Stat 565

Other Topics

Mar 8 2016

Charlotte Wickham

stat565.cwick.co.nz

Today

Announcements

Some other topics:

- Missing data
- Unequally spaced data
- Longitudinal data

Project time @ 11:05am

Final Exam

Monday March 14th 0930-1120 GLSN 100 (here) **Don't forget daylight savings!**

Similar to last time's final:

Q1. Identify SARIMA model

Q2. Frequency domain analysis

Q3. Regression with time series errors

Q4. ? (conceptual)

Office hours

This week as per usual

(Tu 2-3, W 1-2, Th 2-3)

+ Friday 1-2pm

Thursday

Don’t forget peer eval forms

Presentations: Matt’s, Mike’s, Jeremiah’s,

Review

Missing data

With time series data, it is best to be explicit about missing data:

```
> with_miss[19:23, ]
         date         x
19 2015-01-22 -1.997787
20 2015-01-23 -2.351314
21 2015-01-24 -1.960690
22 2015-01-26 -3.330436
23 2015-01-29 -3.762846
```

```
> explicit_miss[22:29, ]
         date         x
22 2015-01-22 -1.997787
23 2015-01-23 -2.351314
24 2015-01-24 -1.960690
25 2015-01-25        NA
26 2015-01-26 -3.330436
27 2015-01-27        NA
28 2015-01-28        NA
29 2015-01-29 -3.762846
```

Why? You will find out quickly when missing values are a problem

If missing values are a problem…

Some objects can still be estimated even with missing values, (i.e. `na.action = na.pass`, in `acf`), but be aware of what is being done with them.

Imputation may be appropriate.

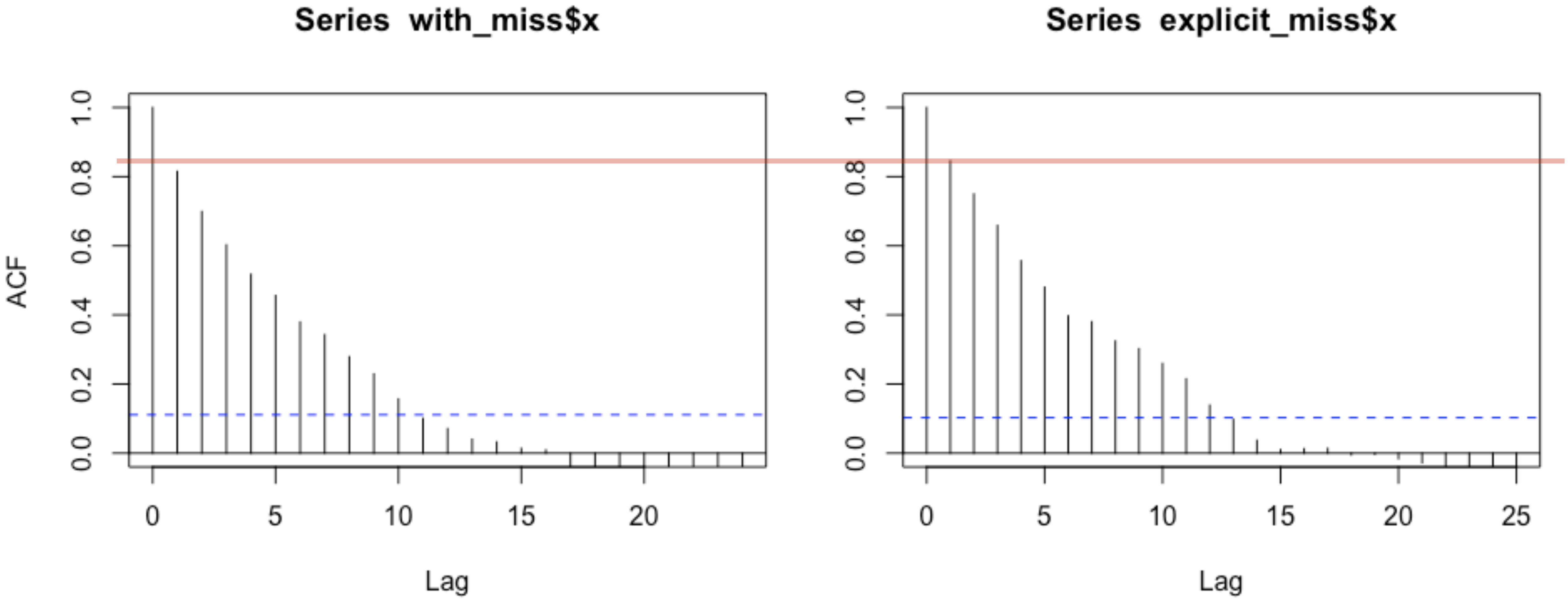
Missing at random?

E.g. sample acf

```
> acf(explicit_miss$x)
Error in na.fail.default(as.ts(x)) : missing values in object
```

```
acf(with_miss$x, ylim = c(0, 1))
```

```
acf(explicit_miss$x, ylim = c(0, 1),
    na.action = na.pass)
```

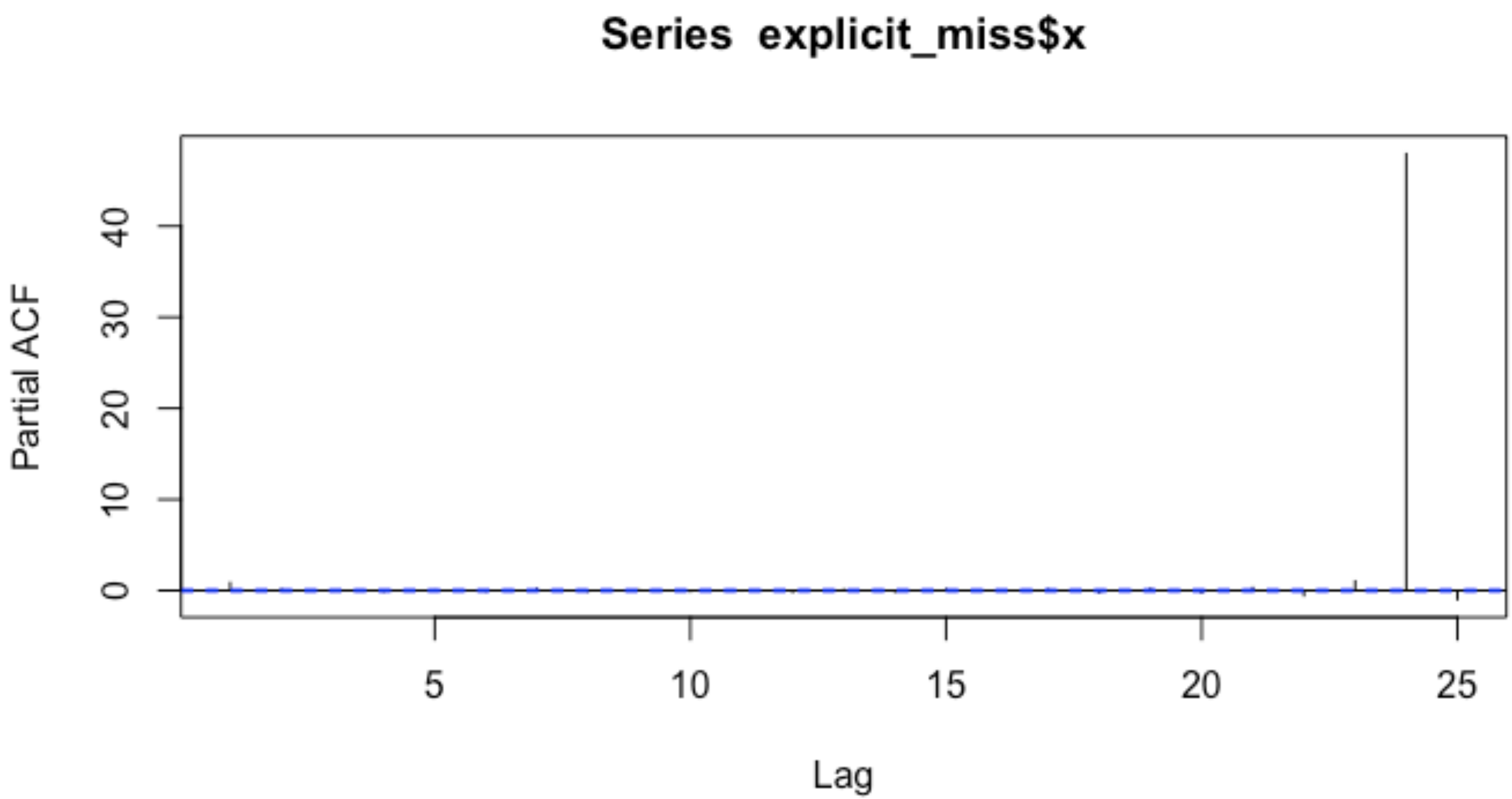


```
> with_miss[19:23, ]
         date         x
19 2015-01-22 -1.997787
20 2015-01-23 -2.351314
21 2015-01-24 -1.960690
22 2015-01-26 -3.330436
23 2015-01-29 -3.762846
```

```
> explicit_miss[22:29, ]
         date         x
22 2015-01-22 -1.997787
23 2015-01-23 -2.351314
24 2015-01-24 -1.960690
25 2015-01-25        NA
26 2015-01-26 -3.330436
27 2015-01-27        NA
28 2015-01-28        NA
29 2015-01-29 -3.762846
```

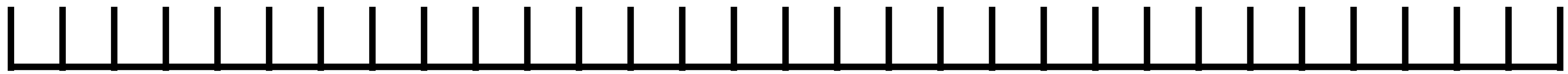
Be careful….

It's your responsibility to investigate how the function is dealing with missing values



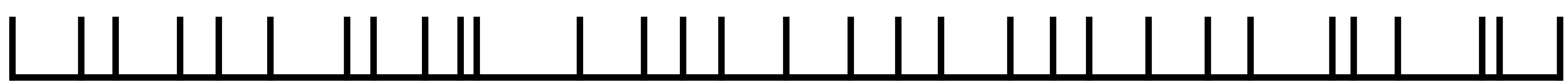
Unequally spaced time series

Equally spaced x_t, $t = 1, 2, \ldots, n$



Unequally spaced (due to missing values)

Unequally spaced $x(t)$, $t \in R$



Continuous time correlation models

The general idea: move from auto-correlation defined at **integer lags**,

$$\text{AR}(1),\ \gamma(h) = \phi^h,\ h = 0, 1, 2, \ldots$$

to auto-correlation function defined at **continuous distances,**

$$\text{CAR}(1)\ \text{model},\ \gamma(s) = \phi^s,\ s \geq 0$$

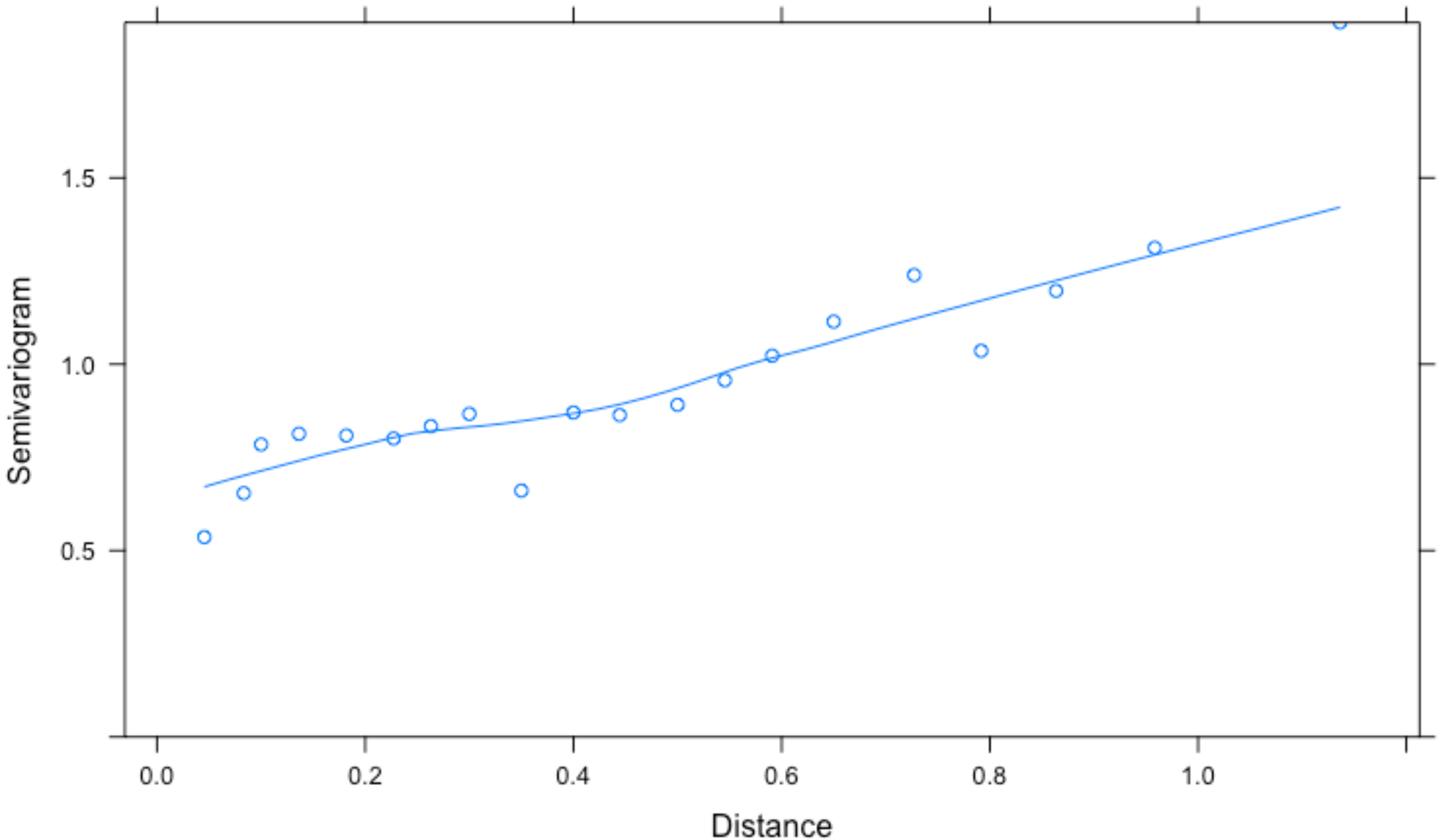
Connection to spatial statistics

The correlogram, $\rho(s)$, is the name of sample version of the ACF in the continuous case.

But by convention people usually look at the (semi)-variogram,

$$V(s) = \gamma(0)\ (1 - \rho(s))$$

this plays the role of the ACF in spatial statistics



Semivariogram
1.5
1.0
0.5
0.0
0.2
0.4
0.6
0.8
1.0
Distance

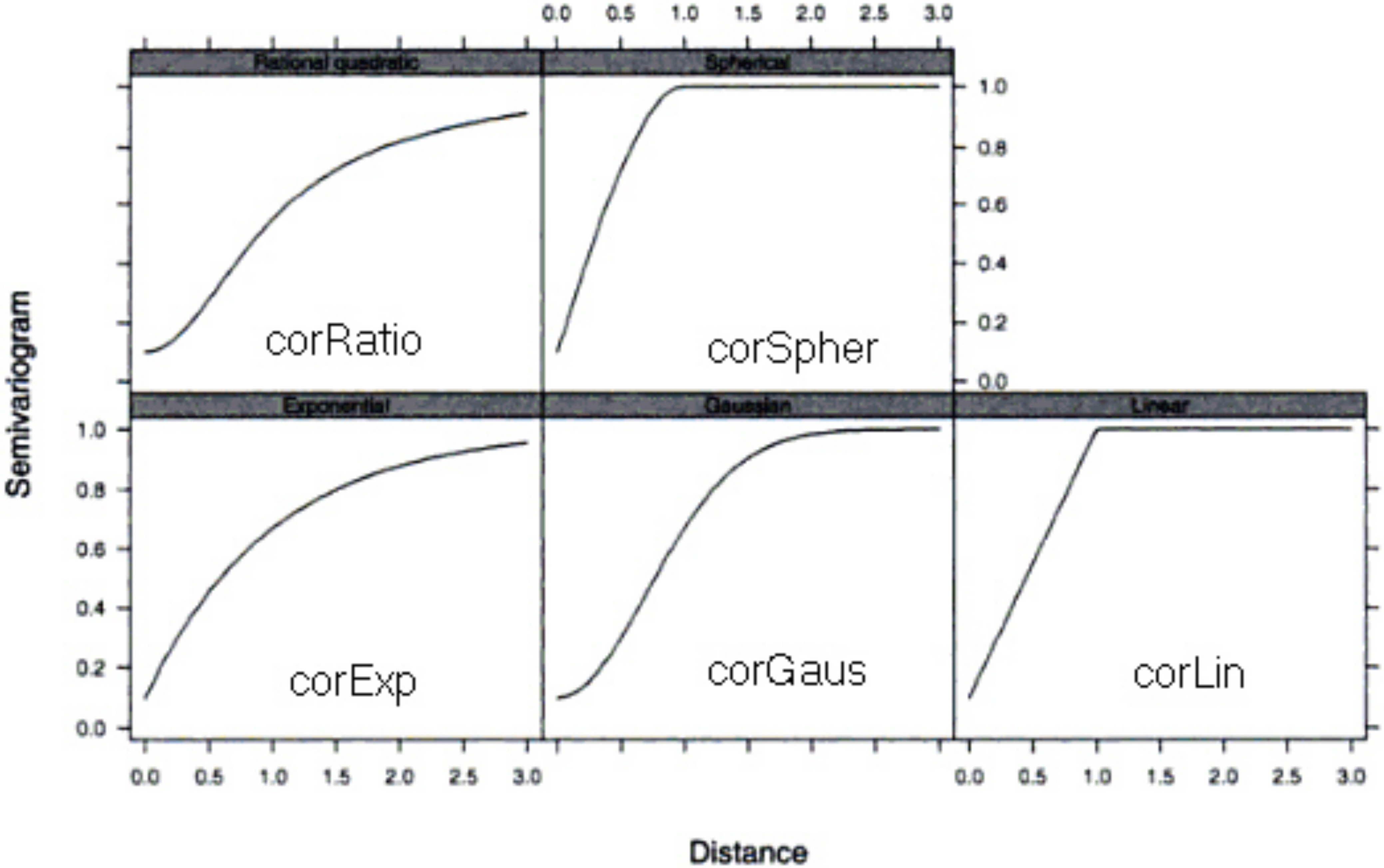


FIGURE 5.9. Plots of semivariogram versus distance for the isotropic spatial correlation models in Table 5.2 with range = 1 and nugget effect = 0.1.

Just like in time series, there are spatial correlation models, that have specified shapes.

Longitudinal Data

Collection of time series, often short.

Usual regression analyses with addition of:

- random effects structure (due to design)
- correlation in time (due to repeated measurements)

Be careful with specifying time correlation structure when you have many series.

Mixed Effects Models in S & S-Plus
Jose C. Pinheiro Douglas M. Bates, Sections 5.1- 5.3

http://search.library.oregonstate.edu/OSU:everything:CP71188863930001451

```
library(nlme)
?Ovary
```

Pierson and Ginther (1987) report on a study of the number of large ovarian follicles detected in different mares at several times in their estrus cycles.

This data frame contains the following columns:

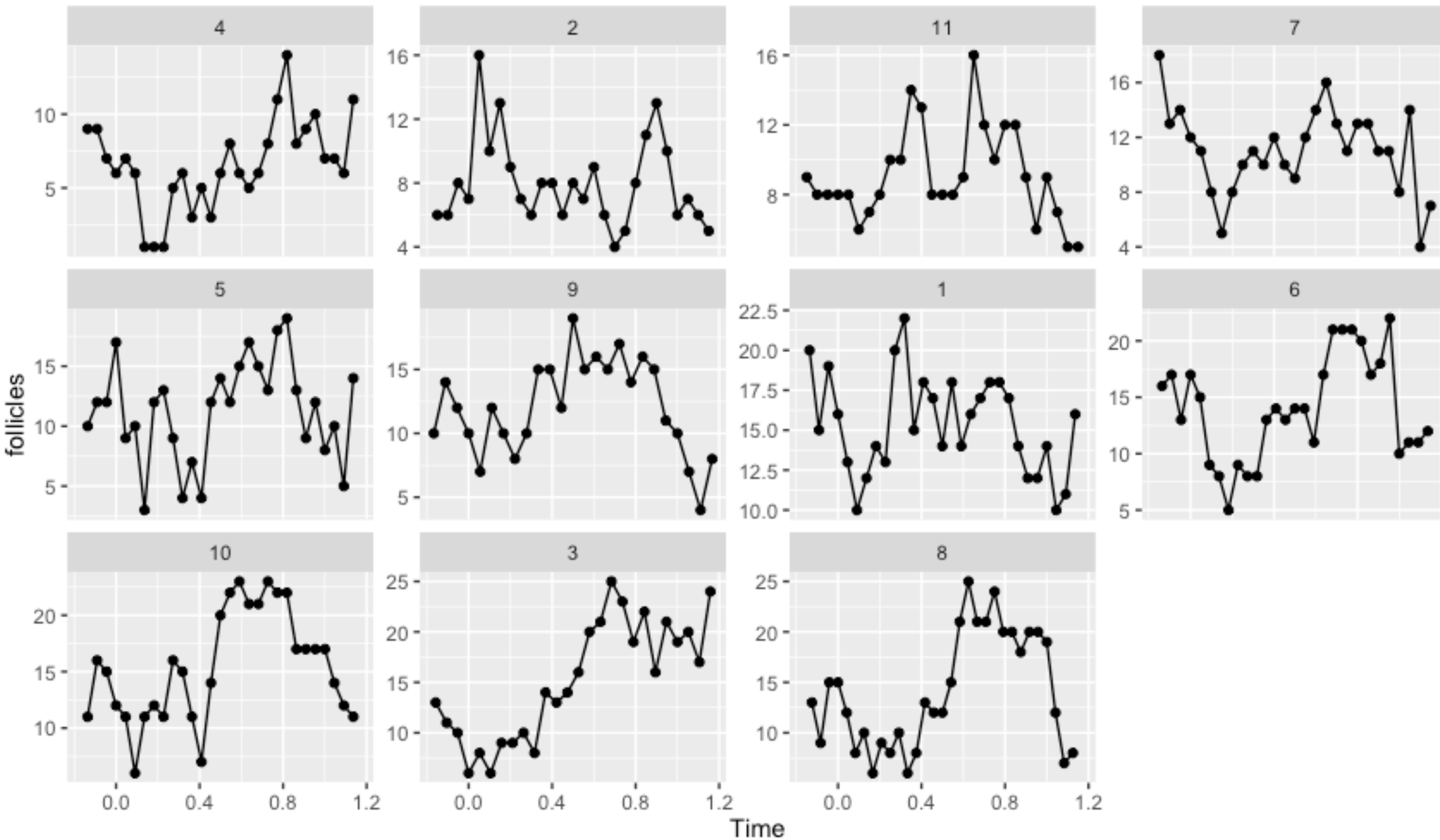
Mare, an ordered factor indicating the mare on which the measurement is made.

Time, time in the estrus cycle. The data were recorded daily from 3 days before ovulation until 3 days after the next ovulation. The measurement times for each mare are scaled so that the ovulations for each mare occur at times 0 and 1.

follicles, the number of ovarian follicles greater than 10 mm in diameter.

```
> head(Ovary)
Grouped Data: follicles ~ Time | Mare
  Mare        Time follicles
1    1 -0.13636360        20
2    1 -0.09090910        15
3    1 -0.04545455        19
4    1  0.00000000        16
5    1  0.04545455        13
6    1  0.09090910        10
              ...
27   1  1.045455          10
28   1  1.090909          11
29   1  1.136364          16
30   2 -0.150000           6
31   2 -0.100000           6
32   2 -0.050000           8
```

What could go wrong if we fed the follicles column in to the acf?



4
2
11
7
5
9
1
6
10
3
8
follicles
Time
0.0
0.4
0.8
1.2

An example

Go through example in R.

Tentative model for *i*th mare at *j*th time

$y_{ij} = (\beta_0 + b_{0i}) +$

$\quad (\beta_1 + b_{1i}) \sin(2\pi t_{ij}) + \beta_2 \cos(2\pi t_{ij}) + \varepsilon_{ij}$

b_{0i}, b_{1i} are random effects for mare

ε_{ij} usual errors