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

Typically used:

- Many repeated measurements of the same construct(s)

The VAR(1) model

Typically used:

- Many repeated measurements of the same construct(s)
- Interested in how the value of A now predicts A in the future
  - **Stability** or **inertia** (autoregressive effects)

The VAR(1) model

Typically used:

- Many repeated measurements of the same construct(s)
- Interested in how the value of A now predicts A in the future
  - **Stability** or **inertia** (autoregressive effects)
- Interested in how the value of A now predicts B in the future
  - Direct effects, influence, granger-causality (**cross-lagged** effects)

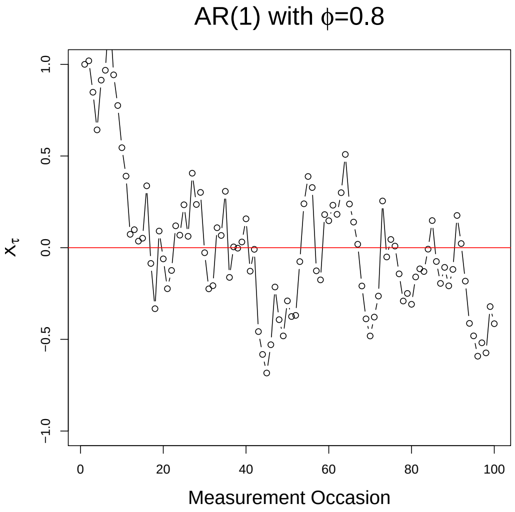
The VAR(1) model

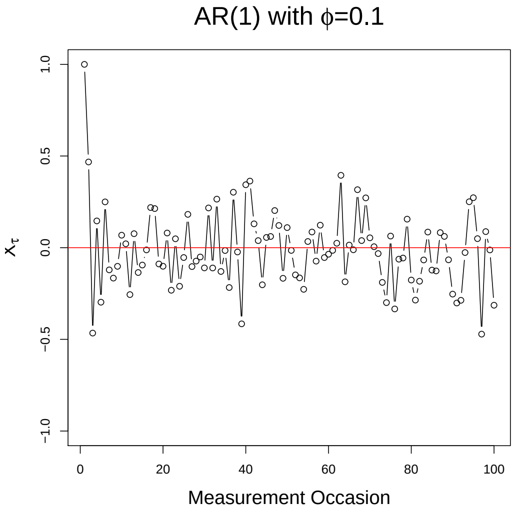
Typically used:

- Many repeated measurements of the same construct(s)
- Interested in how the value of A now predicts A in the future
  - **Stability** or **inertia** (autoregressive effects)
- Interested in how the value of A now predicts B in the future
  - Direct effects, influence, granger-causality (**cross-lagged** effects)
- Fluctuations around an equilibrium which doesn't change
  - Actually much more flexible - growth curves, etc.

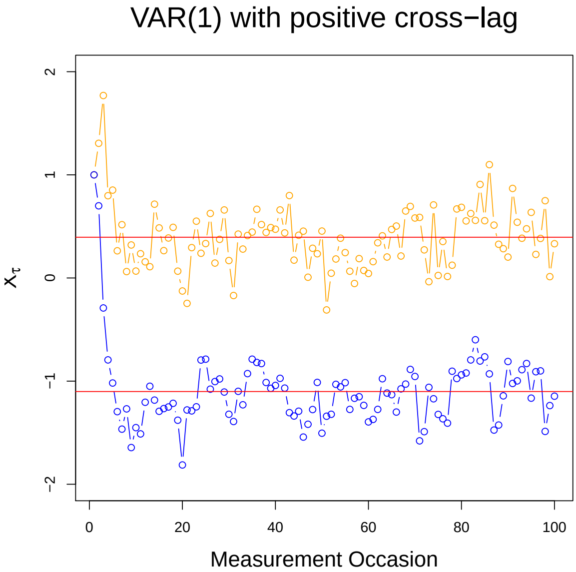
The VAR(1) Model

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



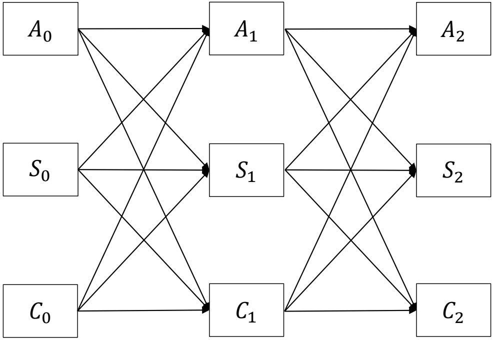


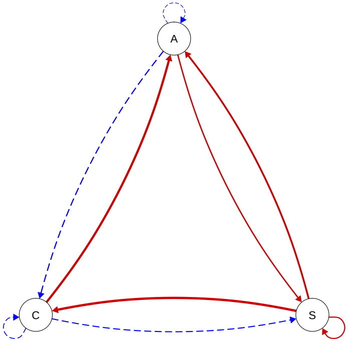
The VAR(1) Model



The VAR(1) Model

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





The VAR(1) Model

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

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**

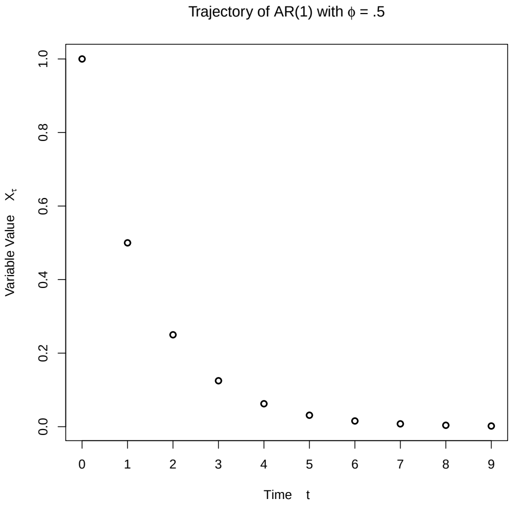
The time-interval dependency problem

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



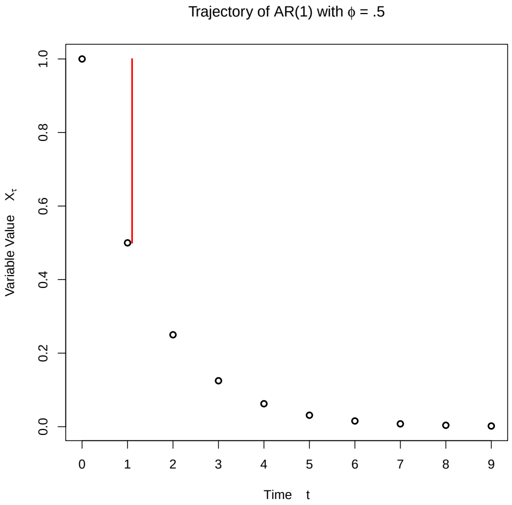
Take

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

with

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

The time-interval dependency problem



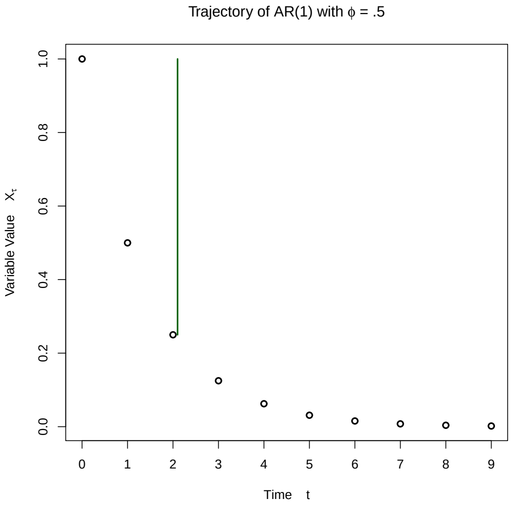
Take

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

with

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

The time-interval dependency problem



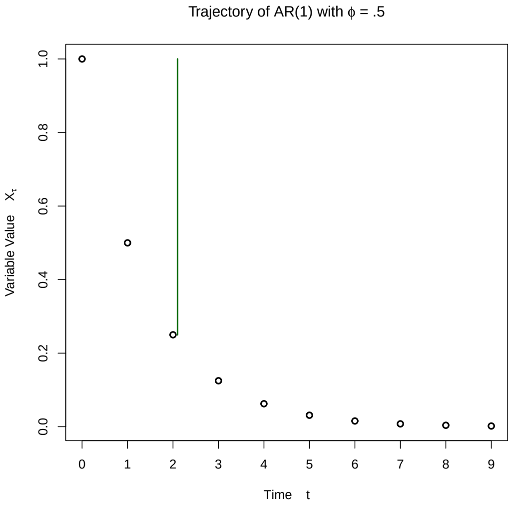
Take

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

with

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

The time-interval dependency problem



Take

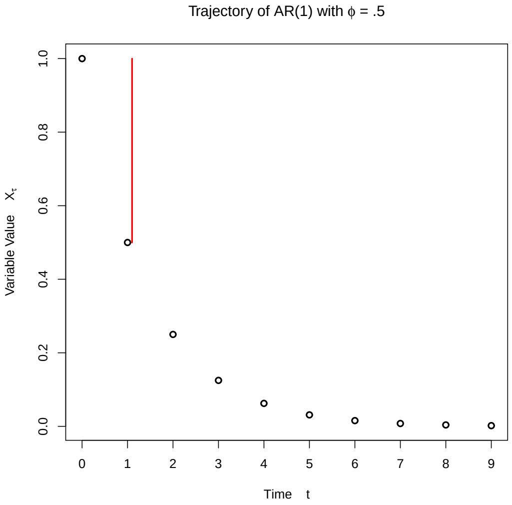
$$E(X_\tau) = \phi^2 X_{\tau-2}$$

with

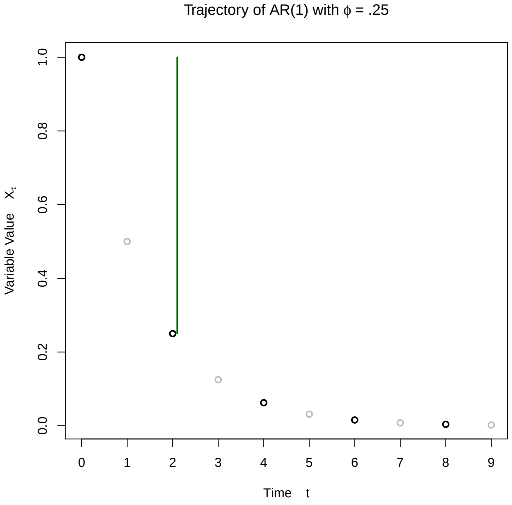
- $\phi = .5$
- initial value $X_0 = 1$
- $\Delta t = 1$

The time-interval dependency problem

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

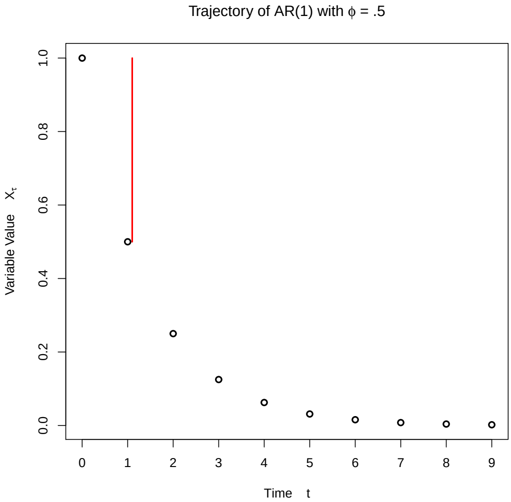


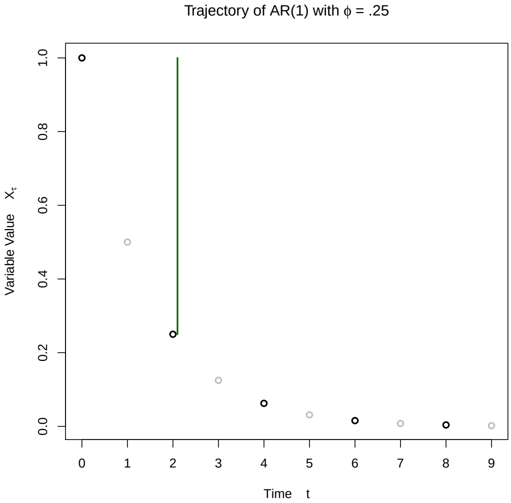
$$E(X_\tau) = \phi(\Delta t = 2)X_{\tau-1}$$



The time-interval dependency problem

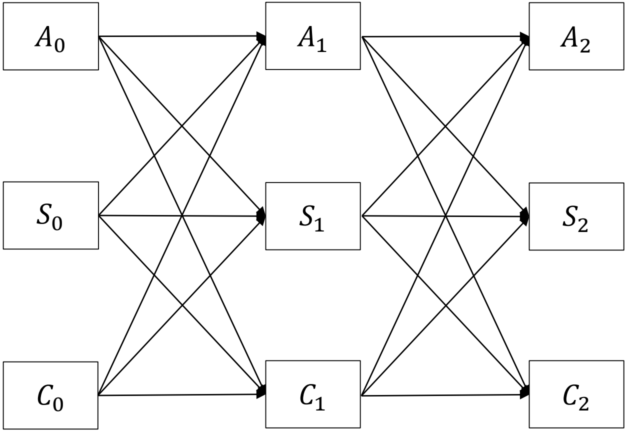
$$\phi(\Delta t = 1)^2 = \phi(\Delta t = 2)$$





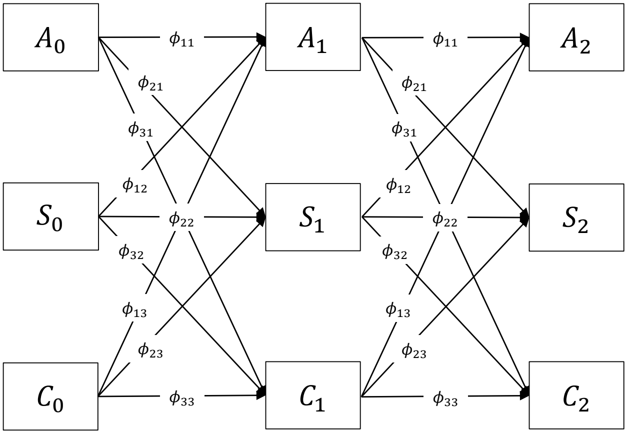
Time-interval dependency: Multivariate case Cole & Maxwell (2003); Reichardt (2011)

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



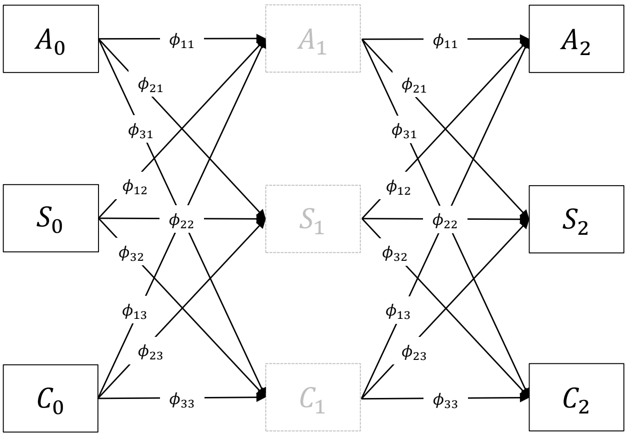
Time-interval dependency: Multivariate case Cole & Maxwell (2003); Reichardt (2011)

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



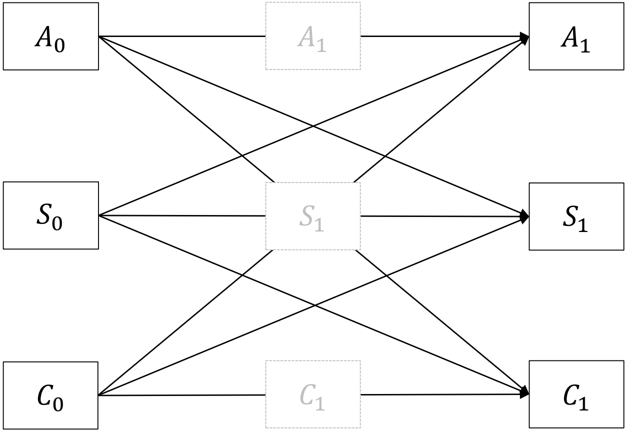
Time-interval dependency: Multivariate case Cole & Maxwell (2003); Reichardt (2011)

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



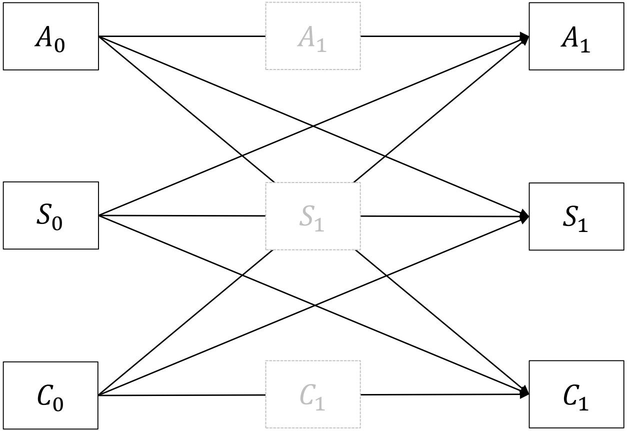
Time-interval dependency: Multivariate case Cole & Maxwell (2003); Reichardt (2011)

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



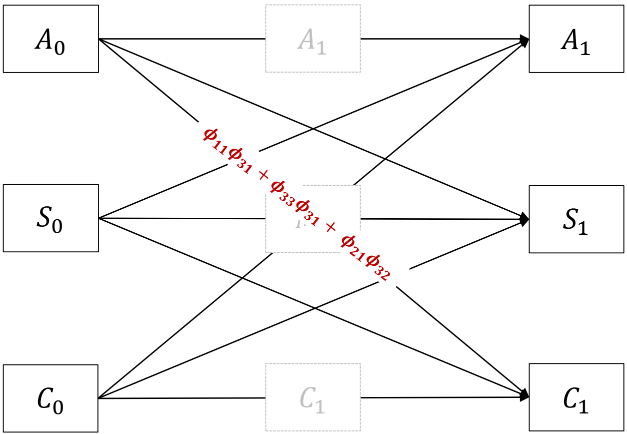
Time-interval dependency: Multivariate case Cole & Maxwell (2003); Reichardt (2011)

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



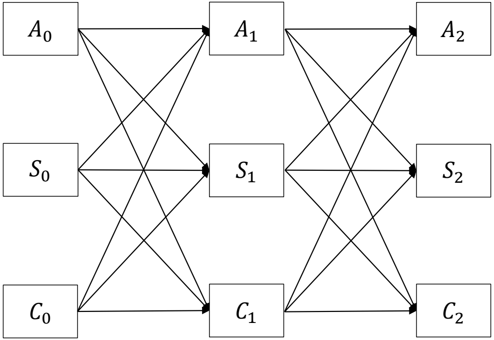
Time-interval dependency: Multivariate case Cole & Maxwell (2003); Reichardt (2011)

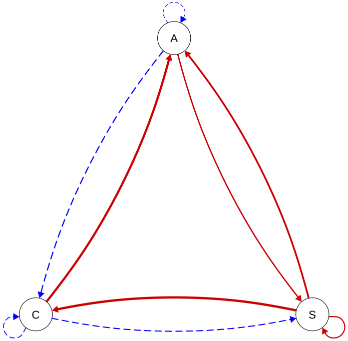
$$\mathbf{\Phi}(\Delta t = 2) = \mathbf{\Phi}(\Delta t = 1)^2$$



Implications of the Time-Interval Problem

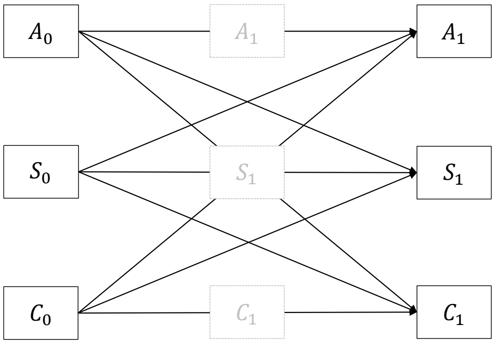
$$\mathbf{\Phi}(\Delta t = 1)$$

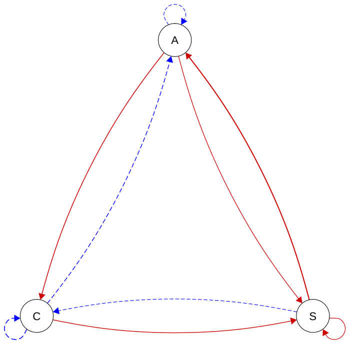




Implications of the Time-Interval Problem

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





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

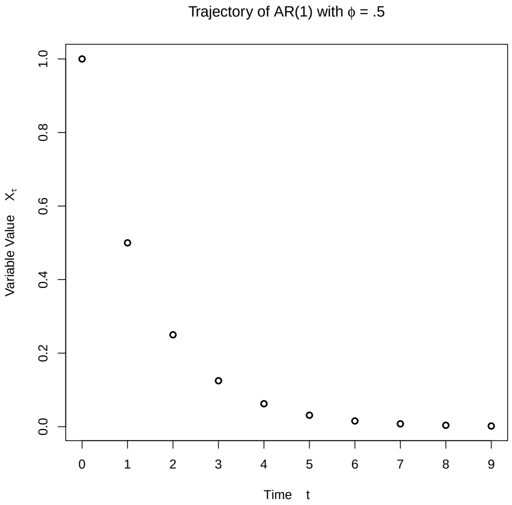
Implications of the Time-Interval Problem

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

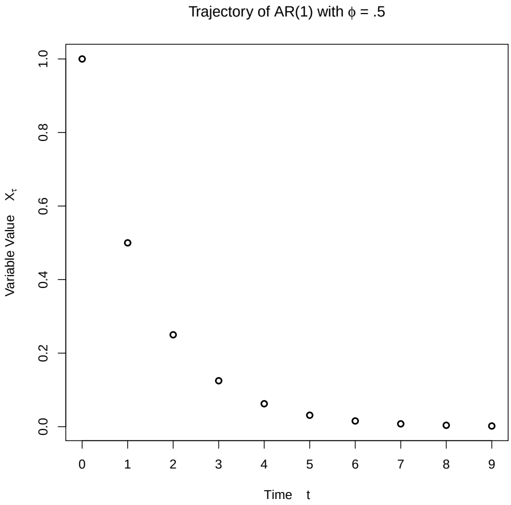
Continuous- and Discrete-Time Models



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"

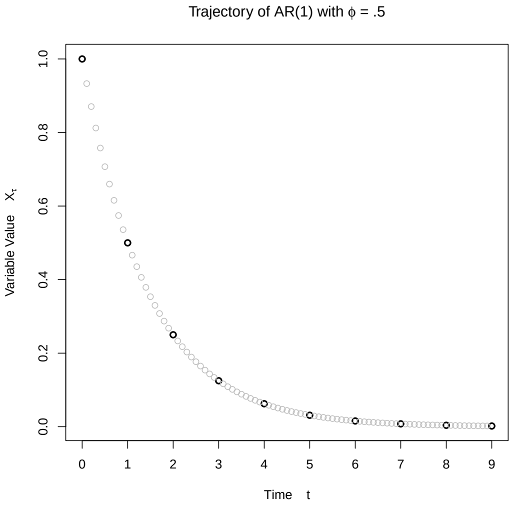
Continuous- and Discrete-Time Models



We can hypothesize that psychological processes

1. Take on some value at all points in time

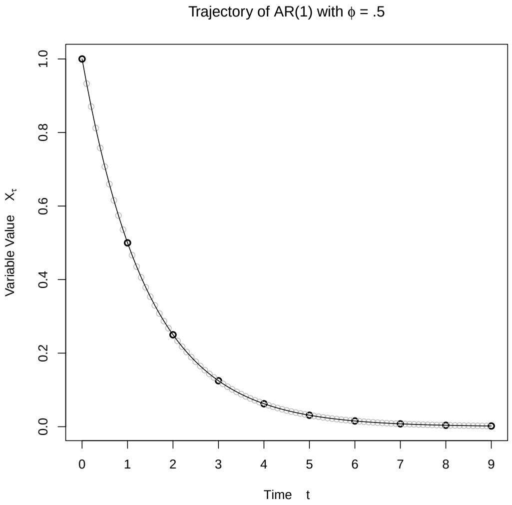
Continuous- and Discrete-Time Models



We can hypothesize that psychological processes

1. Take on some value at all points in time

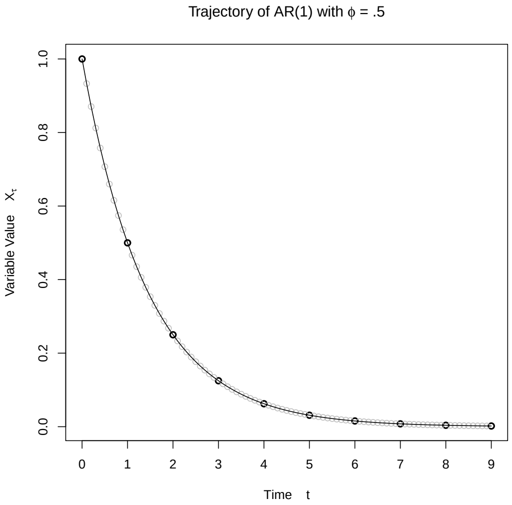
Continuous- and Discrete-Time Models



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

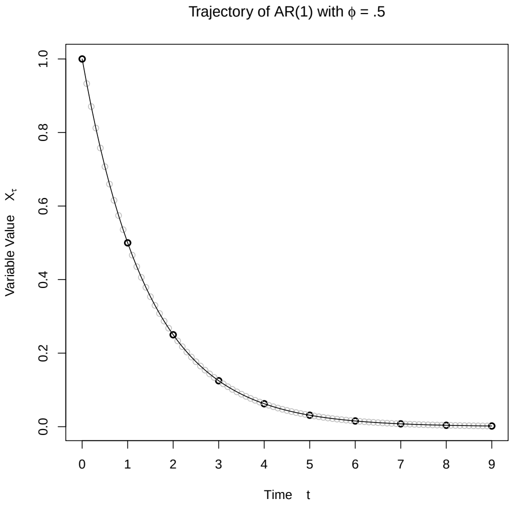
Continuous-Time VAR(1) Models

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
- $\boldsymbol{X}(t)$ is the position of the process at a point in time
- $\frac{d\boldsymbol{X}(t)}{dt}$ 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



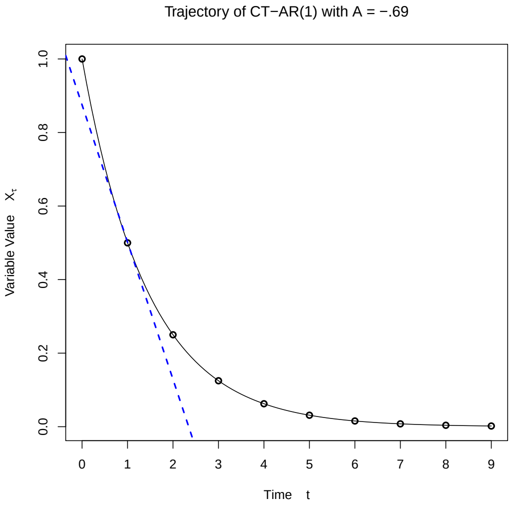
Take

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

with

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

Continuous-Time VAR(1) Models



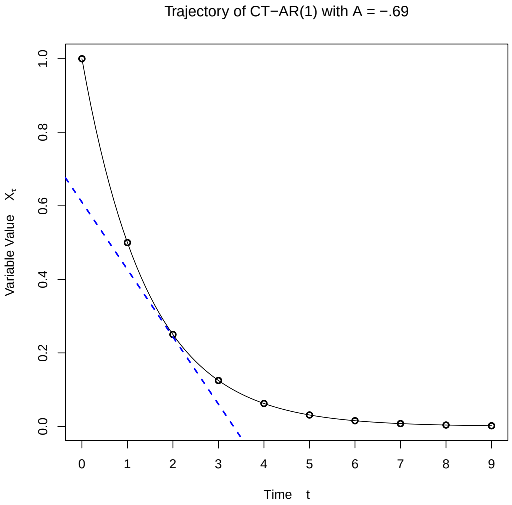
Take

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

with

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

Continuous-Time VAR(1) Models



Take

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

with

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

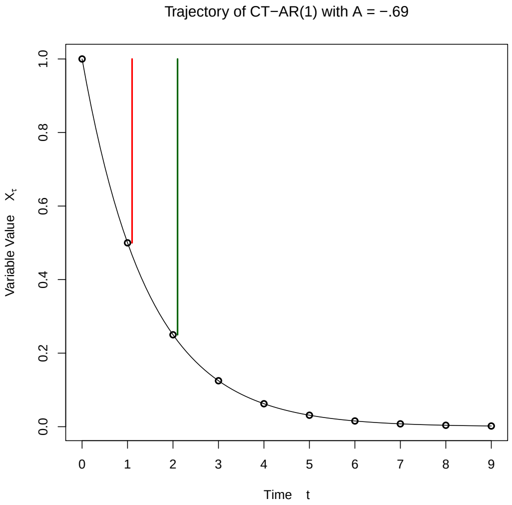
Continuous-Time VAR(1) Models

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

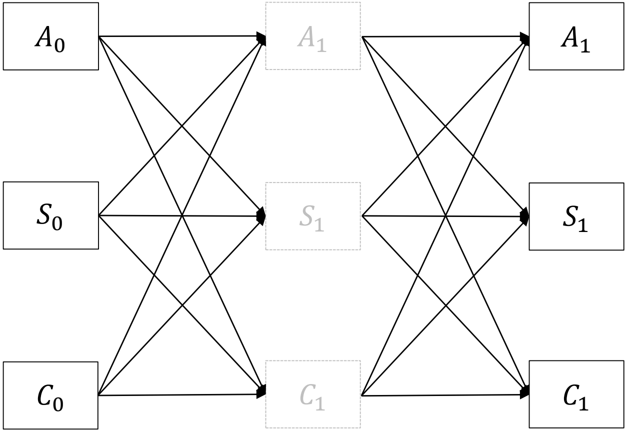
Comparing the CT-VAR(1) and the DT-VAR(1) models (under equally spaced observations)

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

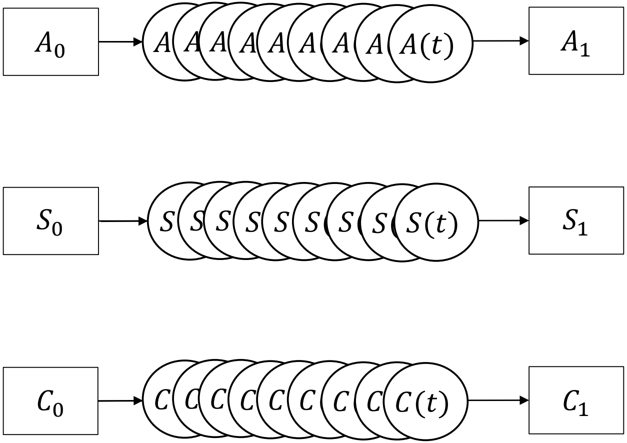
- This relationship is a good explanation for the often observed time-interval problem of DT-VAR(1) models
- CT models allow us to estimate a single effects matrix which is independent of Δt
- We can explore how lagged parameters potentially change as a function of Δt

Continuous-Time VAR(1) Models

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

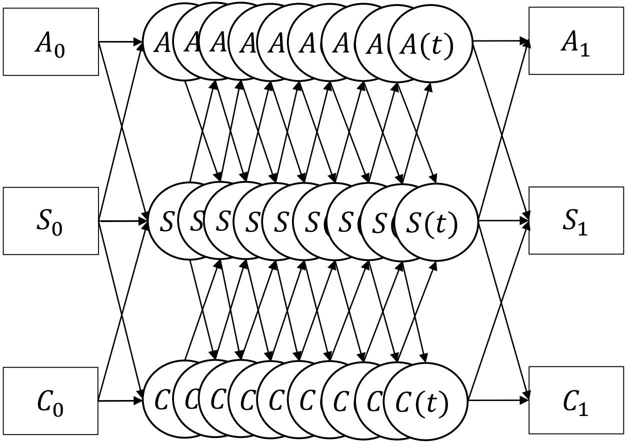


Continuous-Time VAR(1) Models



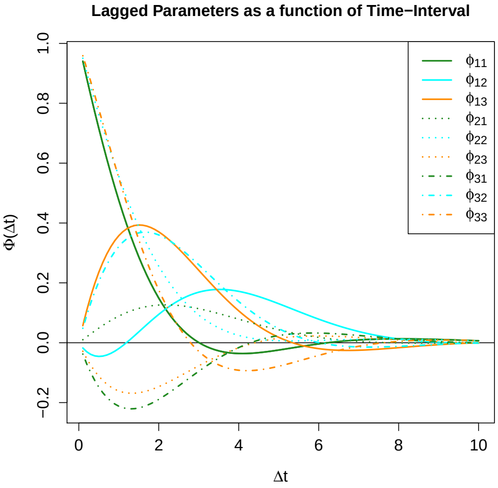
The Continuous-Time VAR(1) Model

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



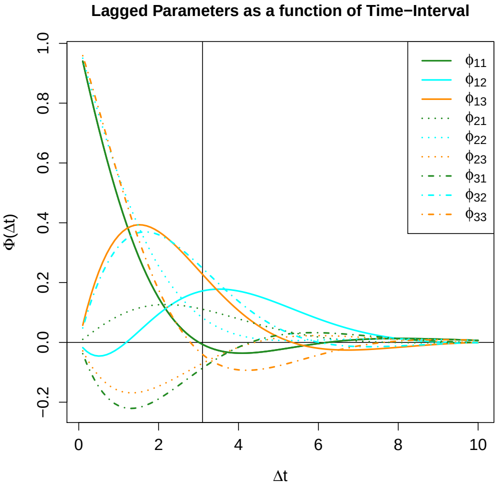
Time-interval dependency of VAR estimates

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



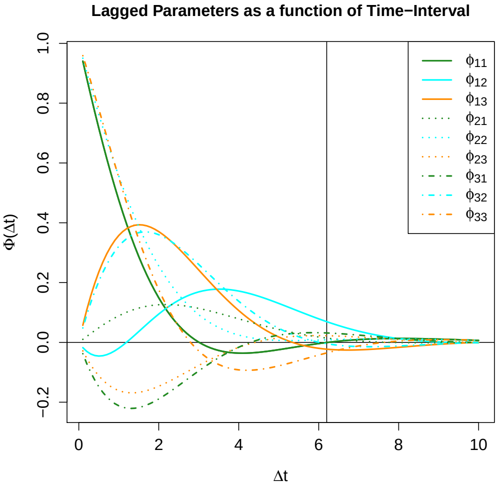
Time-interval dependency of VAR estimates

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

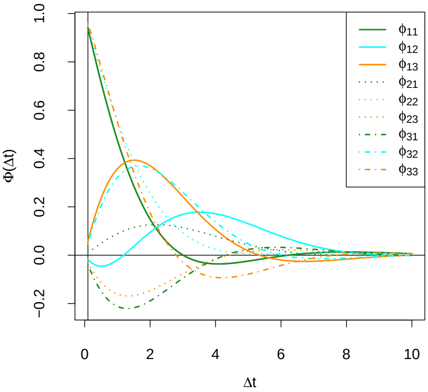


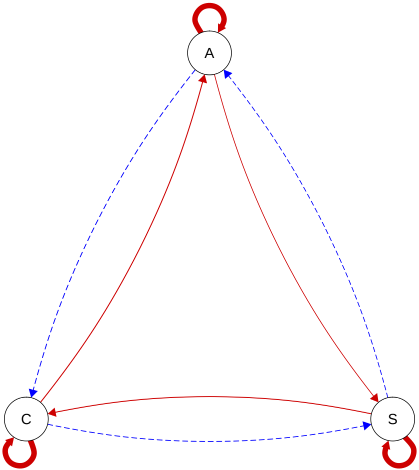
Time-interval dependency of VAR estimates

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



Network structure as a function of lag





Practical Benefits of the CT-VAR(1) model

- Deal with the "problem" of unequally spaced measurements
  - Information about the spacing is used, not ignored!
  - Still estimate only a single effects matrix - no extra parameters!
- Explore how lagged parameters **potentially** change as a function of the time-interval
- Make different predictions about interventions
- Different perspective on direct, indirect and total effects
- Match our substantive ideas about psychological processes!

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
  - Kalman filter for single-subject
  - 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

Analysis Example

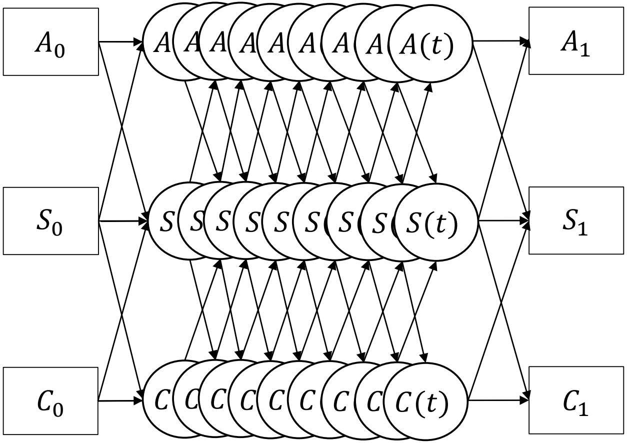
Conceptual Implications of CT models

DT model parameters $\boldsymbol{\Phi}$ 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

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



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