Exploratory Continuous-Time Modeling *expct*: Extracting Dynamic Features from Irregularly Spaced Time Series

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

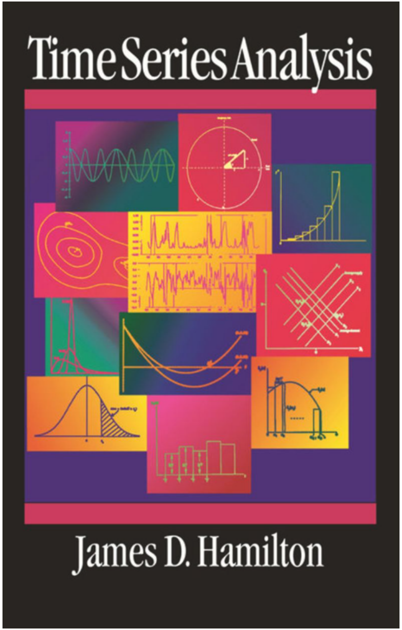
[1]*Department of Methodology and Statistics, Utrecht University, NL*
[2]*Department of Mathematics, University of California, San Diego*
[3]*Departments of Biomedical Data Science, Psychiatry, and Computer Science, Dartmouth College*

CMstatistics 2022


POLAR
120
80
70
MEM
START STOP
SET
I feel anxious right now
0
(not at all)
10
(very much)
I want to binge eat
0
(not at all)
10
(very much)

Time Series Analysis

James D. Hamilton



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

$\dots$

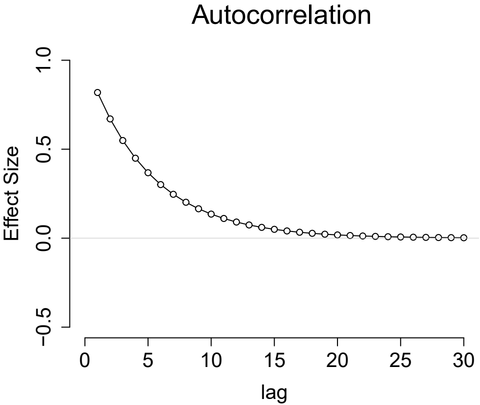
y_T

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

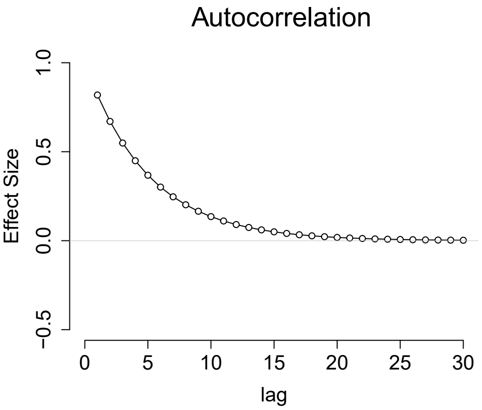
Y	Y at lag 1	Y at lag 2
y_1		
y_2	y_1	
y_3	y_2	y_1
y_4	y_3	y_2
y_5	y_4	y_3
y_6	y_5	y_4
y_7	y_6	y_5
y_8	y_7	y_6
$\dots$	$\dots$	$\dots$
y_T	y_{T-1}	y_{T-2}
	y_T	y_{T-1}
		y_T

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
$\ldots$	$\ldots$	$\ldots$
y_T	y_{T-1}	y_{T-2}
	y_T	y_{T-1}
		y_T

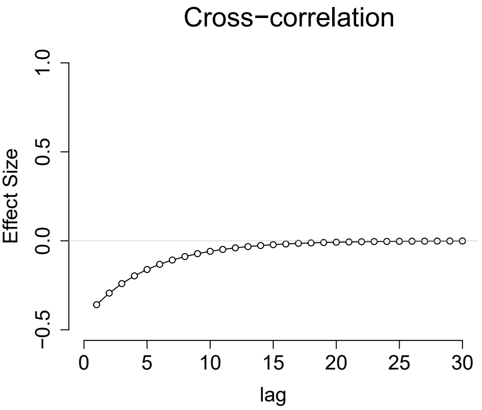
Y	Y at lag 1	Y at lag 2
y_1		
y_2	y_1	
y_3	y_2	y_1
y_4	y_3	y_2
y_5	y_4	y_3
y_6	y_5	y_4
y_7	y_6	y_5
y_8	y_7	y_6
$\dots$	$\dots$	$\dots$
y_T	y_{T-1}	y_{T-2}
	y_T	y_{T-1}
		y_T



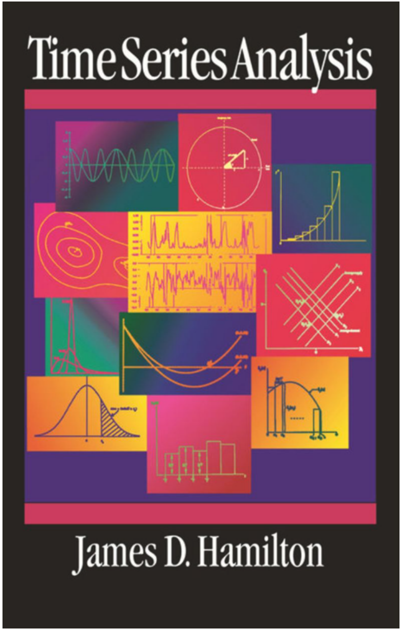
Autocorrelation
Effect Size
−0.5
0.0
0.5
1.0
0
5
10
15
20
25
30
lag



Autocorrelation
Effect Size
1.0
0.5
0.0
−0.5
0
5
10
15
20
25
30
lag



Cross−correlation
Effect Size
1.0
0.5
0.0
−0.5
0
5
10
15
20
25
30
lag

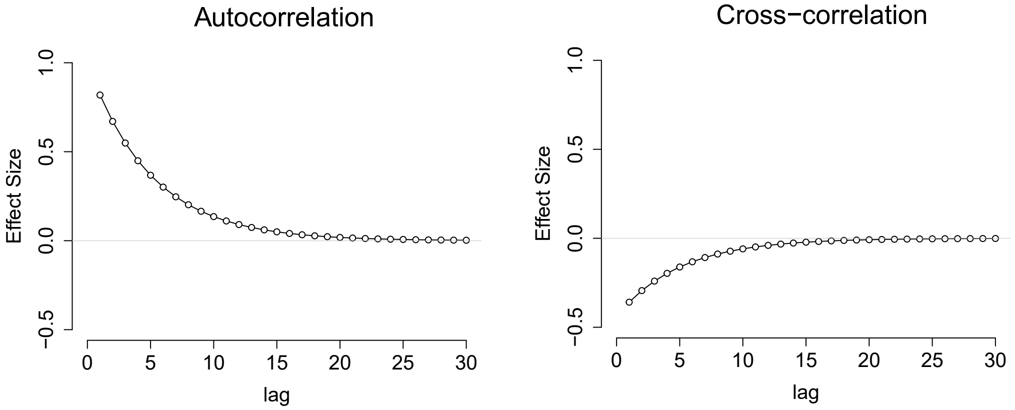


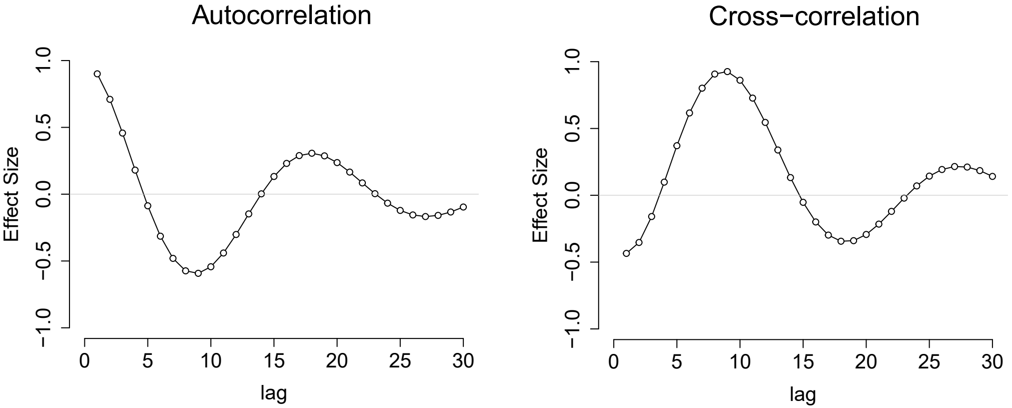
Descriptive / Exploratory Tools

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

Models such as (V)ARIMA

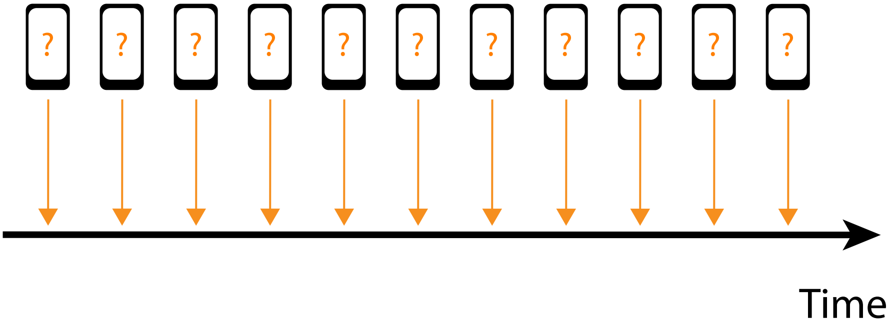
- AR(1), AR(2), VAR(p)
- ACF & CCF to check / find appropriate model


Autocorrelation
Effect Size
lag
Cross−correlation
Effect Size
lag

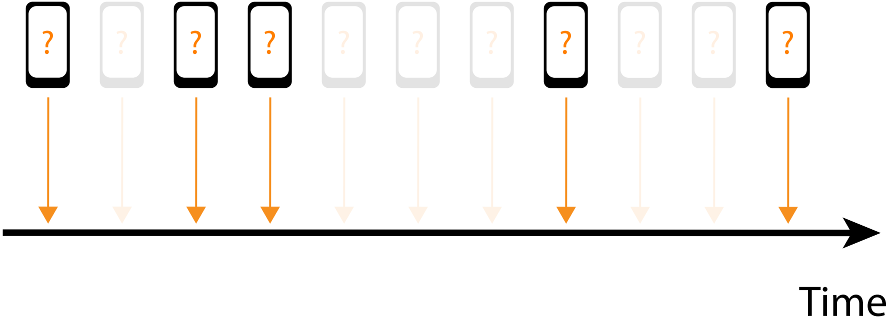


Autocorrelation
Effect Size
lag
1.0
0.5
0.0
−0.5
−1.0
0
5
10
15
20
25
30
Cross−correlation
Effect Size
lag
1.0
0.5
0.0
−0.5
−1.0
0
5
10
15
20
25
30

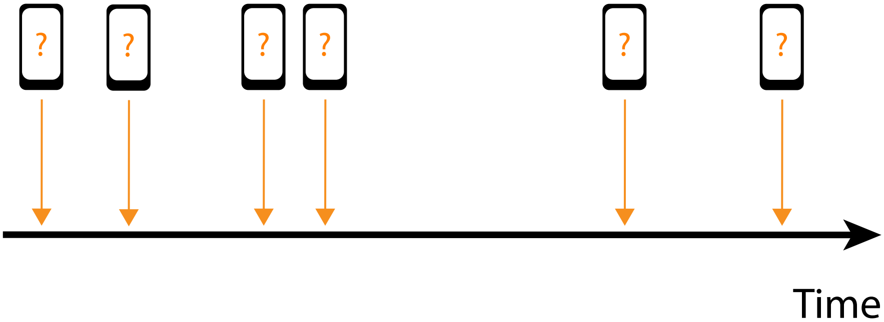
Assumption: Equally Spaced Measurements



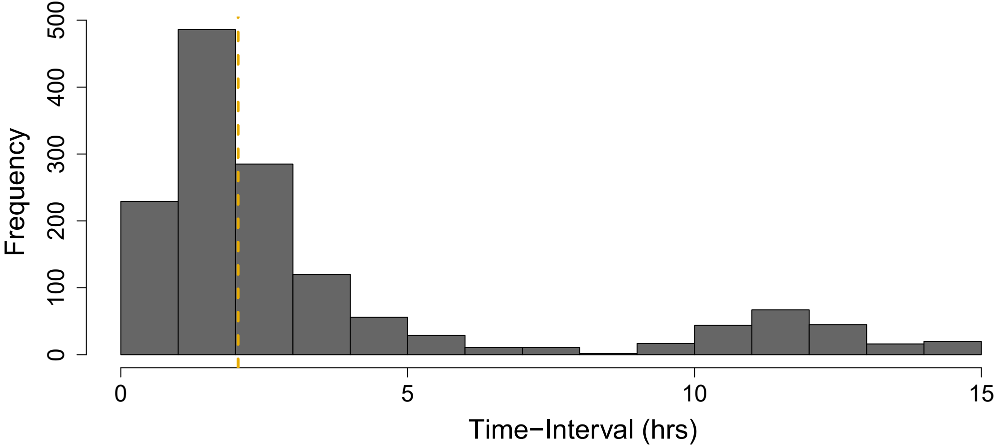
Reality: Irregularly Spaced Measurements



Reality: Irregularly Spaced Measurements

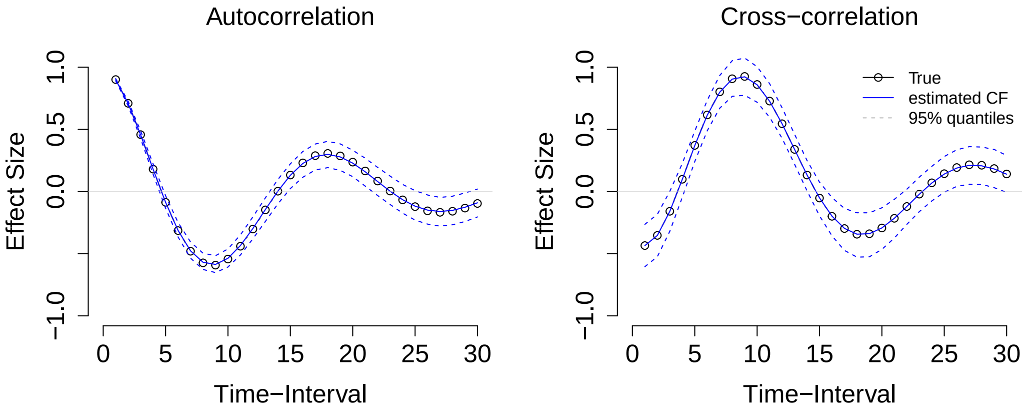


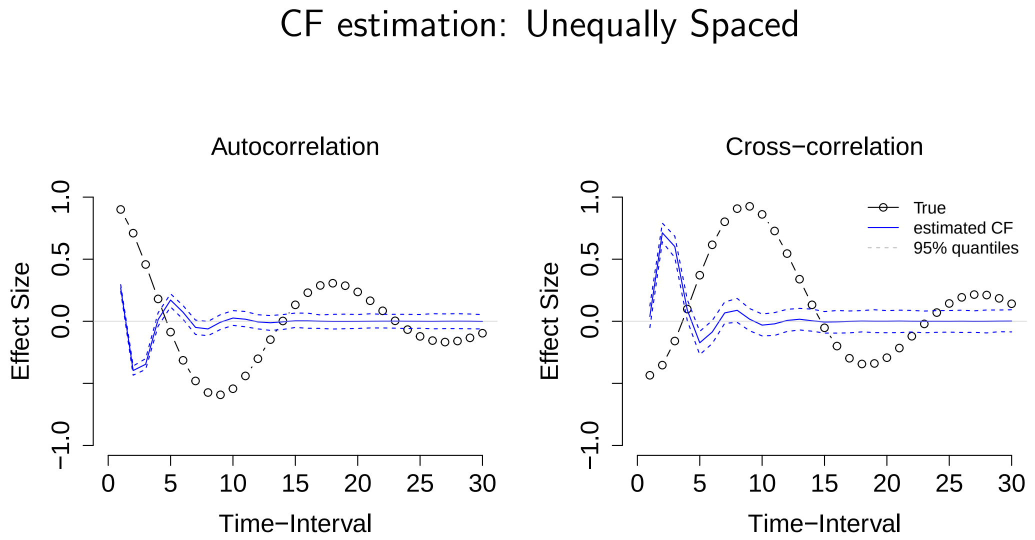
Empirical data



Ryan & Hamaker (2022); Kossakowski, Groot, Haslbeck, Borsboom & Wichers (2017)

CF estimation: Equally Spaced




CF estimation: Unequally Spaced
Autocorrelation
Cross−correlation
Effect Size
Time−Interval
True
estimated CF
95% quantiles
1.0
0.5
0.0
−1.0
0
5
10
15
20
25
30

Continuous-Time Modeling

So far, people often suggest using **Continuous-Time** models

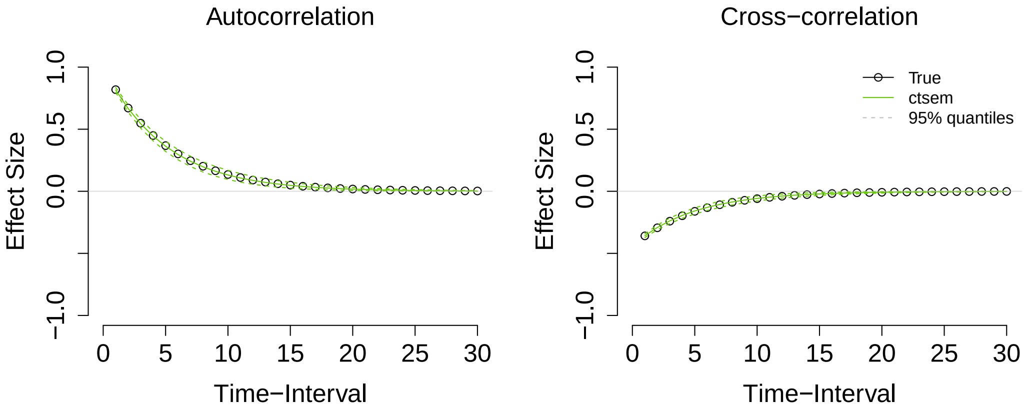
$$\frac{d\boldsymbol{Y}(t)}{dt} = \boldsymbol{A}\ \boldsymbol{Y}(t) + \boldsymbol{G}\frac{d\boldsymbol{W}(t)}{dt}$$

In auto-regressive form, this looks like a VAR(1) model

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

Auto- and cross-regressions are a specific **non-linear** function (matrix exponential $\boldsymbol{e}$) of the time-interval between measurements Δt.

- This model can be estimated, e.g., using the *ctsem* package (Driver et al. 2017) ctsem estimation: unequally spaced



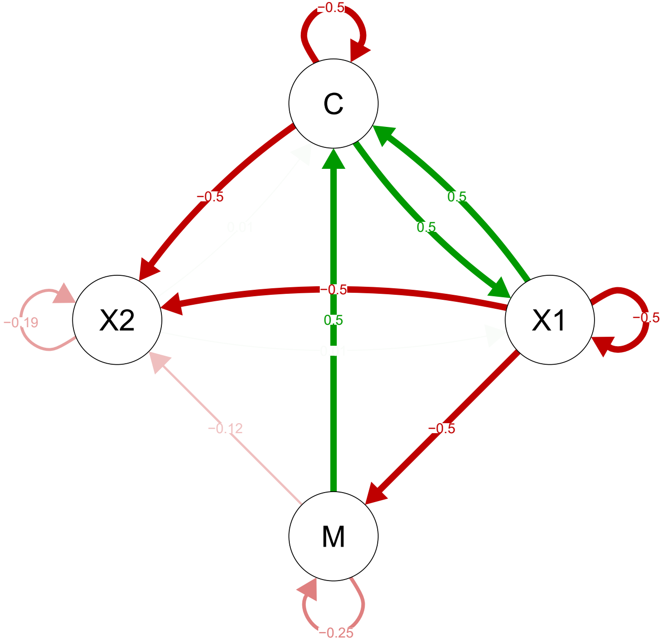
Problem: Model Misspecification

With a CT model, the auto- and cross- relations are **derived** based on the estimated drift matrix $\boldsymbol{A}$.

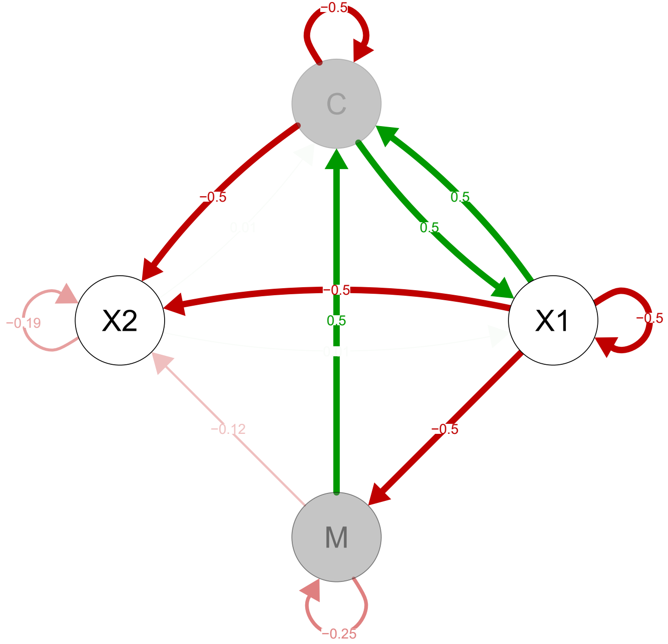
- Model-based estimate. Not (entirely) data-driven / exploratory

But this will only match **reality** if the (simple, low-D, linear) model is **correctly specified**.

- If the **order** of the model (first vs second) is wrong, or if we have **unobserved confounding** these will be incorrect

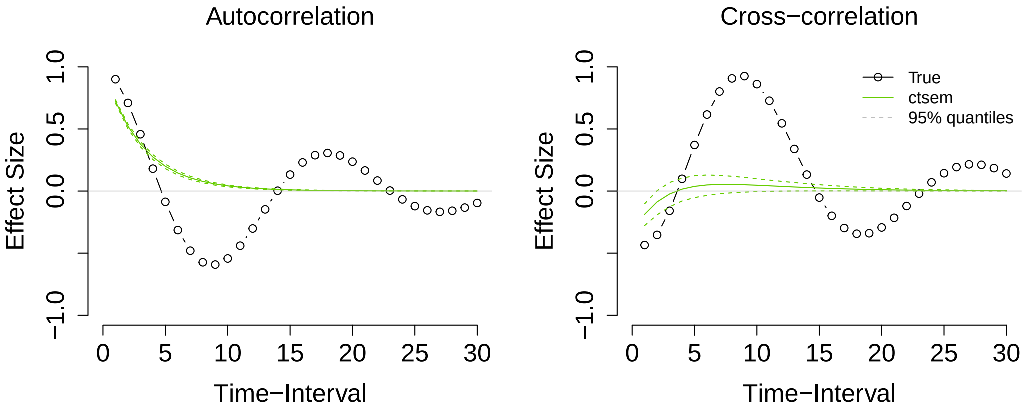


−0.5
C
−0.5
0.01
0.5
0.5
X2
−0.5
X1
−0.19
0.5
−0.5
−0.12
−0.5
M
−0.25



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

ctsem estimation: model misspecified



Traditional ACF and CCF estimation:

- **Data-driven and exploratory** (relatively model-free) method for exploring dynamic features
- Does not perform well with irregularly spaced data

CT model estimation:

- Can be estimated from irregularly spaced data and in principle capture non-linear patterns of correlations
- But relies on unrealistic assumption of correct model specification, which likely never holds in practice
- Kicker: without a data-driven way of computing correlations, no way to **check** the model misspecification

expct: Exploratory Continuous Time Modeling

Estimate ACF and CCF functions from data taken with any arbitrary sampling scheme

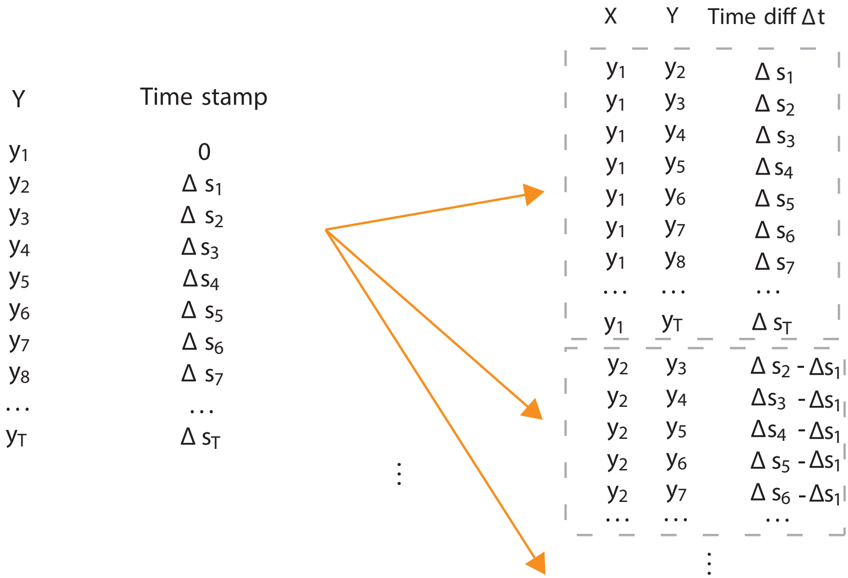
- Development version available github: ryanoisin/expct

Two-step procedure

1. Create a “stacked” data frame: Every observation acts as a predictor for every future observation, with the time-interval Δt as additional variable

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

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
. . . . . .
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
. . . .
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 $\Delta s_6 - \Delta s_1$
. . . .

expct: Exploratory Continuous Time Modeling

Estimate ACF and CCF functions from data taken with any arbitrary sampling scheme

- Development version available github: ryanoisin/expct

Two-step procedure

1. Create a "stacked" data frame: Every observation acts as a predictor for every future observation, with the time-interval Δt as additional variable
2. Use a Generalized Additive Mixed Models (GAMM) to estimate auto- and cross-correlations

$$Y = f(\Delta t)X + \epsilon$$

expct: Exploratory Continuous Time Modeling

Estimate ACF and CCF functions from data taken with any arbitrary sampling scheme

- Development version available github: ryanoisin/expct

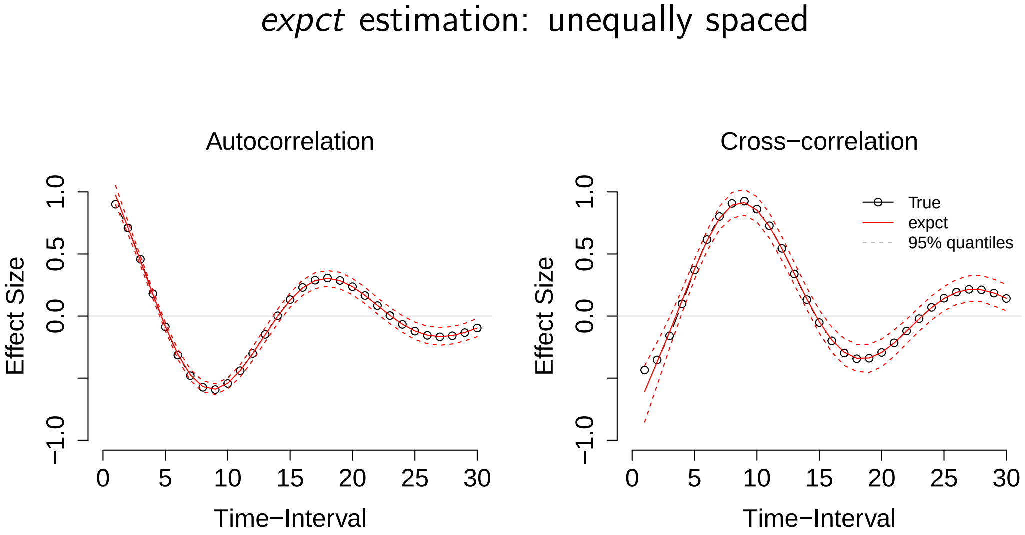
Two-step procedure

1. Create a "stacked" data frame: Every observation acts as a predictor for every future observation, with the time-interval Δt as additional variable
2. Use a Generalized Additive Mixed Models (GAMM) to estimate auto- and cross-correlations

$$Y = f(\Delta t)X + \epsilon$$

Auto and cross-correlation functions are estimated for arbitrary Δt by fitting seperate bivariate GAMMs. In this way we approximate

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


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

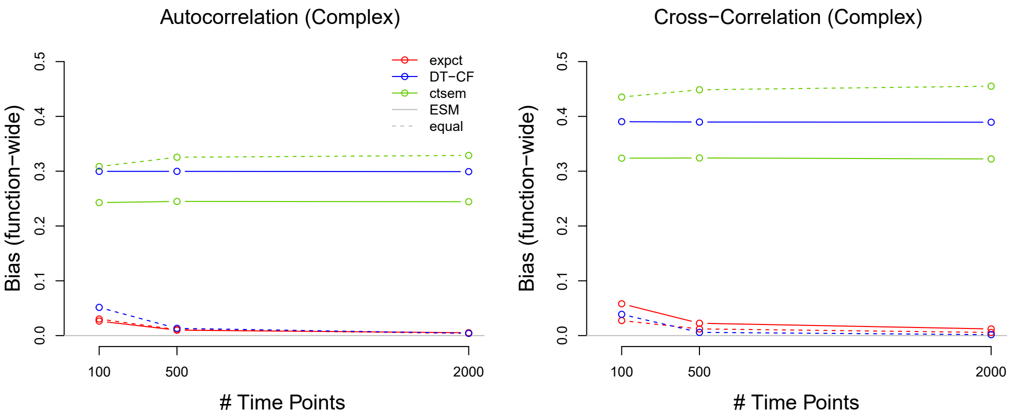
Simulation Study

Time-series length: [100, 500, 2000]
Sampling Scheme: [Equal, “ESM” bimodal, Uniform]
Data-generating models: [AR(1), Bivariate Oscillating, Misspecified]

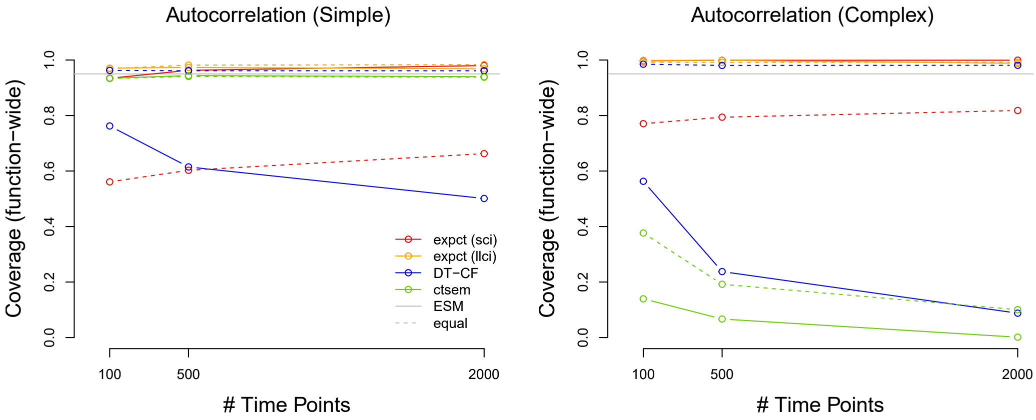
We also compared a number of different methods for computing confidence intervals:

- Point-wise, Simultaneous, Simultaneous + large lag error correction, Bootstrap

We use “function-wide” averages to compute bias, coverage, etc.



Autocorrelation (Complex)
Cross−Correlation (Complex)
Bias (function−wide)
0.0
0.1
0.2
0.3
0.4
0.5
expct
DT−CF
ctsem
ESM
equal
100
500
2000
Time Points


Autocorrelation (Simple)
Autocorrelation (Complex)
Coverage (function−wide)
0.0
0.2
0.4
0.6
0.8
1.0
100
500
2000
Time Points
expct (sci)
expct (llci)
DT−CF
ctsem
ESM
equal

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)
- Multi-subject low T data common in social sciences, if we assume shared ACF and CCF

Extensions TBD:

- Multi-level multi-subject data (random effects, work underway)
- Partial relationships (PACF, PCCF)
- Empirical Examples

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*
- GAMM-based, unbiased, good coverage with novel llci method

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

o.ryan@uu.nl | oisinryan.org