New Approaches To Analysing Psychological Time Series

SAA 2023 Symposium

Oisín Ryan, Jonas Haslbeck, Nick Jacobson, Marieke Helmich, Merlijn Olthof

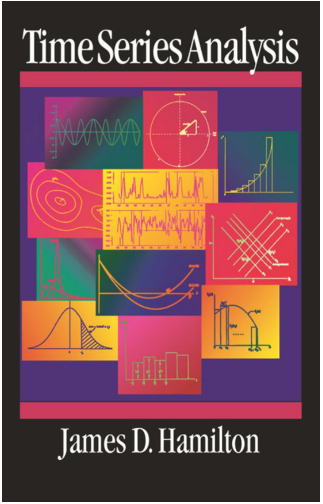
Extracting Dynamic Features from Irregularly Spaced Time Series

R package *expct*

Oisín Ryan[1,2], Kejin Wu[3], Nicholas C. Jacobson[4]

[1]*Department of Methodology and Statistics, Utrecht University*
[2]*Department of Data Science and Biostatistics, University Medical Centre Utrecht*
[3]*Department of Mathematics, University of California, San Diego*
[4]*Departments of Biomedical Data Science, Psychiatry, and Computer Science, Dartmouth College*

SAA 2023 Amsterdam



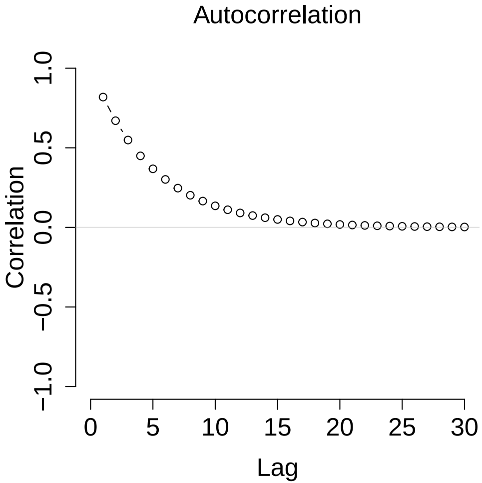
Descriptive / Exploratory Tools

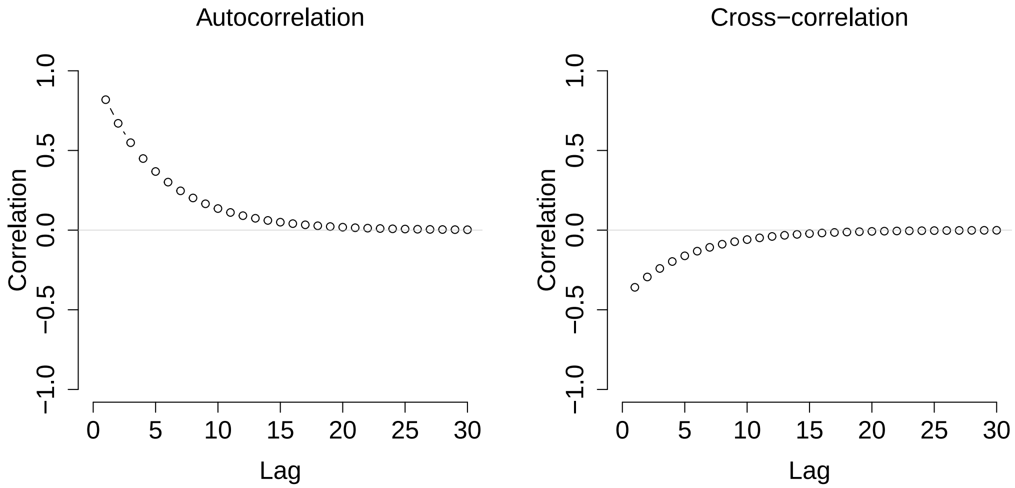
- Autocorrelation function (ACF)
- Cross-Correlation function (CCF)

Y

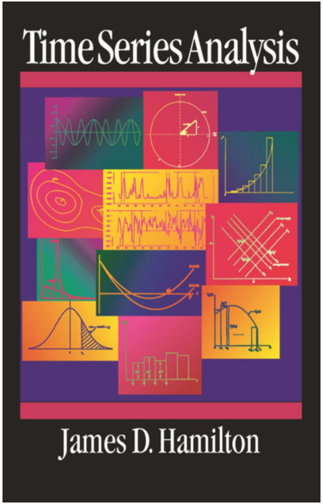
y_1
y_2
y_3
y_4
y_5
y_6
y_7
y_8
$\ldots$
y_T

Y	Y at lag 1	Y at lag 2
y_1		
y_2	y_1	
y_3	y_2	y_1
y_4	y_3	y_2
y_5	y_4	y_3
y_6	y_5	y_4
y_7	y_6	y_5
y_8	y_7	y_6
$\dots$	$\dots$	$\dots$
y_T	y_{T-1}	y_{T-2}
	y_T	y_{T-1}
		y_T


Autocorrelation
Correlation
1.0
0.5
0.0
−0.5
−1.0
0
5
10
15
20
25
30
Lag



Autocorrelation
Cross−correlation
Correlation
Lag
1.0
0.5
0.0
−0.5
−1.0
0
5
10
15
20
25
30

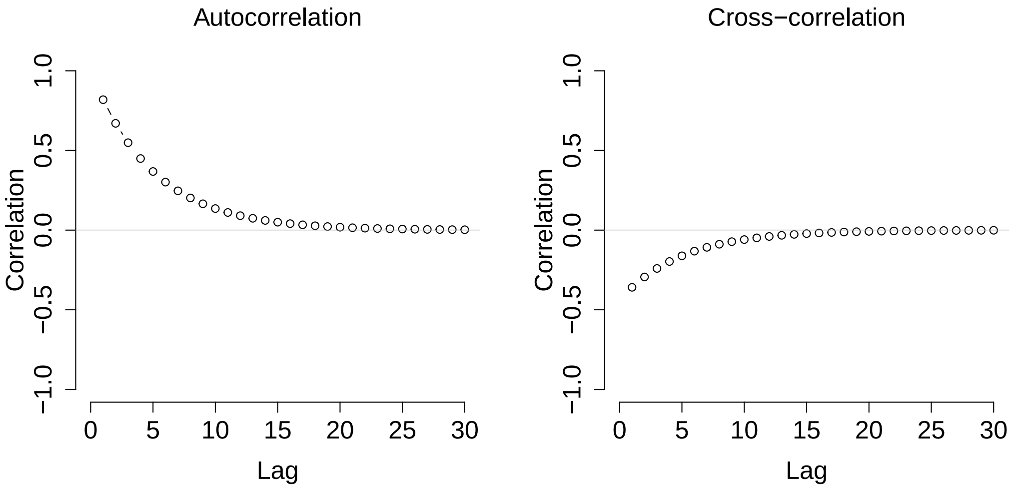


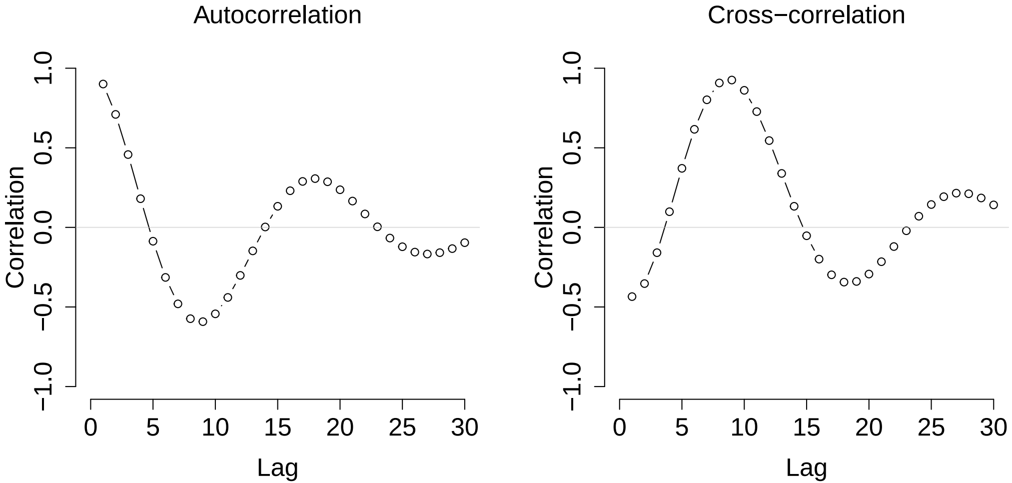
Descriptive / Exploratory Tools

- Autocorrelation function (ACF)
- Cross-Correlation function (CCF)

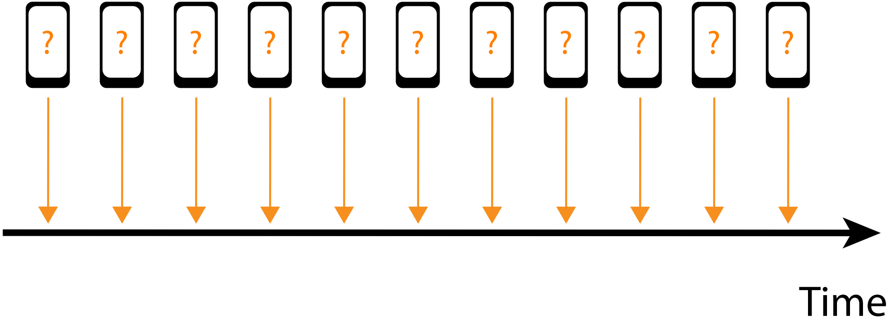
Models such as VAR and ARIMA

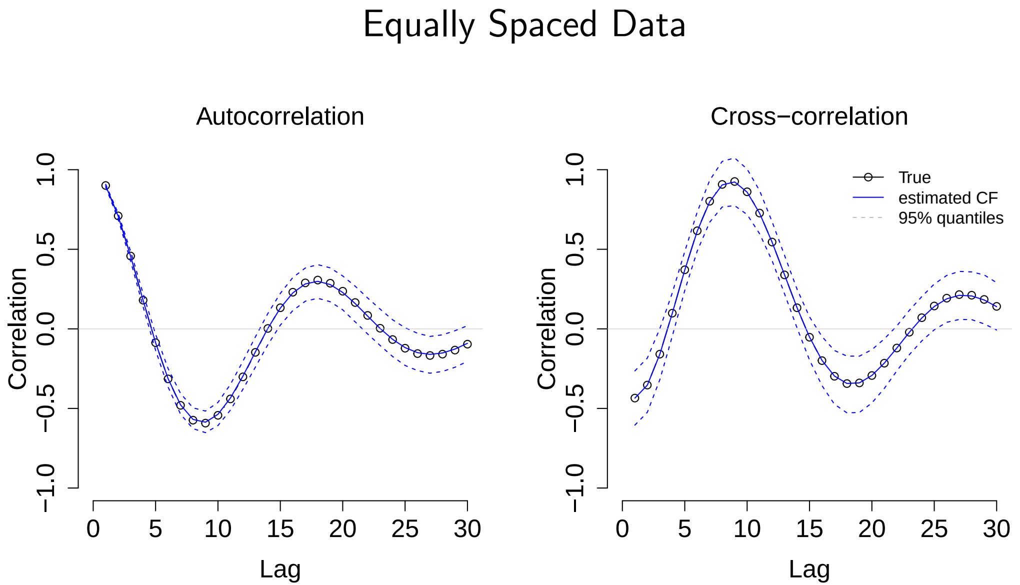
- AR(1), AR(2), VAR(p)


Autocorrelation
Cross−correlation
Correlation
1.0
0.5
0.0
−0.5
−1.0
0
5
10
15
20
25
30
Lag

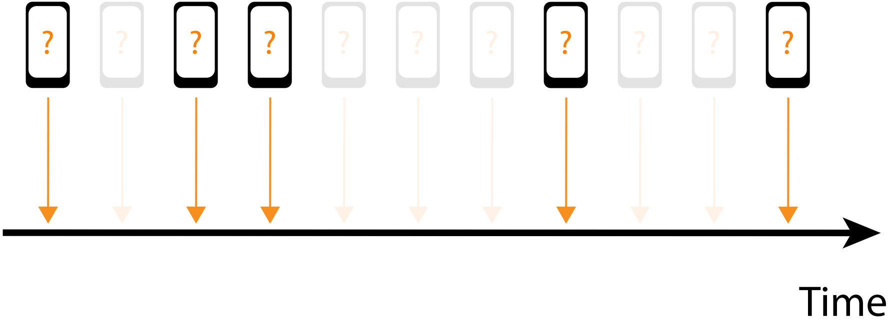

Autocorrelation
Cross−correlation
Correlation
1.0
0.5
0.0
−0.5
−1.0
0
5
10
15
20
25
30
Lag

Assumption: Equally Spaced Measurements

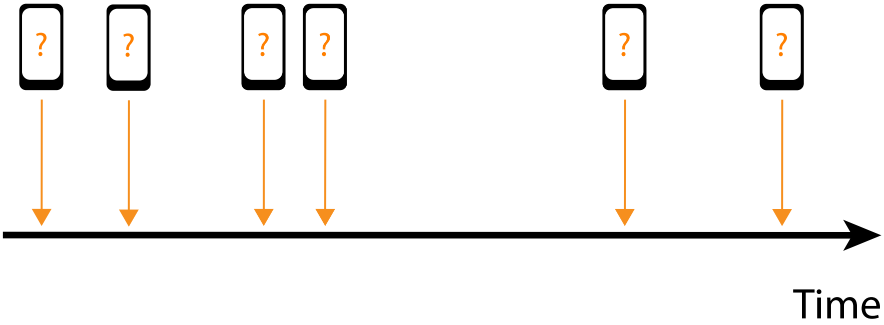



Equally Spaced Data
Autocorrelation
Cross−correlation
Correlation
Lag
−1.0
−0.5
0.0
0.5
1.0
0
5
10
15
20
25
30
True
estimated CF
95% quantiles

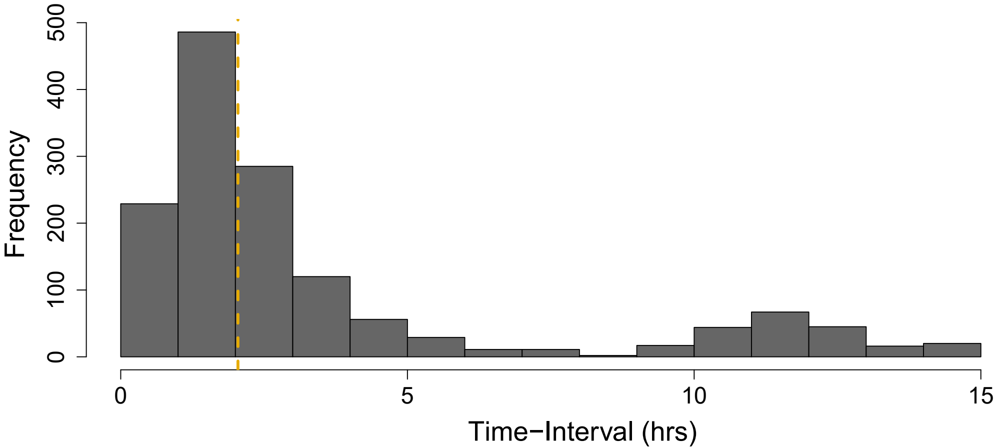
Reality: Irregularly Spaced Measurements



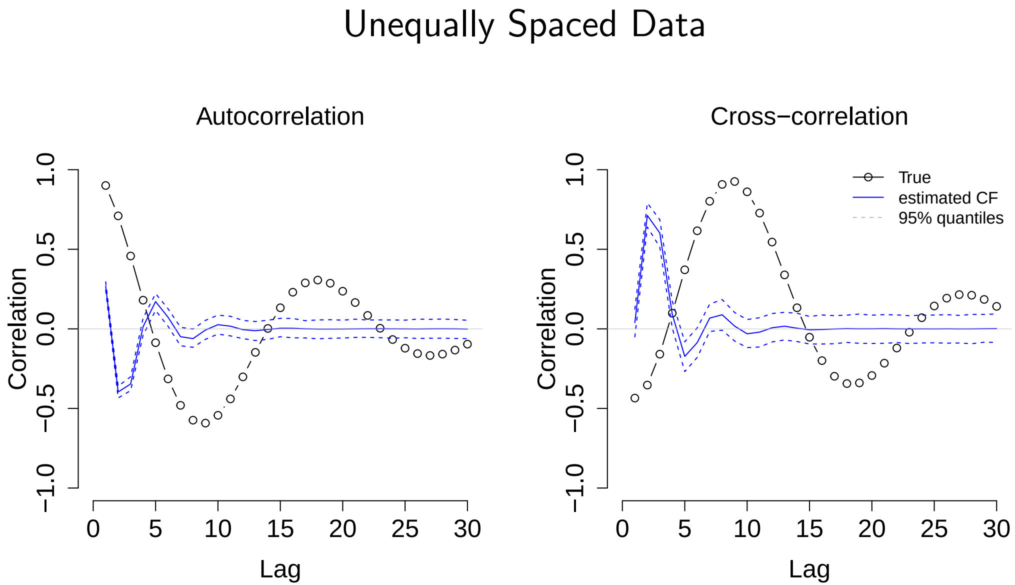
Reality: Irregularly Spaced Measurements



Measurement Spacing in Empirical Data



Wichers & Groot (2016)


Unequally Spaced Data
Autocorrelation
Cross−correlation
Correlation
Lag
1.0
0.5
0.0
−0.5
−1.0
0
5
10
15
20
25
30
True
estimated CF
95% quantiles

Continuous-Time Modeling

Avoids time-interval problems by modelling moment-to-moment dynamics directly:

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

Re-written allows us to model variable relations at different time-intervals (Δt)

$$\boldsymbol{Y}(t+\Delta t) = \mathbf{e}^{\boldsymbol{A}\Delta t}\boldsymbol{Y}(t) + \epsilon(\Delta t)$$

Can be estimated, e.g., using the *ctsem* package (Driver et al. 2017)

- Use $\mathbf{e}^{\hat{\boldsymbol{A}}\Delta t}$ to inspect auto-and cross regression effects at different time-intervals
- Or transform to find model-implied auto and cross **correlations**

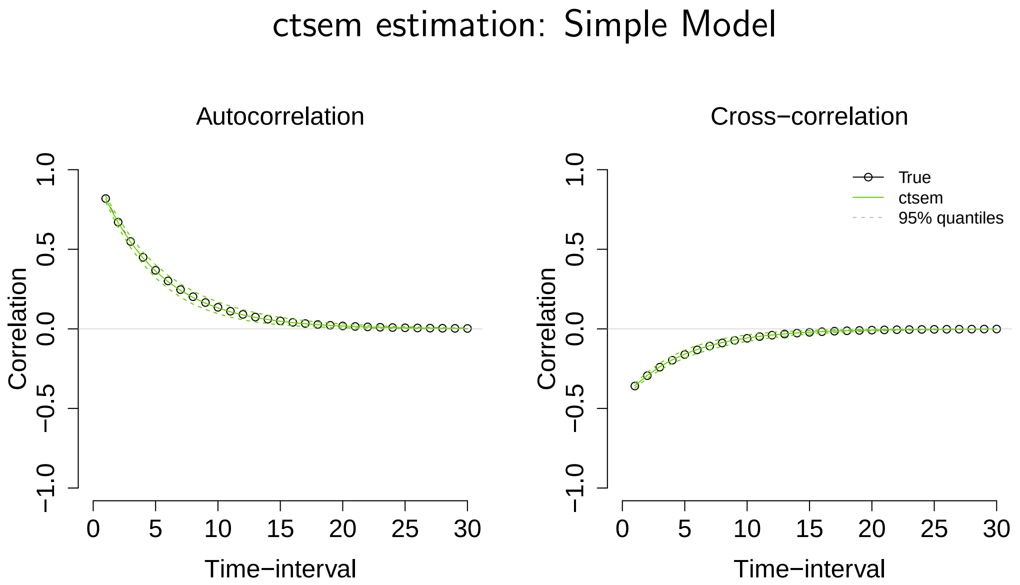
Boker (2002); Oud & Delsing (2010); Voelkle et al (2012); Ryan & Hamaker (2022)

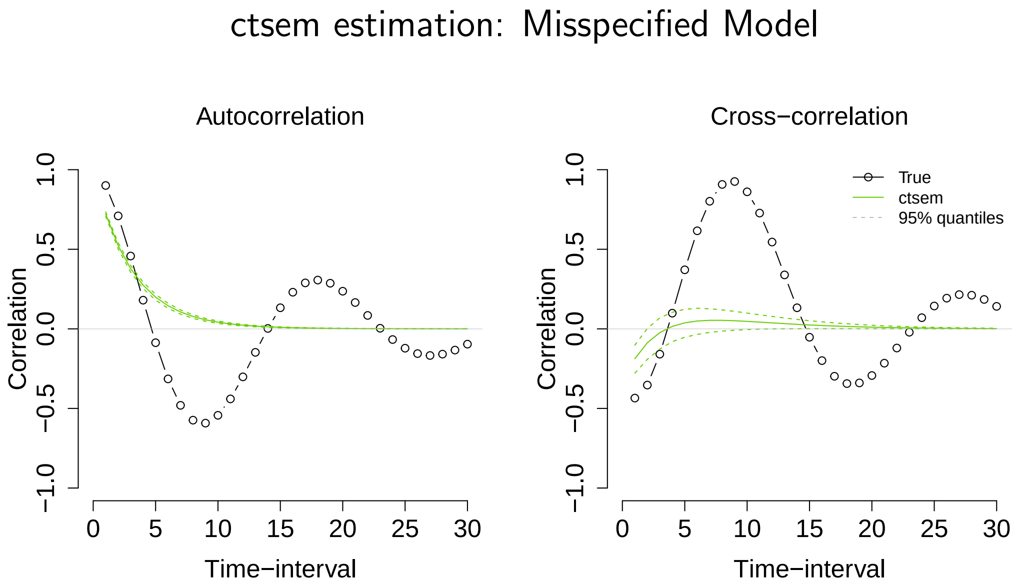
Problem: Model Misspecification

Model-based correlations accurate only if the model is **correctly specified**

In **reality** the model is **never correctly specified**.

- We may have the **order** of the model (first vs second) wrong
- Functional form of the relationships may not be **linear** at the DE level
- Even if we have a linear first order model, if we have unobserved **confounders** or even **mediators**, we may run into problems


ctsem estimation: Simple Model
Autocorrelation
Cross−correlation
Correlation
Time−interval
1.0
0.5
0.0
−0.5
−1.0
0
5
10
15
20
25
30
True
ctsem
95% quantiles



ctsem estimation: Misspecified Model
Autocorrelation
Cross−correlation
Correlation
Time−interval
−1.0
−0.5
0.0
0.5
1.0
0
5
10
15
20
25
30
True
ctsem
95% quantiles

Traditional ACF estimation:

- **Data-driven and exploratory** method for exploring dynamic features
- Does not perform well with unequally spaced time series

Continuous-Time (CT) model estimation:

- Can be estimated from **unequally spaced time series**
- Inferences rely on correct model specification
- Without a data-driven way of computing auto- and cross- correlations: **no easy way to check model misspecification**

expct: Exploratory Continuous Time Modeling

Method to estimate ACF and CCFs from unequally spaced data

R package: github: ryanoisin/expct

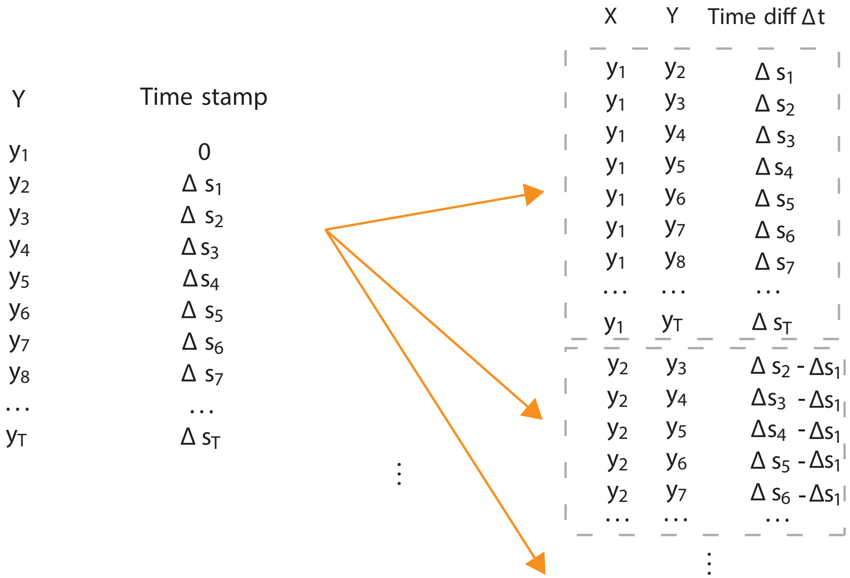
expct: Exploratory Continuous Time Modeling

Method to estimate ACF and CCFs from unequally spaced data
R package: github: ryanoisin/expct

Two-step procedure

1. Create a “stacked” data frame

Y	Time stamp
y_1	0
y_2	Δs_1
y_3	Δs_2
y_4	Δs_3
y_5	Δs_4
y_6	Δs_5
y_7	Δs_6
y_8	Δs_7
$\dots$	$\dots$
y_T	Δs_T


Y
Time stamp
y_1 0
y_2 Δs_1
y_3 Δs_2
y_4 Δs_3
y_5 Δs_4
y_6 Δs_5
y_7 Δs_6
y_8 Δs_7
$\dots$ $\dots$
y_T Δs_T
X
Y
Time diff Δt
y_1 y_2 Δs_1
y_1 y_3 Δs_2
y_1 y_4 Δs_3
y_1 y_5 Δs_4
y_1 y_6 Δs_5
y_1 y_7 Δs_6
y_1 y_8 Δs_7
$\dots$ $\dots$ $\dots$
y_1 y_T Δs_T
y_2 y_3 $\Delta s_2 - \Delta s_1$
y_2 y_4 $\Delta s_3 - \Delta s_1$
y_2 y_5 $\Delta s_4 - \Delta s_1$
y_2 y_6 $\Delta s_5 - \Delta s_1$
y_2 y_7 $\Delta s_6 - \Delta s_1$
$\dots$ $\dots$ $\dots$

expct: Exploratory Continuous Time Modeling

Method to estimate ACF and CCFs from unequally spaced data
R package: github: ryanoisin/expct

Two-step procedure

1. Create a "stacked" data frame
2. Use Generalized Additive Model (GAM) to estimate how lagged correlations depend on the time-interval

$$Y_{t+\Delta t} = f(\Delta t)Y_t + \epsilon$$

expct: Exploratory Continuous Time Modeling

Method to estimate ACF and CCFs from unequally spaced data
R package: github: ryanoisin/expct

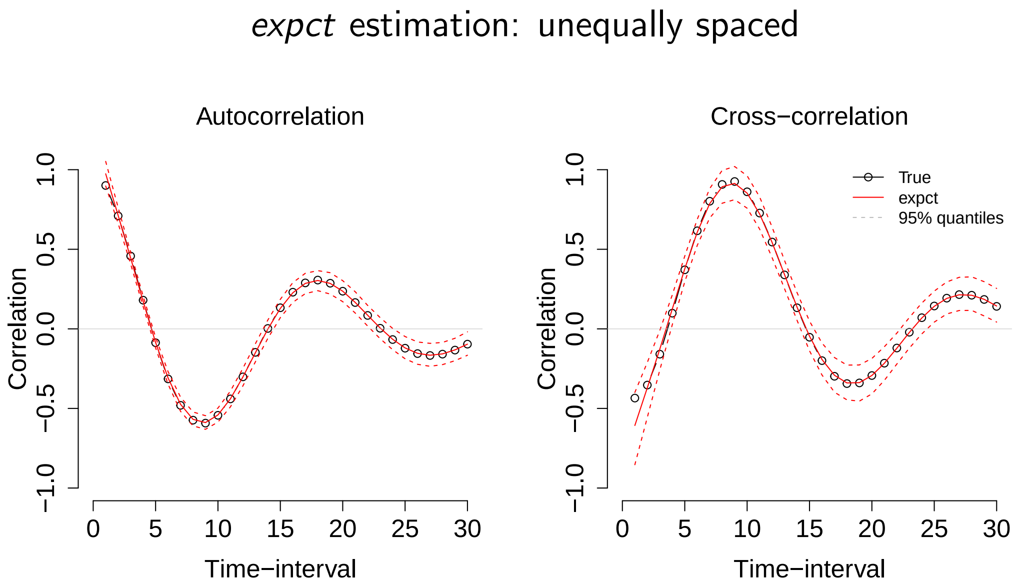
Two-step procedure

1. Create a "stacked" data frame
2. Use Generalized Additive Model (GAM) to estimate how lagged correlations depend on the time-interval

$$Y_{t+\Delta t} = f(\Delta t) Y_t + \epsilon$$

By rescaling $f(\Delta t)$ we estimate

$$cor(Y_t, Y_{t+\Delta t}) \propto f(\Delta t)$$



expct estimation: unequally spaced
Autocorrelation
Cross−correlation
Correlation
−1.0
−0.5
0.0
0.5
1.0
0
5
10
15
20
25
30
Time−interval
True
expct
95% quantiles

Simulation Study

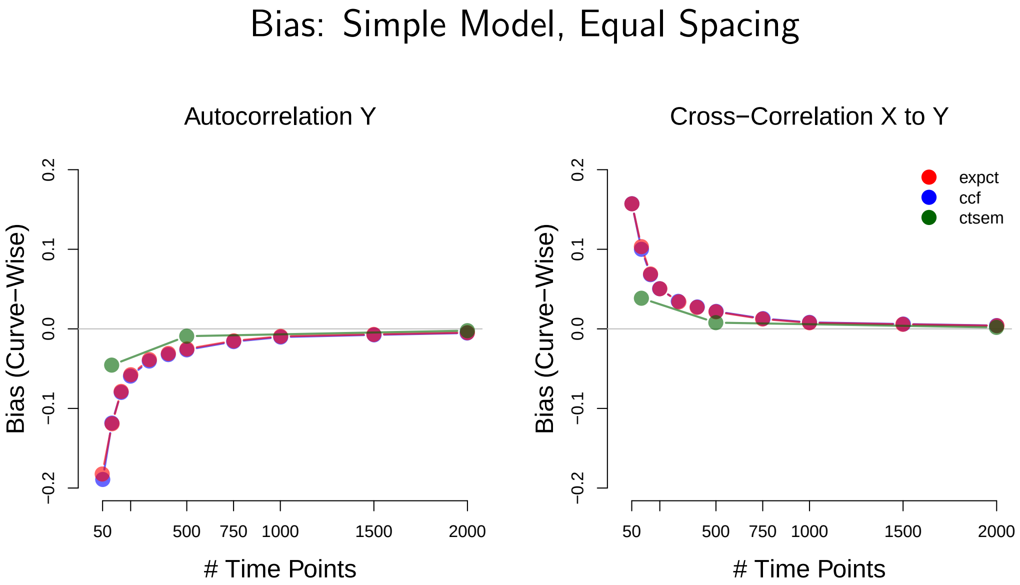
Time-series length: [50 - 2000]
Sampling Scheme: { Equal, Unequal Bimodal, Unequal Uniform }
Data-generating models: { Simple, Oscillating, Complex (missing variables) }

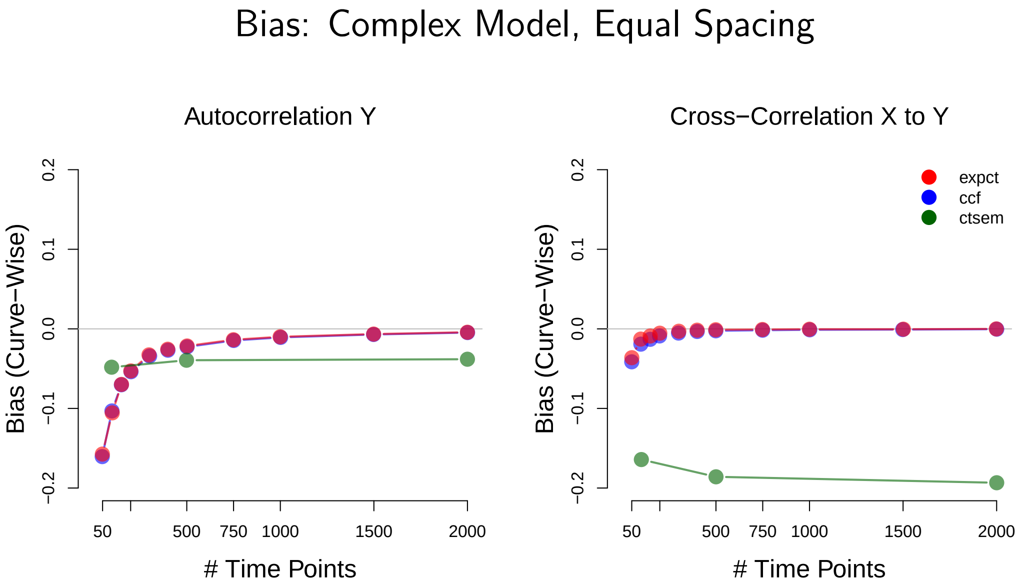
Methods for computing CIs:

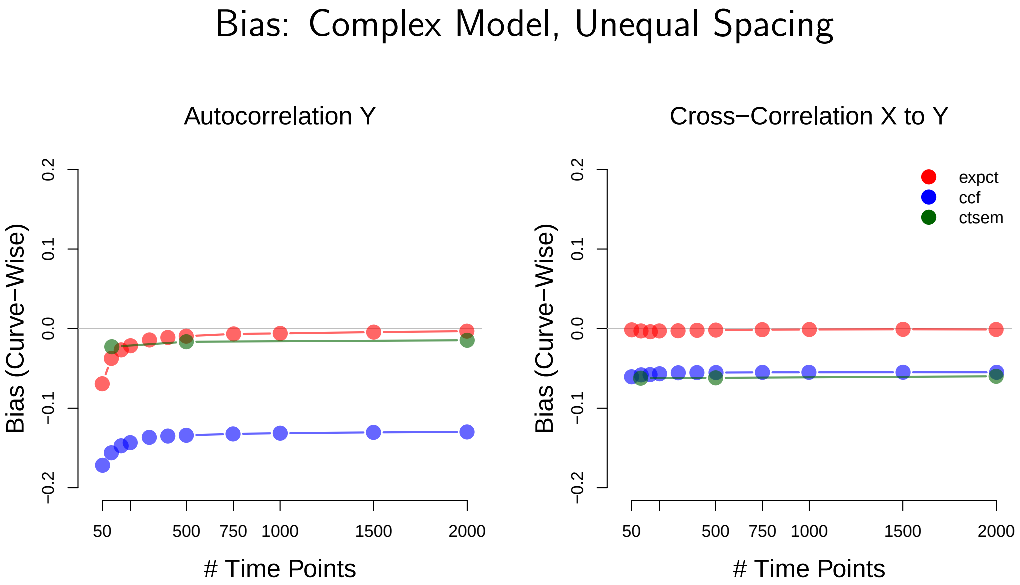
- Point-wise, Simultaneous, Analytic

We use "function-wide" averages to summarize performance:

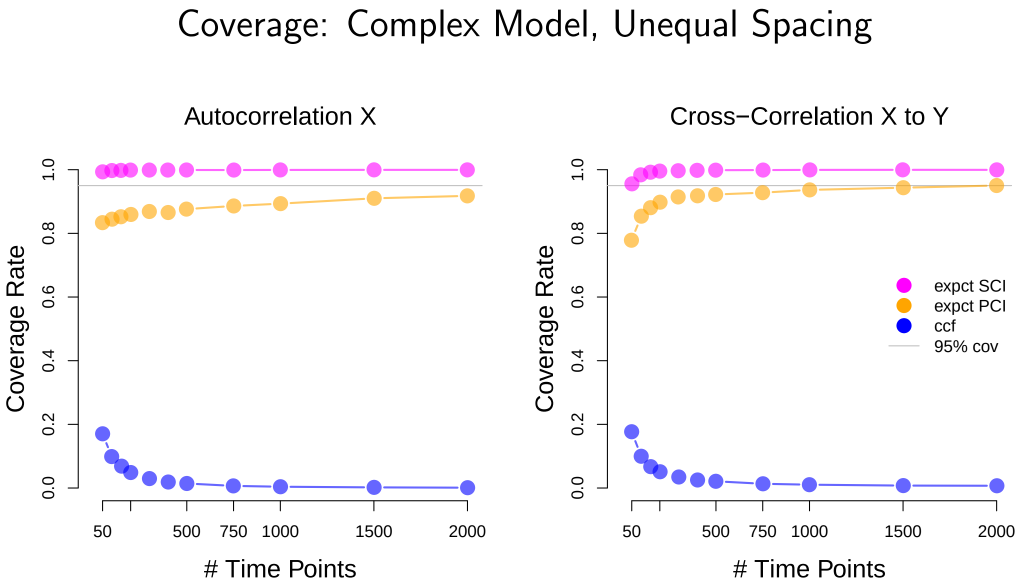
- Function-wide bias: Average distance between true and estimated correlation function evaluated at a range of time-intervals / lags


Bias: Simple Model, Equal Spacing
Autocorrelation Y
Cross−Correlation X to Y
Bias (Curve−Wise)
Time Points
0.2
0.1
0.0
−0.1
−0.2
50
500
750
1000
1500
2000
expct
ccf
ctsem

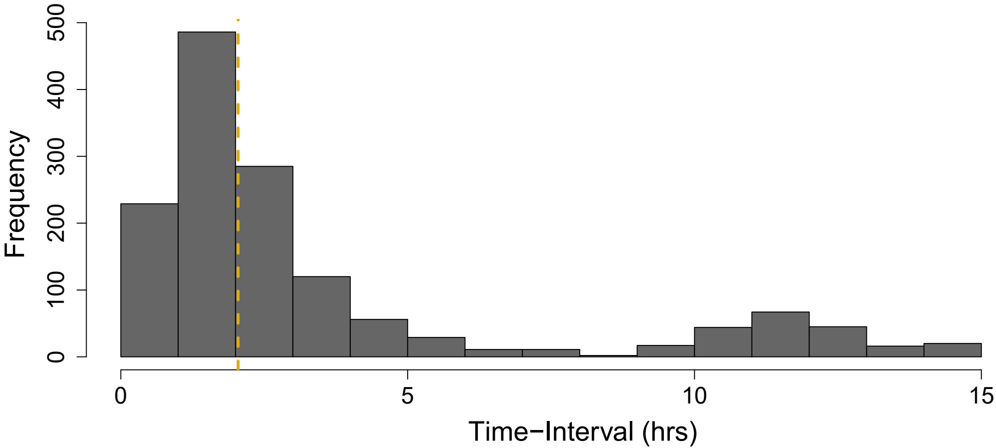

Bias: Complex Model, Equal Spacing
Autocorrelation Y
Cross−Correlation X to Y
Bias (Curve−Wise)
−0.2
−0.1
0.0
0.1
0.2
50
500
750
1000
1500
2000
Time Points
expct
ccf
ctsem



Bias: Complex Model, Unequal Spacing
Autocorrelation Y
Cross−Correlation X to Y
Bias (Curve−Wise)
0.2
0.1
0.0
−0.1
−0.2
50
500
750
1000
1500
2000
Time Points
expct
ccf
ctsem

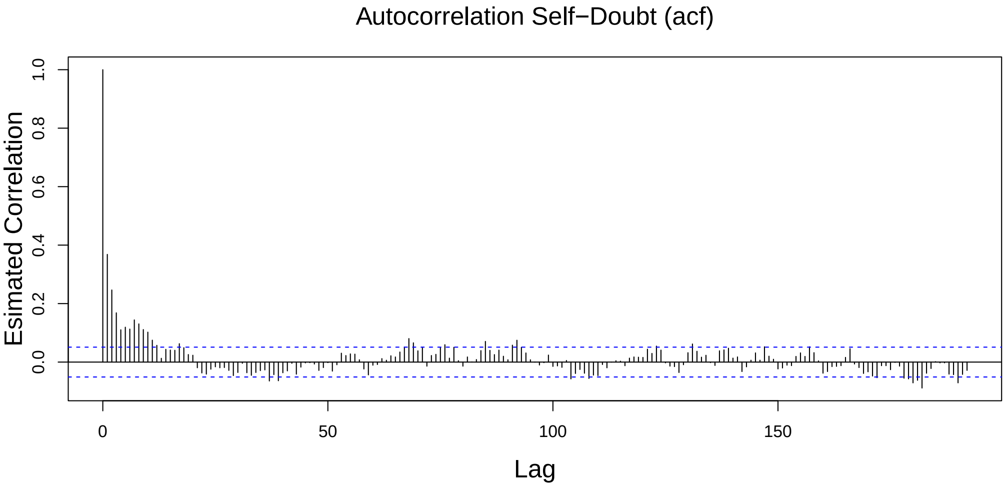

Coverage: Complex Model, Unequal Spacing
Autocorrelation X
Cross−Correlation X to Y
Coverage Rate
Coverage Rate
0.0
0.2
0.4
0.6
0.8
1.0
50
500
750
1000
1500
2000
Time Points
Time Points
expct SCI
expct PCI
ccf
95% cov

Empirical Data

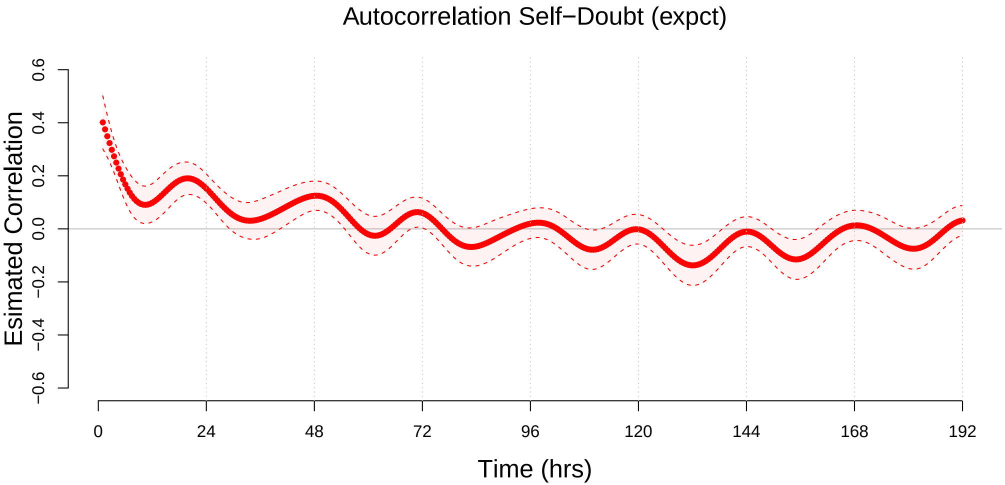


Wichers & Groot (2016)

Empirical Data



Empirical Data



Future work

In principle this method can be used in other situations than those studied here

- Systems of variables measured at different timescales (e.g., daily diary vs hourly ratings vs minute-to-minute physiological measurements)
- "Panel" data: multi-subject low repeated measures

Extensions in progress:

- Multi-level time-series data (random effects)
- Partial relationships (PACF, PCCF)

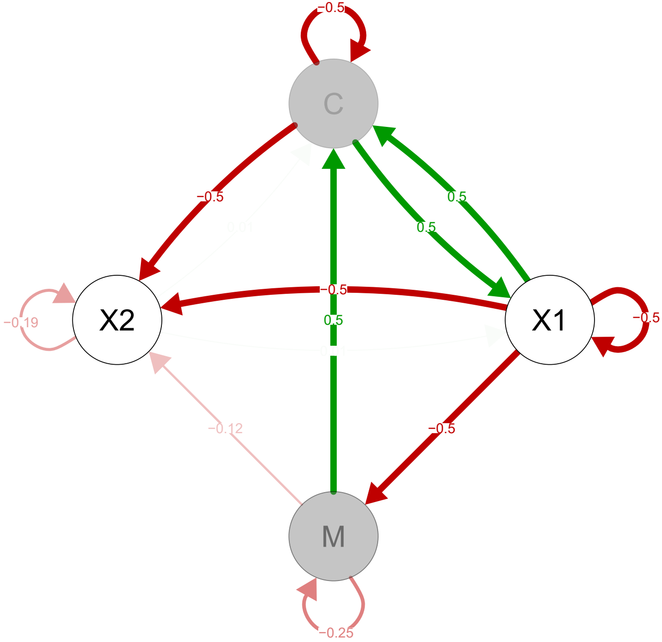
Extracting Dynamic Features from Irregularly Spaced Time Series

expct: Exploratory continuous-time modeling

- Available as an R package github: ryanoisin/expct
- Overcomes equal-interval limitation of traditional ACF/CCF estimation
- Avoids reliance on correct lagged model specification in confirmatory continuous-time models
  - *ctsem*, *dynr*

Ryan O., Wu, K., & Jacobson, N.K. (in preperation). Exploratory Continuous-Time Modeling (expct): Extracting Dynamic Features from Irregularly Spaced Time Series

oisin.ryan.stats@gmail.com | oisinryan.org



−0.5
C
−0.5
0.01
0.5
0.5
0.5
−0.5
X2
X1
−0.19
0.5
−0.5
−0.12
−0.5
M
−0.25