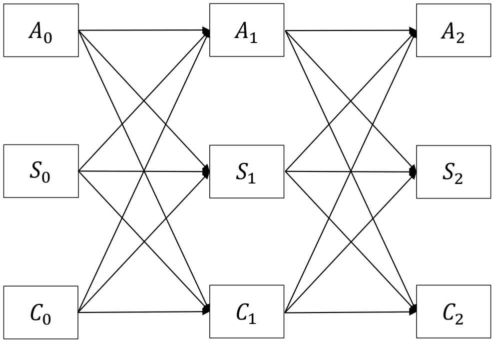
$$\boldsymbol{X}_\tau = \boldsymbol{c} + \boldsymbol{\Phi}\boldsymbol{X}_{\tau-1} + \epsilon_\tau$$

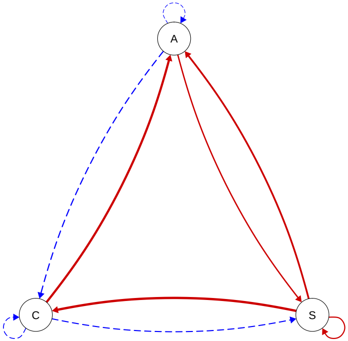
where

- $\boldsymbol{c} + \boldsymbol{\Phi}\boldsymbol{X}_{\tau-1}$ is referred to as the **structural part**
- ϵ_τ is referred to as the **stochastic innovation**

The VAR(1) Model

$$\boldsymbol{X}_\tau = \boldsymbol{c} + \boldsymbol{\Phi}\boldsymbol{X}_{\tau-1} + \boldsymbol{\epsilon}_\tau$$





The VAR(1) Model

$$\boldsymbol{X}_\tau = \boldsymbol{c} + \boldsymbol{\Phi}\boldsymbol{X}_{\tau-1} + \epsilon_\tau$$

Some simplifications I will make without loss of generality:

- Single-subject case: no i subscript
- Let $\boldsymbol{c}$ be a vector of zero's - each X_t is centered around its stable mean

The VAR(1) Model

$$\boldsymbol{X}_\tau = \boldsymbol{\Phi}\boldsymbol{X}_{\tau-1} + \boldsymbol{\epsilon}_\tau$$

Some simplifications I will make without loss of generality:

- Single-subject case: no i subscript
- Let $\boldsymbol{c}$ be a vector of zero's - each X_τ is centered around its stable mean

The VAR(1) Model

$$\boldsymbol{X}_\tau = \boldsymbol{\Phi}\boldsymbol{X}_{\tau-1} + \epsilon_\tau$$

Key Assumptions of the VAR(1) model:

- Stationarity - constant mean and variance
  - Parameters do not change during observation period
- Stable, mean reverting process
  - Eigenvalues of $\boldsymbol{\Phi}$ fall within $(-1, 1)$
- Equal time-intervals between measurements

The VAR(1) Model

$$\boldsymbol{X}_{\tau} = \boldsymbol{\Phi}\boldsymbol{X}_{\tau-1} + \epsilon_{\tau}$$

Key Assumptions of the VAR(1) model:

- Stationarity - constant mean and variance
  - Parameters do not change during observation period
- Stable, mean reverting process
  - Eigenvalues of $\boldsymbol{\Phi}$ fall within $(-1, 1)$
- **Equal time-intervals between measurements**

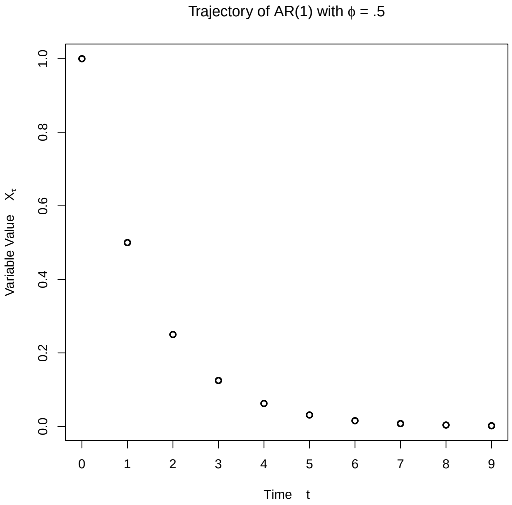
The time-interval dependency problem

Long recognized in the time-series literature (cf Gollob & Reichardt, 1987).

Refers to the phenomenon that:

- $\mathbf{\Phi}$ differs depending on the time-interval between measurements Δt
- $\mathbf{\Phi} \rightarrow \mathbf{\Phi}(\Delta t)$

The time-interval dependency problem



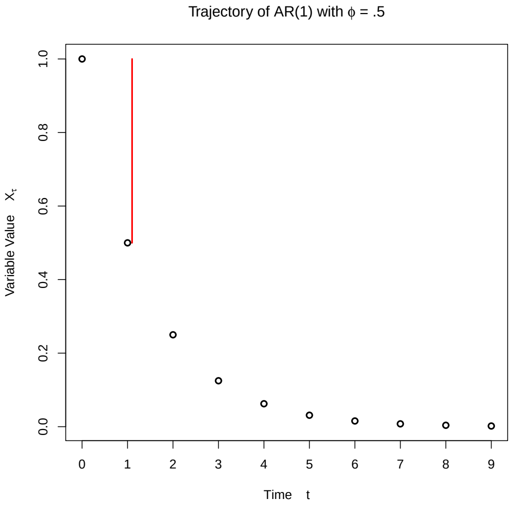
Take

$$E(X_\tau) = \phi X_{\tau-1}$$

with

- $\phi = .5$
- initial value $X_0 = 1$
- $\Delta t = 1$

The time-interval dependency problem



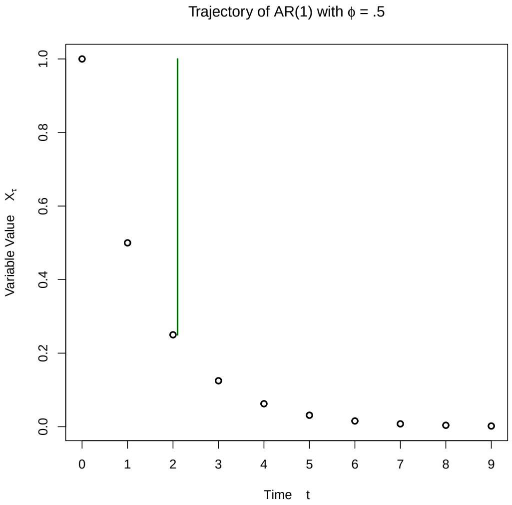
Take

$$E(X_\tau) = \phi X_{\tau-1}$$

with

- $\phi = .5$
- initial value $X_0 = 1$
- $\Delta t = 1$

The time-interval dependency problem



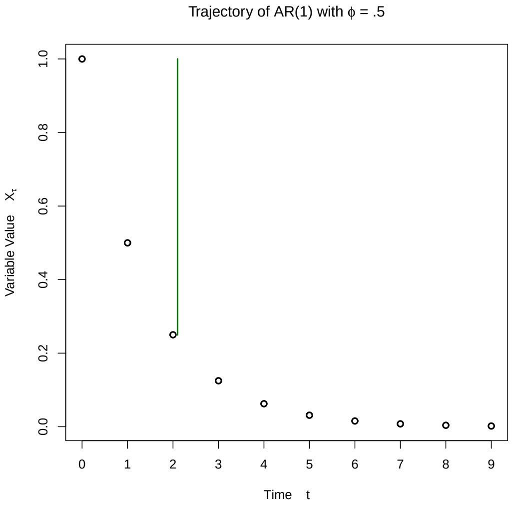
Take

$$E(X_\tau) = \phi.\phi X_{\tau-2}$$

with

- $\phi = .5$
- initial value $X_0 = 1$
- $\Delta t = 1$

The time-interval dependency problem



Take

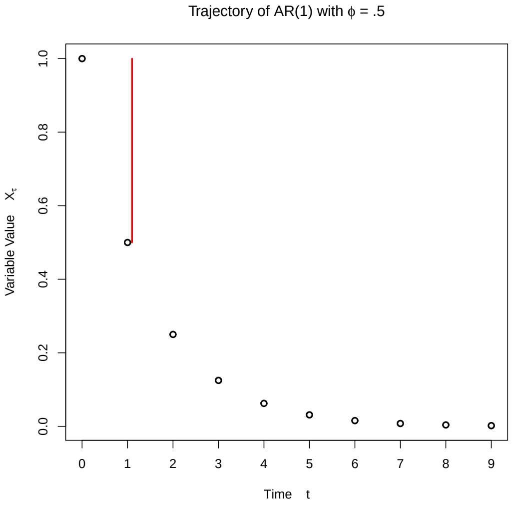
$$E(X_\tau) = \phi^2 X_{\tau-2}$$

with

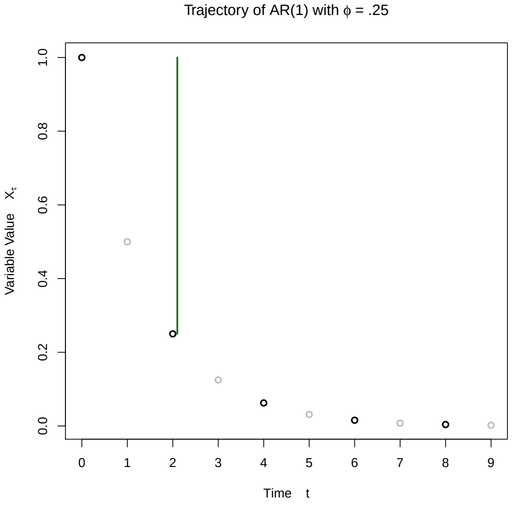
- $\phi = .5$
- initial value $X_0 = 1$
- $\Delta t = 1$

The time-interval dependency problem

$$E(X_\tau) = \phi(\Delta t = 1)X_{\tau-1}$$

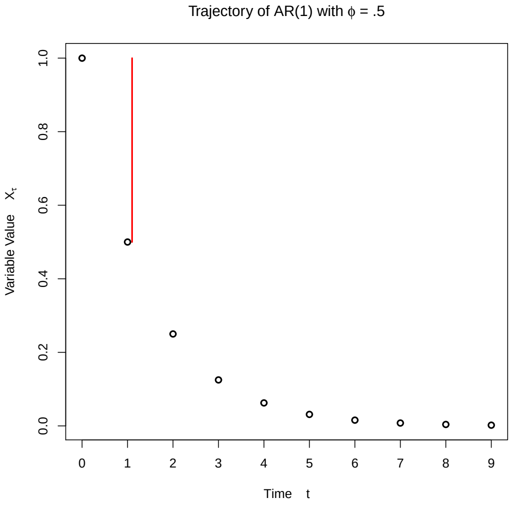


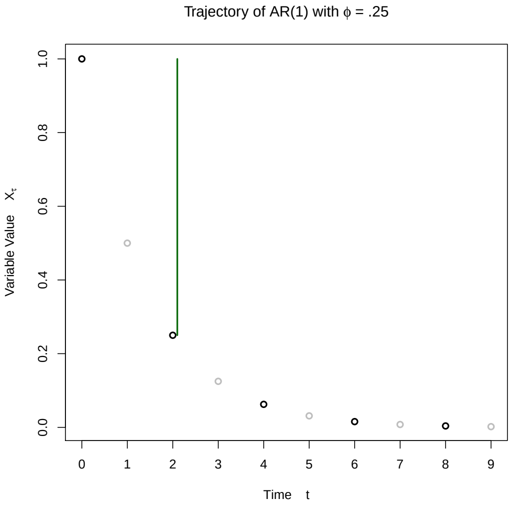
$$E(X_\tau) = \phi(\Delta t = 2)X_{\tau-1}$$



The time-interval dependency problem

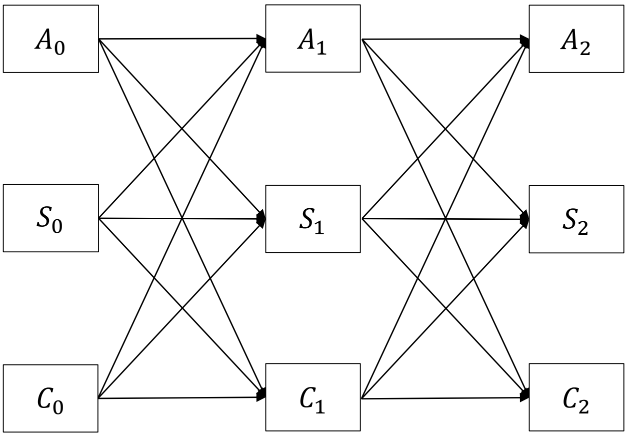
$$\phi(\Delta t = 1)^2 = \phi(\Delta t = 2)$$





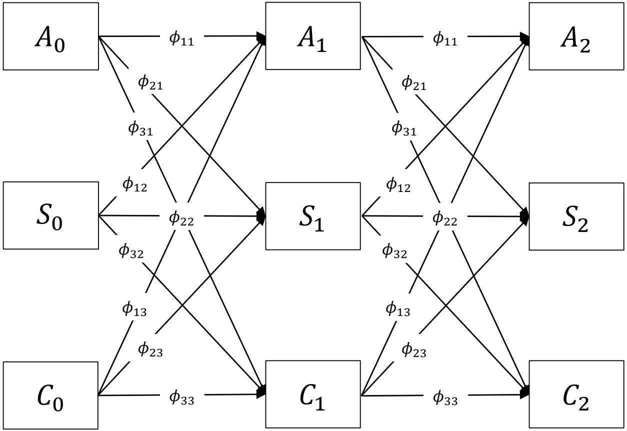
Time-interval dependency: Multivariate case Cole & Maxwell (2003); Reichardt (2011)

$$\boldsymbol{X}_\tau = \boldsymbol{\Phi}(\Delta t = 1)\boldsymbol{X}_{\tau-1} + \boldsymbol{\epsilon}_\tau$$



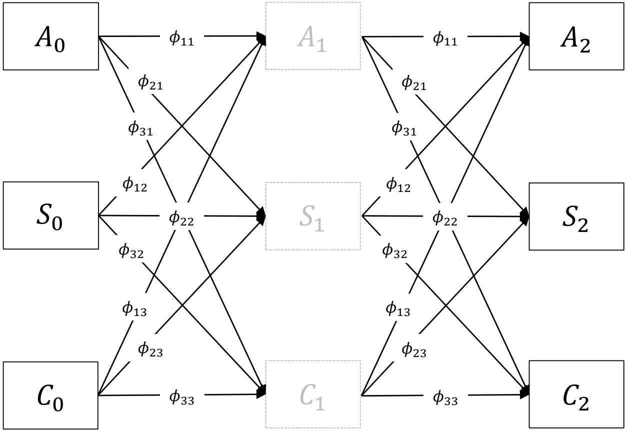
Time-interval dependency: Multivariate case Cole & Maxwell (2003); Reichardt (2011)

$$\boldsymbol{X}_\tau = \boldsymbol{\Phi}(\Delta t = 1)\boldsymbol{X}_{\tau-1} + \boldsymbol{\epsilon}_\tau$$



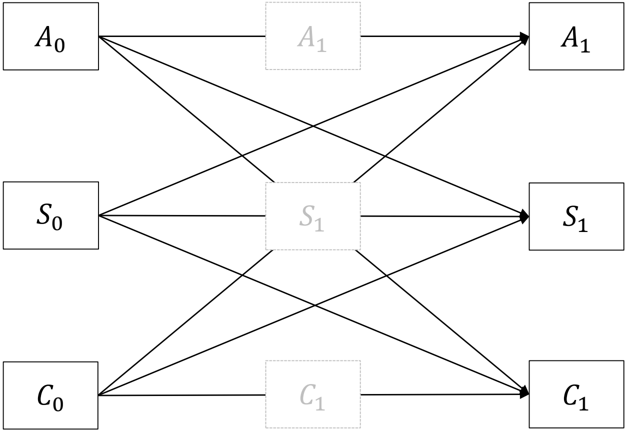
Time-interval dependency: Multivariate case Cole & Maxwell (2003); Reichardt (2011)

$$\boldsymbol{X}_\tau = \boldsymbol{\Phi}(\Delta t = 1)\boldsymbol{X}_{\tau-1} + \boldsymbol{\epsilon}_\tau$$



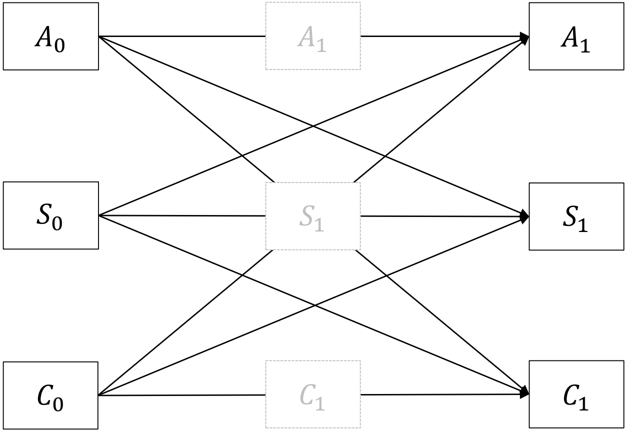
Time-interval dependency: Multivariate case Cole & Maxwell (2003); Reichardt (2011)

$$\boldsymbol{X}_\tau = \boldsymbol{\Phi}(\Delta t = 2)\boldsymbol{X}_{\tau-1} + \epsilon_\tau$$



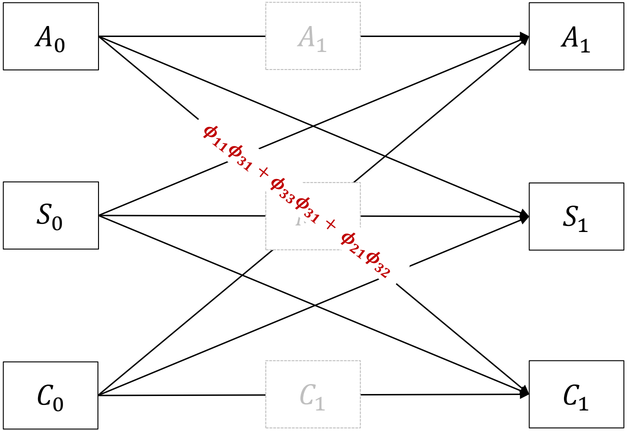
Time-interval dependency: Multivariate case Cole & Maxwell (2003); Reichardt (2011)

$$\mathbf{\Phi}(\Delta t = 2) = \mathbf{\Phi}(\Delta t = 1)^2$$



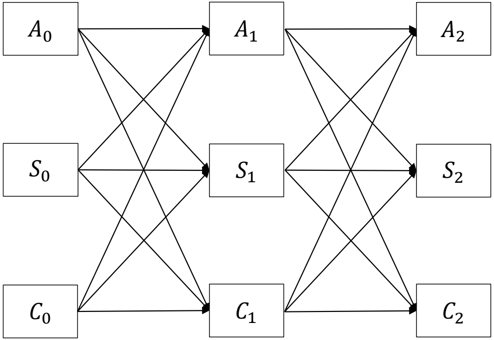
Time-interval dependency: Multivariate case Cole & Maxwell (2003); Reichardt (2011)

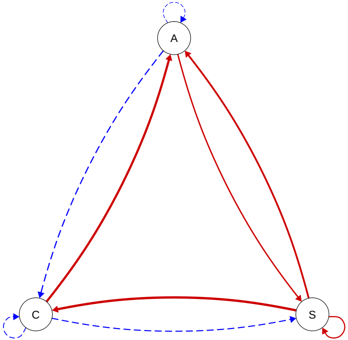
$$\mathbf{\Phi}(\Delta t = 2) = \mathbf{\Phi}(\Delta t = 1)^2$$



Implications of the Time-Interval Problem

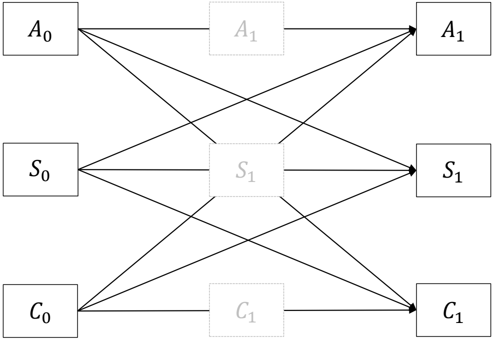
$$\mathbf{\Phi}(\Delta t = 1)$$

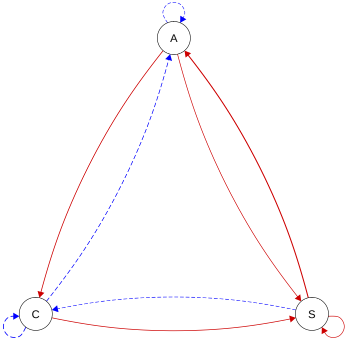




Implications of the Time-Interval Problem

$$\mathbf{\Phi}(\Delta t = 2)$$





Implications of the Time-Interval Problem

1. For a uniform time-interval Δt
   - $\hat{\mathbf{\Phi}} = \mathbf{\Phi}(\Delta t)$ *only for the given time-interval*
   - Depending on Δt, lagged effects may differ in terms of size, sign, and relative magnitude

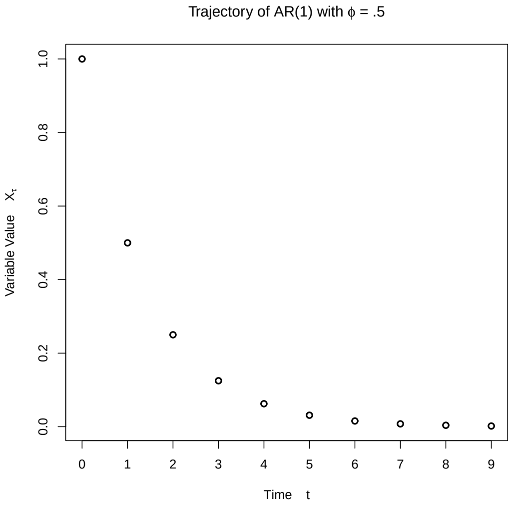
Implications of the Time-Interval Problem

1. For a uniform time-interval Δt
   - $\hat{\mathbf{\Phi}} = \mathbf{\Phi}(\Delta t)$ *only for the given time-interval*
   - Depending on Δt, lagged effects may differ in terms of size, sign, and relative magnitude
2. Unevenly spaced observations
   - $\hat{\mathbf{\Phi}} \neq \mathbf{\Phi}(\Delta t)$ *for any time-interval*

Implications of the Time-Interval Problem

1. For a uniform time-interval Δt
   - $\hat{\mathbf{\Phi}} = \mathbf{\Phi}(\Delta t)$ *only for the given time-interval*
   - Depending on Δt, lagged effects may differ in terms of size, sign, and relative magnitude
2. Unevenly spaced observations
   - $\hat{\mathbf{\Phi}} \neq \mathbf{\Phi}(\Delta t)$ *for any time-interval*
3. Interpretation of $\mathbf{\Phi}$ parameters as direct effects is questionable
   - Deboeck & Preacher, 2015; Aalen et al. 2012.
   - We will return to this in the end if we have time

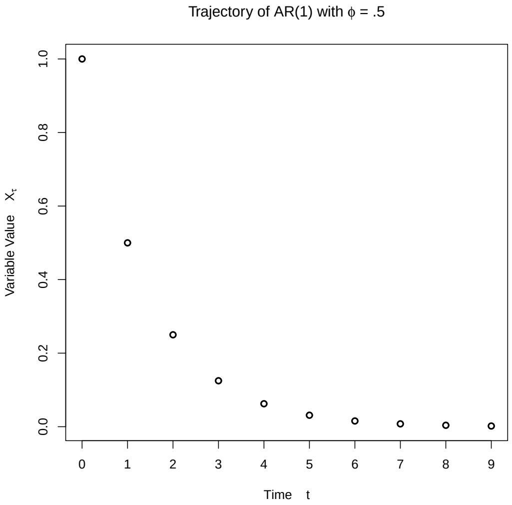
Continuous- and Discrete-Time Models



The traditional VAR(1) model is a **Discrete-Time** model

- Does not explicitly take into account time-interval between measurement
- As a data-generating model suggests that processes evolve in discrete "jumps"

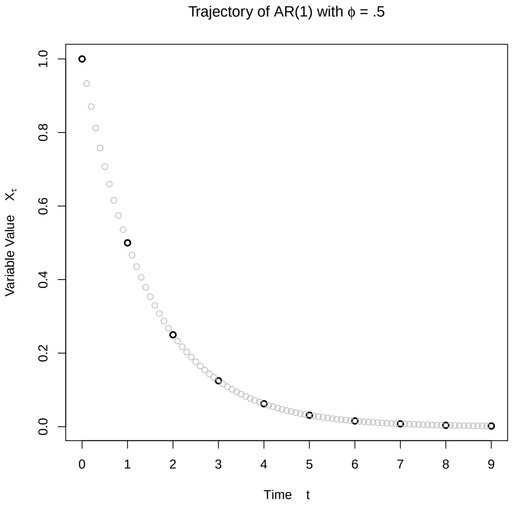
Continuous- and Discrete-Time Models



We can hypothesize that psychological processes

1. Take on some value at all points in time

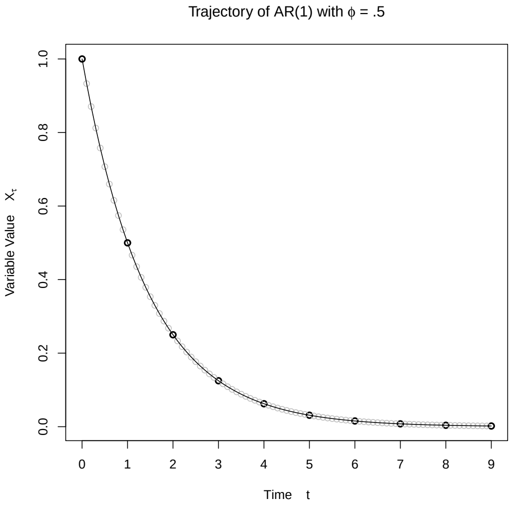
Continuous- and Discrete-Time Models



We can hypothesize that psychological processes

1. Take on some value at all points in time

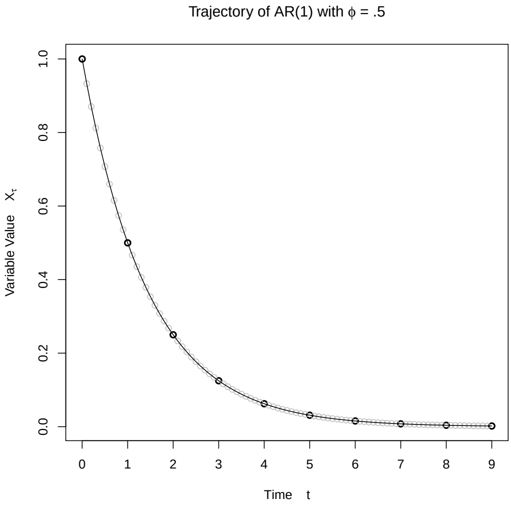
Continuous- and Discrete-Time Models



We can hypothesize that psychological processes

1. Take on some value at all points in time
2. Are smooth and differentiable

Continuous- and Discrete-Time Models



We can hypothesize that psychological processes

1. Take on some value at all points in time
2. Are smooth and differentiable
3. Exert influence on another at every moment in time

see Boker(2001) for a further discussion

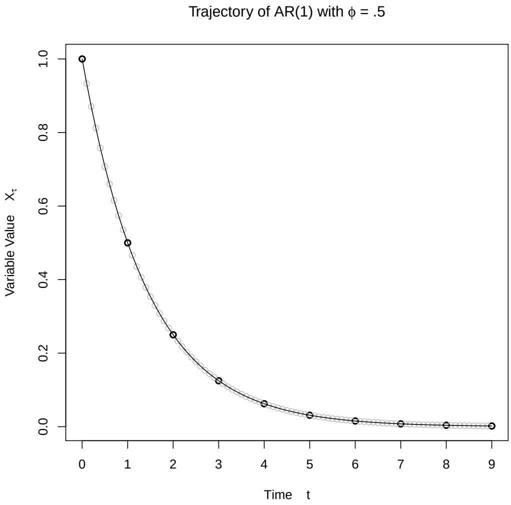
Continuous-Time VAR(1) Models

We can model such processes using **Continuous-Time Models**

$$\frac{d\boldsymbol{X}(t)}{dt} = \boldsymbol{A}\boldsymbol{X}(t) + \boldsymbol{\Gamma}(t)$$

- Univariate First-order Stochastic Differential Equation
- $\boldsymbol{Y}(t)$ is the position of the process at a point in time
- $\frac{d\boldsymbol{Y}(t)}{dt}$ is the rate of change of position (or velocity) at that point in time
- $\boldsymbol{A}$ is the drift matrix relating these two
- $\boldsymbol{\Gamma(t)}$ is the stochastic innovation part (see Voelkle et al. 2012)

Continuous-Time VAR(1) Models



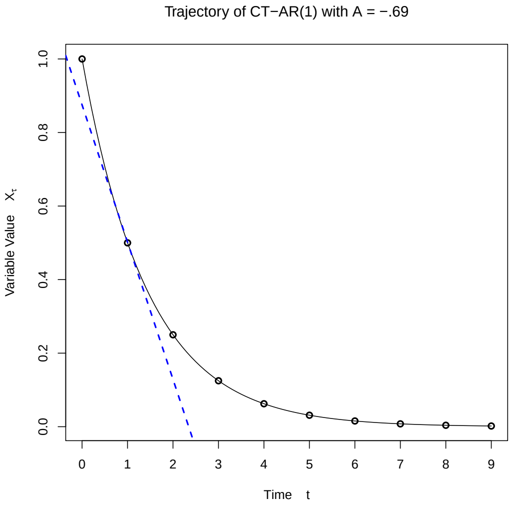
Take

$$E(\frac{dX(t)}{dt}) = AX(t)$$

with

- $A = -.69$
- initial value $X_0 = 1$

Continuous-Time VAR(1) Models



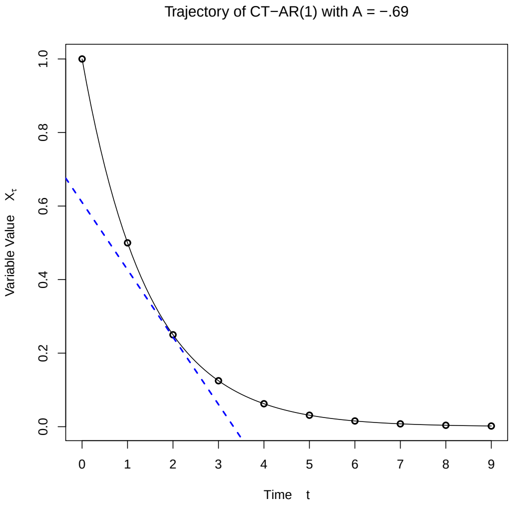
Take

$$E(\frac{dX(t)}{dt}) = AX(t)$$

with

- $A = -.69$
- initial value $X_0 = 1$

Continuous-Time VAR(1) Models



Take

$$E(\frac{dX(t)}{dt}) = AX(t)$$

with

- $A = -.69$
- initial value $X_0 = 1$

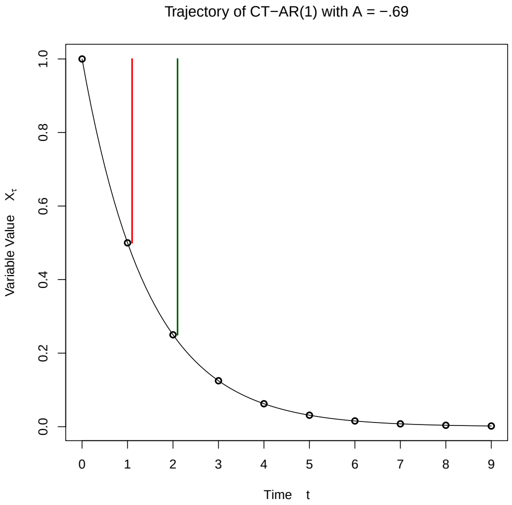
Continuous-Time VAR(1) Models

This differential equation can also be written as a CT-VAR(1) model

$$\boldsymbol{X}(t_\tau) = \boldsymbol{e}^{\boldsymbol{A}\Delta t}\boldsymbol{X}(t_{\tau-1}) + \epsilon(t_\tau)$$

- $\boldsymbol{X}(t_\tau)$ is the value of the process at a point in time corresponding to some measurement occassion
- $\boldsymbol{A}$ is the drift matrix from the differential equation
- Δt is the time interval between measurements τ and $\tau - 1$
- $\boldsymbol{e}$ is the matrix exponential
- $\epsilon(t_\tau)$ is the stochastic innovation - normally distributed with variance a function of Δt, $\boldsymbol{A}$ and γ (see Voelkle et al. 2012)

Continuous-Time VAR(1) Models



Take

$$E(X(t_\tau)) = e^{A\Delta t}X(t_{\tau-1})$$

with

- $A = -.69$
- initial value $X_0 = 1$

Continuous-Time VAR(1) Models

Comparing the CT-VAR(1) and the DT-VAR(1) models (under equally spaced observations)

$$\boldsymbol{\Phi}(\Delta t) = \boldsymbol{e}^{\boldsymbol{A}\Delta t}$$

Continuous-Time VAR(1) Models

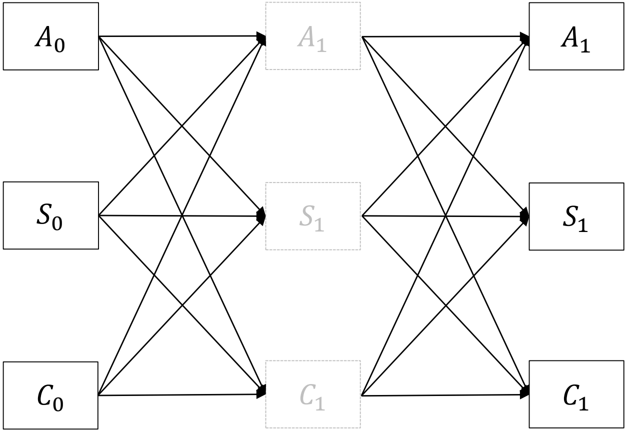
Comparing the CT-VAR(1) and the DT-VAR(1) models (under equally spaced observations)

$$\boldsymbol{\Phi}(\Delta t) = \boldsymbol{e}^{\boldsymbol{A}\Delta t}$$

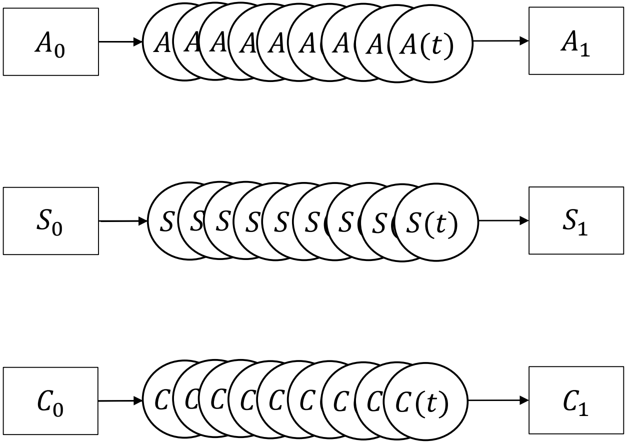
- This relationship is a good explanation for the often observed time-interval problem of DT-VAR(1) models
- CT models allow us to estimate a single effects matrix which is independent of Δt
- We can explore how lagged parameters potentially change as a function of Δt

Continuous-Time VAR(1) Models

$$\boldsymbol{X}_\tau = \boldsymbol{\Phi}(\Delta t = 1)\boldsymbol{X}_{\tau-1} + \boldsymbol{\epsilon}_\tau$$

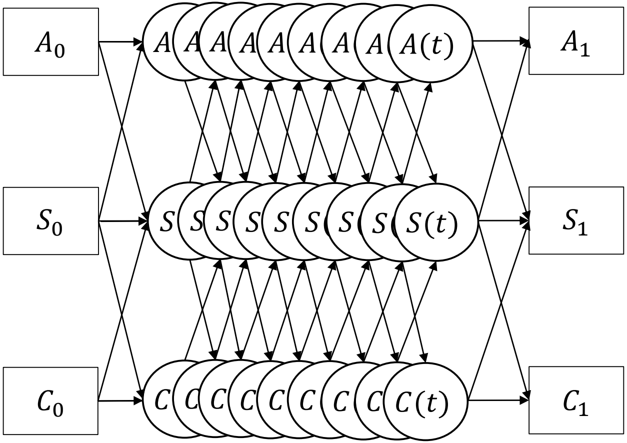


Continuous-Time VAR(1) Models



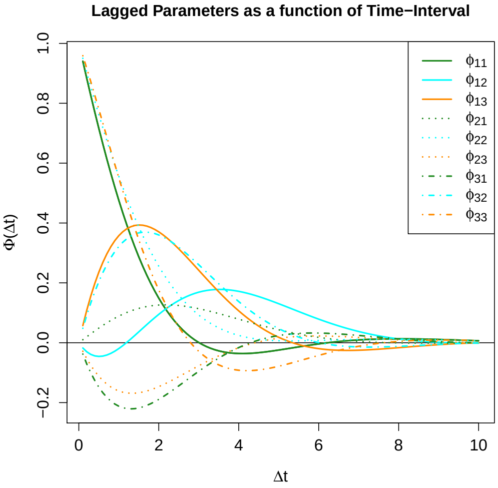
The Continuous-Time VAR(1) Model

$$\boldsymbol{X}(t_\tau) = \boldsymbol{e}^{\boldsymbol{A}\Delta t}\boldsymbol{X}(t_{\tau-1}) + \epsilon(t_\tau)$$



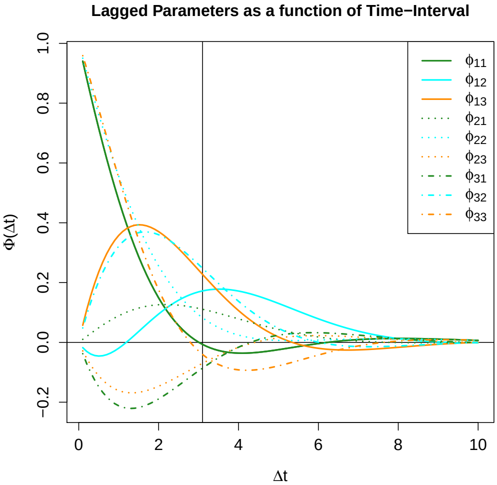
Time-interval dependency of VAR estimates

$$e^{\boldsymbol{A}\Delta t} = \boldsymbol{\Phi}(\Delta t)$$



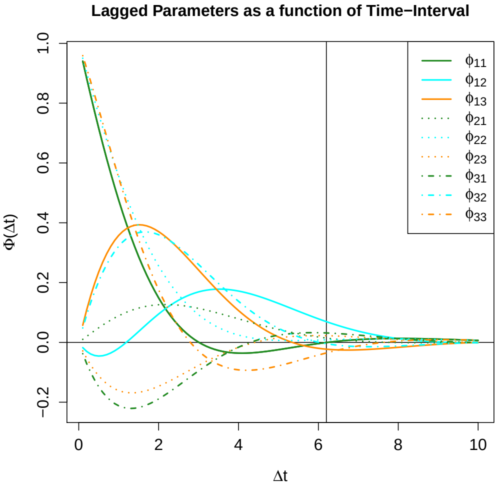
Time-interval dependency of VAR estimates

$$e^{\boldsymbol{A}\Delta t} = \boldsymbol{\Phi}(\Delta t)$$

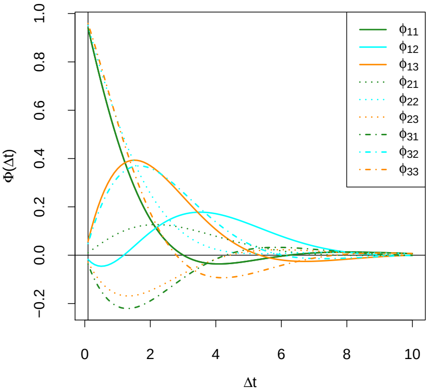


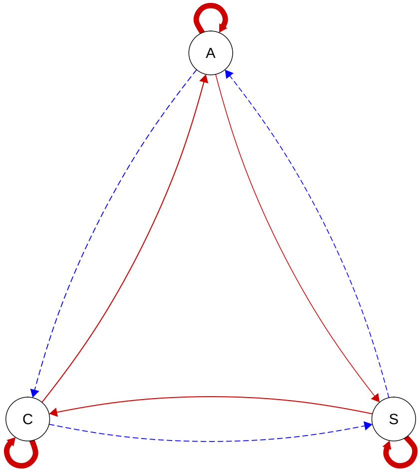
Time-interval dependency of VAR estimates

$$e^{\boldsymbol{A}\Delta t} = \boldsymbol{\Phi}(\Delta t)$$



Network structure as a function of lag





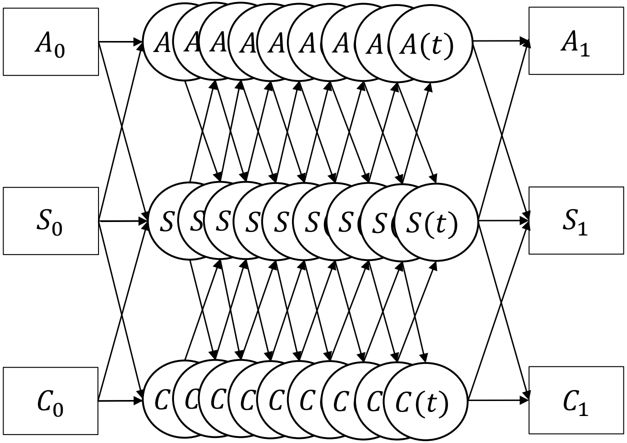
Conceptual Implications of CT models

DT model parameters $\mathbf{\Phi}$ generally interpreted as *direct effects*

- If the underlying process is a CT model these effects are not “direct” in an intuitive sense
- May be better conceptualised as “total” effects through unobserved pathways
- Cross-lagged parameters change sign due to competing negative and positive direct and indirect links in $\boldsymbol{A}$
- Implications for indirect effects centrality measures and network interpretation

The Continuous-Time VAR(1) Model

$$\boldsymbol{X}(t_\tau) = \boldsymbol{e}^{\boldsymbol{A}\Delta t}\boldsymbol{X}(t_{\tau-1}) + \boldsymbol{\epsilon}(t_\tau)$$

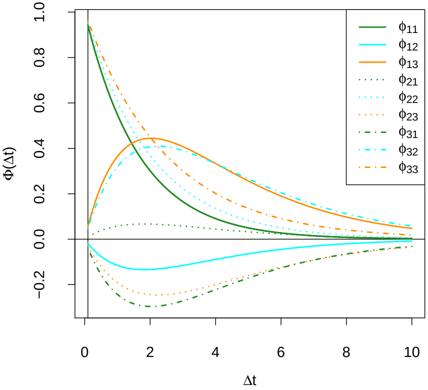


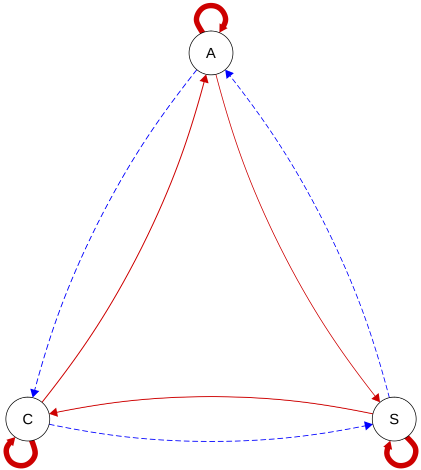
Conceptual Implications of CT models

Alternative calculations possible for CT path-specific effects

- CT direct lagged effects do not change sign across Δt
- Deboeck & Preacher 2016, Aalen et al. (2012, 2017)
- Ryan & Hamaker (in preperation 1) - further exploration and derivations in the context of interventionist causality
- Ryan & Hamaker (in preperation 2) - investigation of centrality measures for CT dynamical networks

CT Direct Effect Network structure as a function of lag





Estimation Possiblities

1. The Indirect Method
   - First estimate a DT-VAR(1) model with equally spaced observations
   - Alternatively correct for unequally spaced observations by inserting missing values
     - Pick the smallest time interval, and make sure the "rows" in your dataset are spaced with this time-interval
     - Only works with estimation methods which do not use listwise deletion (e.g. Bayesian approaches)
     - Implemented in DSEM in Mplus
   - Once estimated, "solve" for the CT matrix $\hat{\boldsymbol{A}}$ using $\hat{\boldsymbol{\Phi}}$
     - Use the logm() function in the R-package expm
     - Only works if the eigenvalues of $\boldsymbol{\Phi}$ are positive, real, and less than 1
     - Otherwise the matrix logarithm doesn't have a unique solution

Estimation Possiblities

2 Fit the Differential Equation Directly

- Generalised Linear Local Approximation (GLLA) and Latent Different Equations (LDE)
    - R-scripts created by Steve Boker and Colleagues
    - Use OpenMx, an R-based SEM package
    - First estimate the derivatives themselves, using kalman filter or latent variable loading constraints
    - Then fits the differential equation directly
    - Advantages: higher-order differential equations sometimes do not have easy-to-work with integral form
    - Disadvantages: multi-level extensions a little bit limited

Estimation Possiblities

3 Fit the CT-VAR(1) (Integral Form)
- ctsem by Driver, Voekle and Oud
  - R package
  - Uses OpenMx and STAN for bayesian estimation
  - Single subject and multi-level extensions
  - Can model higher-order systems (oscillation) through the VAR(1) model without constraints to real eigenvalues
  - Some of the newest features are still being tested - yet to see extensive published simulations

Estimation Possiblities

4. Packages just being released now
   - dynr by Ou, Hunter and Chow
     - R package
     - Does both DT and CT models, regime-switching models
     - Seemingly does everything
     - version 0.1.11-8 Released on CRAN 21st August 2017
5. BHOUM extensions by Kuiper and Oravecz
   - Alternative method of estimating multi-level integral form CT models
   - Development focusing on hypothesis testing and model comparisons

Summary

The CT-VAR(1) model is advantageous over the DT-VAR(1) model because

- Assumes a more realistic data-generating model for psychological processes
- May in turn lead to more realistic interpretation of causal structure
- Deals well with unequal spacing, gaps for night time etc.
- Gives us a single effects matrix which is independent of Δt
- Explains the time-interval dependency problem
- Allows us to explore how the lagged parameters potentially change as a function of Δt

Summary

The CT-VAR(1) model is disadvantageous over the DT-VAR(1) model because

- More difficult/involved to estimate
- Knowledge gap regarding differential equations and their interpretation
- Easier ways to deal with unequal spacing - e.g. inserting missing values
- Makes assumption of smoothness/differentiability

Upcoming work

- Ryan, Kuiper & Hamaker (under review) A continuous time approach to intensive longitudinal data: The what, why and how. In M.C. Voekle, K. von Montfort and J.H. Oud (Eds.) *Continuous time modeling in the behavioural and related sciences*
- Kuiper & Ryan (under review) Drawing conclusions from cross-lagged relationships: re-considering the role of the time-interval. *Structural Equation Modeling*
- Ryan & Hamaker (in preperation 1) Path-specific effects in CT modeling: re-visiting continuous-time mediation analysis from a causal perspective
- Ryan & Hamaker (in preperation 2) A continuous-time approach to dynamical network models

Get in Touch

- http://dml.sites.uu.nl/
- o.ryan@uu.nl

Key References

- Aalen, O. O., Rysland, K., Gran, J. M., & Ledergerber, B. (2012). Causality, mediation and time: a dynamic viewpoint. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 175*(4), 831-861.
- Aalen, O. O., Gran, J. M., Rysland, K., Stensrud, M. J., & Strohmaier, S. (2017). Feedback and Mediation in Causal Inference Illustrated by Stochastic Process Models. *Scandinavian Journal of Statistics.*
- Boker, S.M. (2001) Consequences of continuity: The hunt for intrinsic properties within parameters of dynamics in psychological processes *Multivariate Behavioral Research, 37*(3), 405-422
- Boker, S. M., Montpetit, M. A., Hunter, M. D., & Bergeman, C. S. (2010). Modeling resilience with differential equations. In *Learning and Development: Individual Pathways of Change. Washington, DC: American Psychological Association*, 183-206.
- Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling. Journal of abnormal psychology, 112(4), 558-577.

Key References

- Deboeck, P. R., & Preacher, K. J. (2016). No need to be discrete: A method for continuous time mediation analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 61-75.
- Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child development*, 80-92.
- Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research, 40*(2), 207-233.
- Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton: Princeton university press.
- Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2009). A hierarchical OrnsteinUhlenbeck model for continuous repeated measurement data. *Psychometrika, 74*(3), 395-418.
- Voelkle, M. C., Oud, J. H., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: relating authoritarianism and anomia. *Psychological methods, 17*(2), 176.

Continuous Time Model

First-Order Stochastic Differential Equation

$$\frac{d\boldsymbol{Y}(t)}{dt} = \boldsymbol{A}(\boldsymbol{Y}(t) - \boldsymbol{\mu}) + \gamma \frac{d\boldsymbol{W}(t)}{dt}$$

CT VAR(1) Model

$$\boldsymbol{Y}(t) = \boldsymbol{e}^{\boldsymbol{A}\Delta t}\boldsymbol{Y}(t - \Delta t) + \boldsymbol{w}(\Delta t)$$

Numerical Example Network

$$\boldsymbol{A} = \begin{bmatrix} -6 & -.2 & .6 \\ .1 & -.5 & -.3 \\ -.4 & .5 & -.4 \end{bmatrix}$$