Stat 565

Spectrum

Feb 18 2016

Charlotte Wickham

stat565.cwick.co.nz

Correlated errors model

Works for incorporating trend and seasonality estimates too.

Think back to week 1 & 2

$$y_t = m_t + s_t + Z_t$$

Variable measured at time t

Trend

Seasonality

Noise

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_p x_{tp} + Z_t$$

write our trend and seasonal parts as linear functions

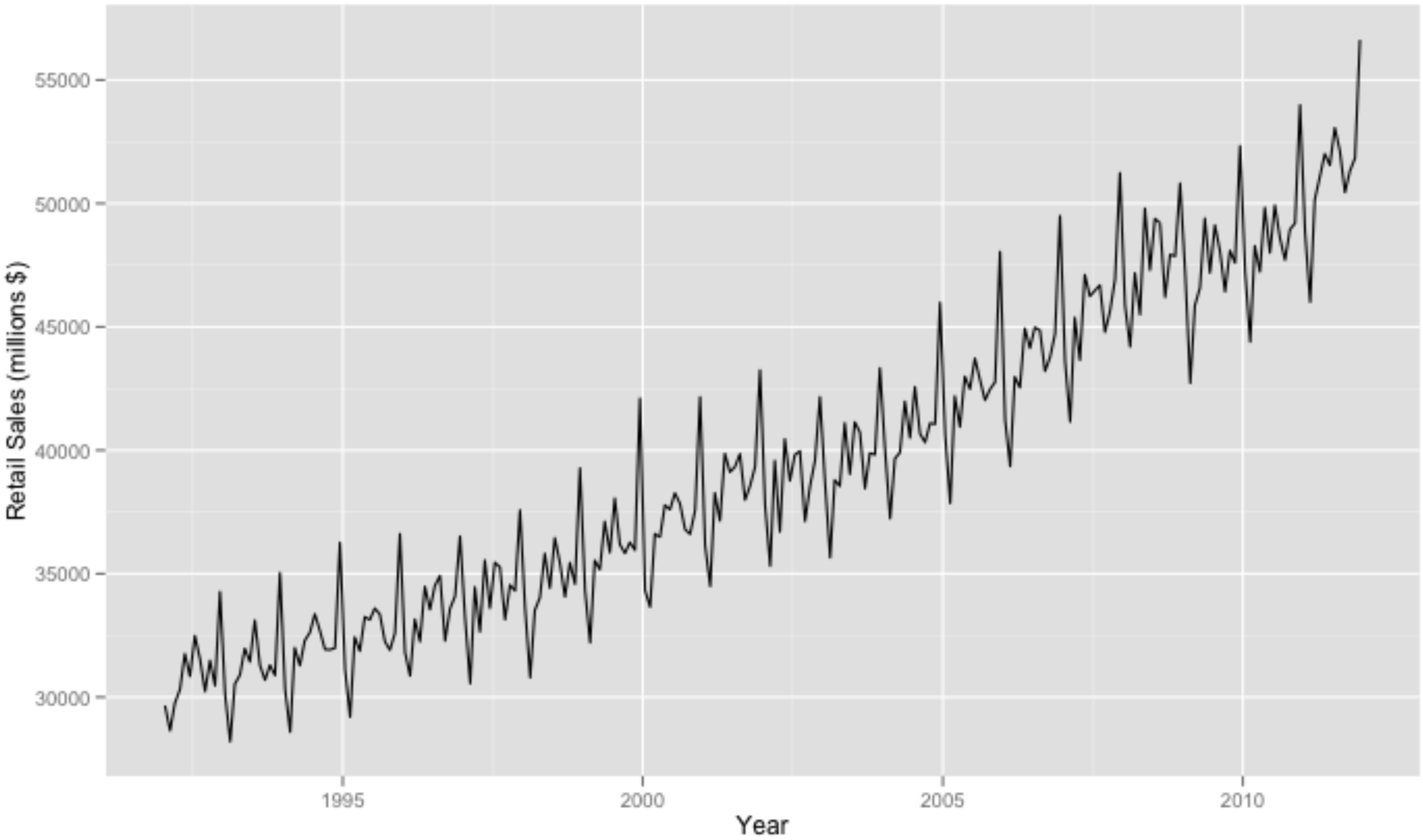
ARMA(p,q)

Trend and seasonality in regression models

Now we have lots of options:

trend	**seasonality**
linear function of time (i.e. straight line)	fixed effects for season (months, quarters etc.)
polynomial function of time	sinusoids
splines	splines
smooths?	smooths?

Retail sales in food and beverage stores



Model

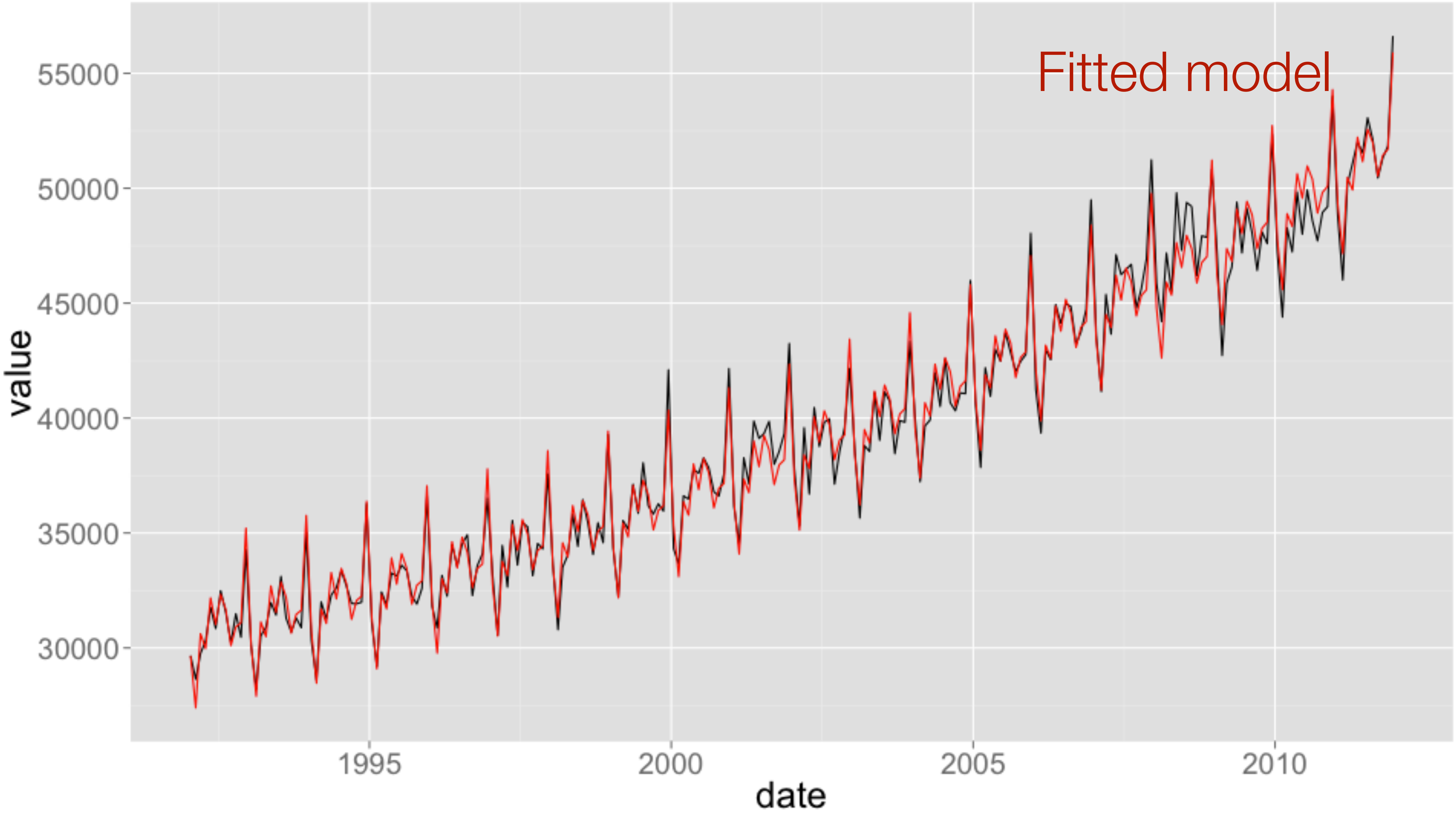
$$\text{sales}_t = \alpha + \beta_1 t + \beta_2 t^2 + \mu_{\text{month}} + z_t$$

z_t is AR(3)

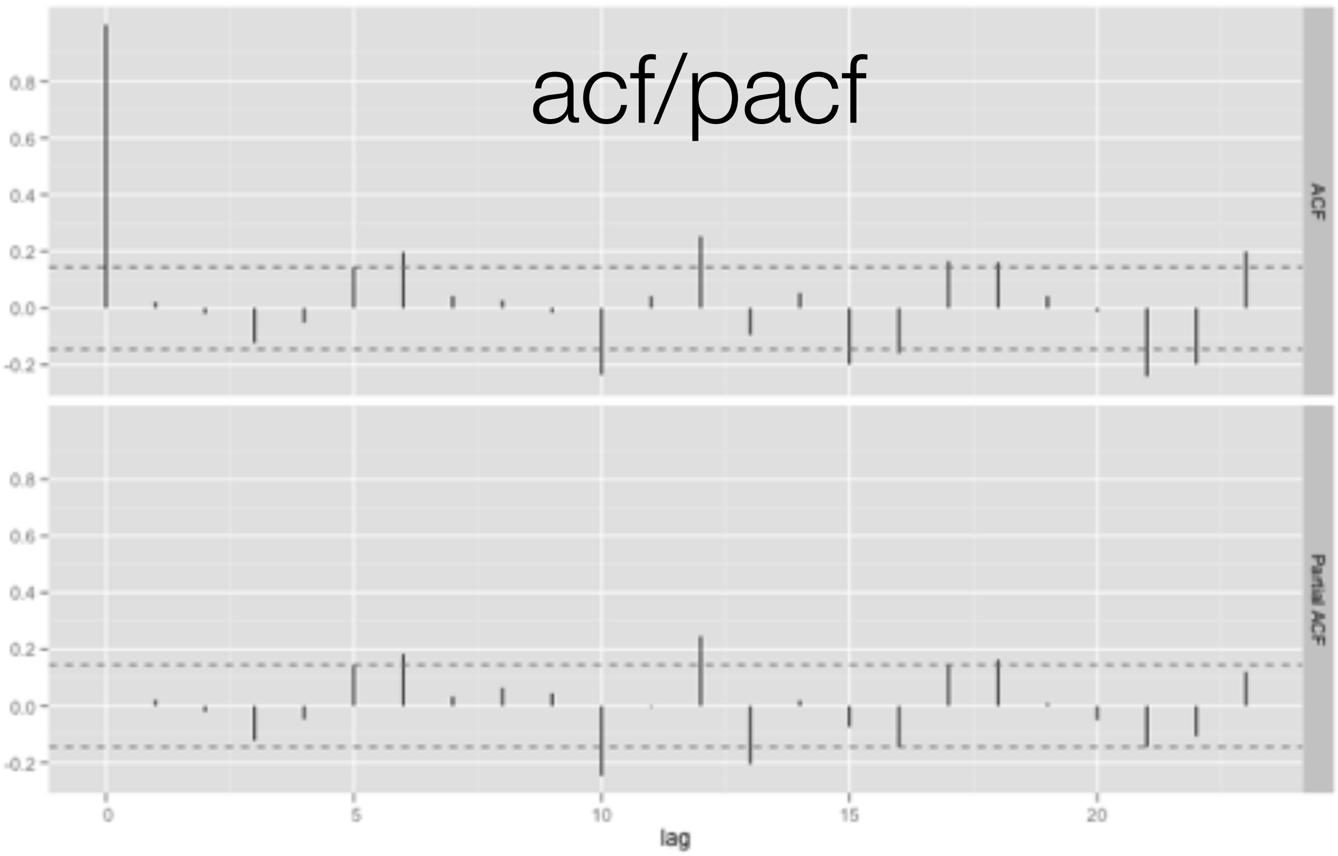
$$z_t = \phi_3 z_{t-3} + \phi_2 z_{t-2} + \phi_1 z_{t-1} + w_t$$

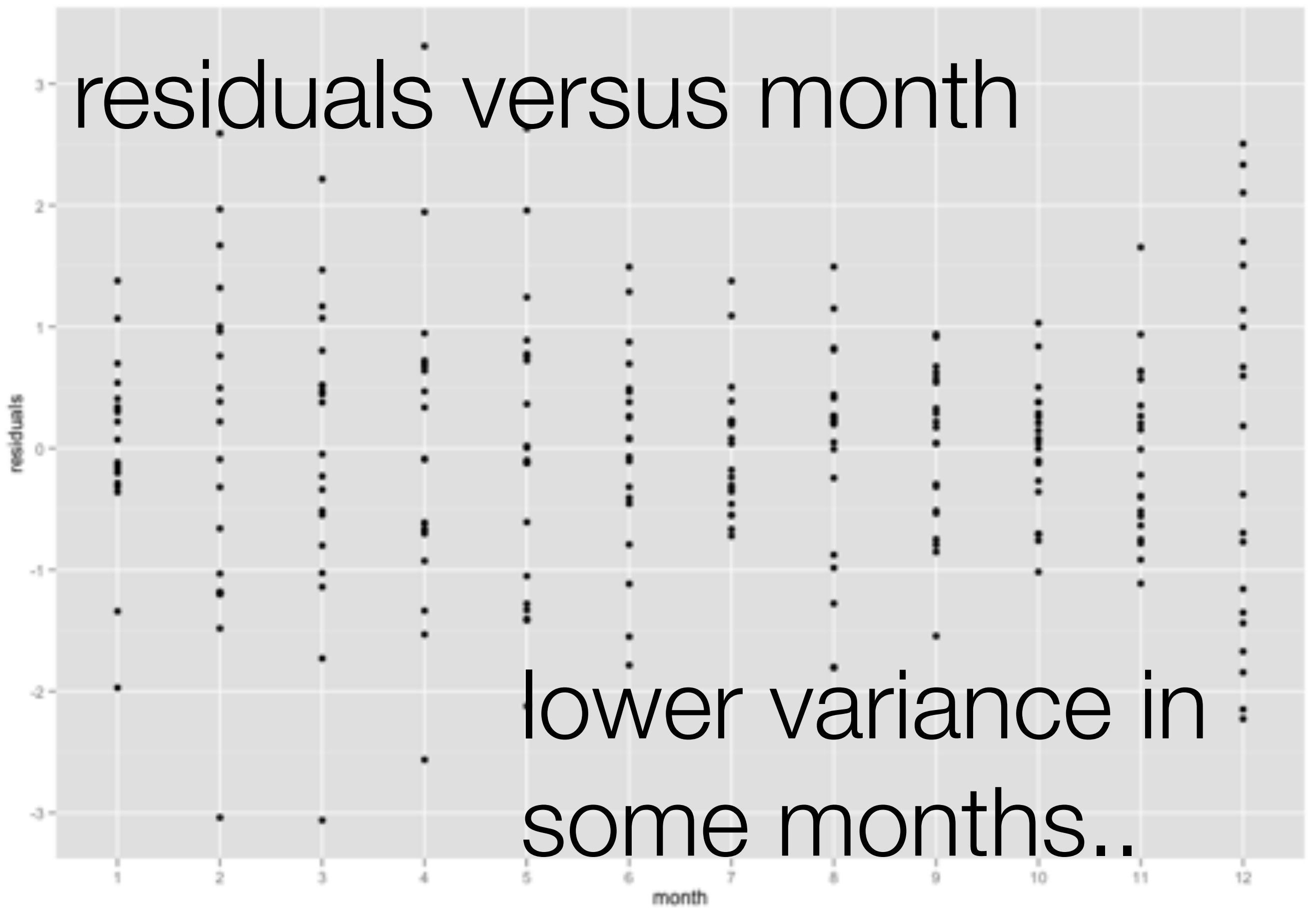
$$w_t \sim \text{Normal}(0, \sigma^2)$$

Fitted model

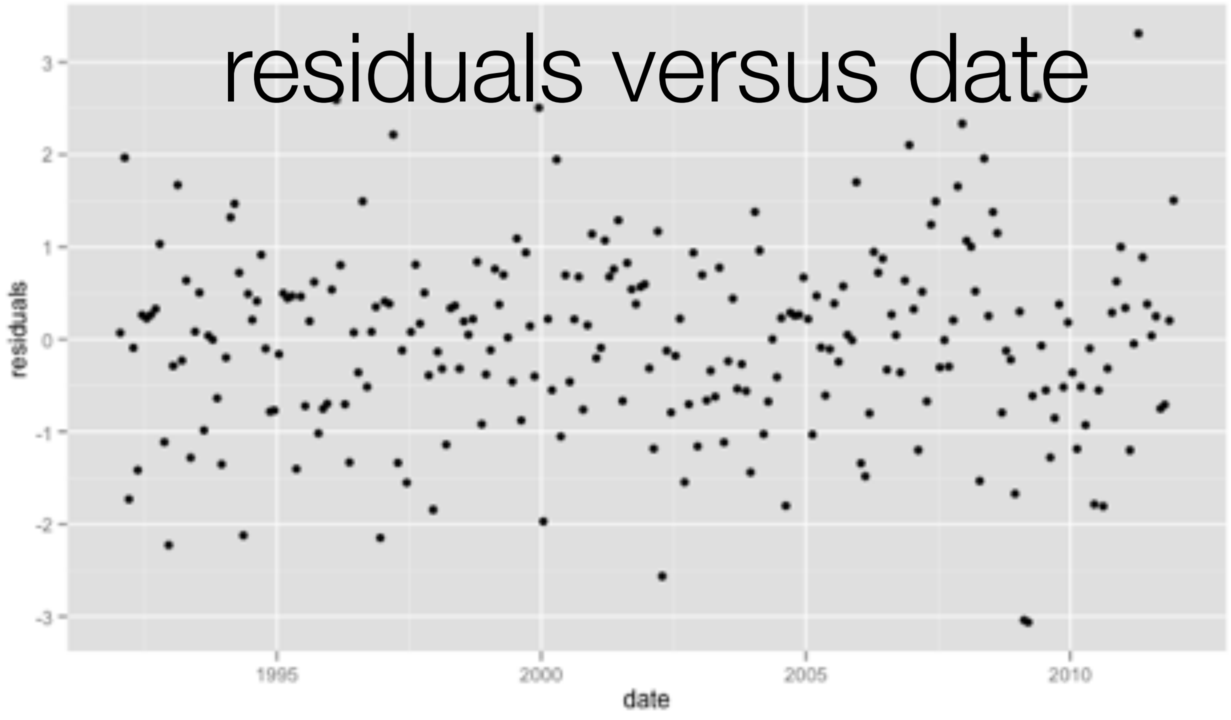


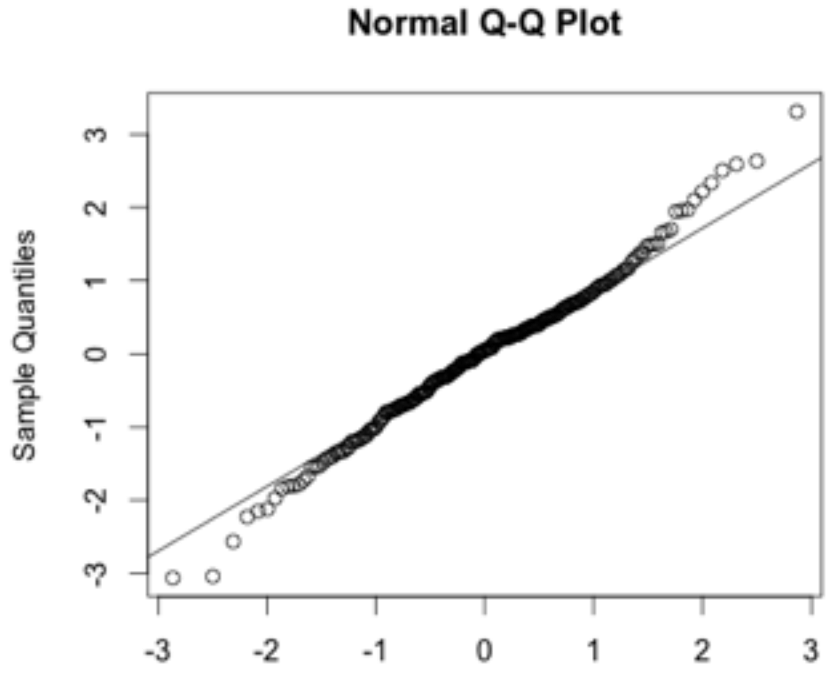
Diagnostics


acf/pacf


residuals versus month

lower variance in some months..


residuals versus date



Using smooths in linear models

Generalized additive models (GAM)

are an extension of linear models that removes the linearity assumption, i.e.

$$y_t = f(x_t) + z_t$$

where f is a smooth function

If z_t is not white noise, then it's called a generalized additive mixed model (GAMM)

`gam` and `gamm` functions in the `mgcv` package

ARMA models for stationary data provide a flexible way to explain autocorrelation structure

SARIMA	Regression with correlated errors
Difference to remove non-stationarity	Directly model non-stationarity
Use ARMA model to capture autocorrelation after differencing.	Use ARMA model to capture autocorrelation in the error term.

Moving to the frequency domain

	Time Domain	**Frequency Domain**
	x_t linear combination of past	x_t linear combination of periodic components
Object of interest	population ACF and PACF	Spectral Density
Data analysis tool	sample acf and pacf	Periodogram

Identify dominant freqenc(y/ies)

Where are we going?

Today: Spectral Density

motivation, examples

Then: Periodogram

the periodogram is an estimate of the spectral density

A quick trig review

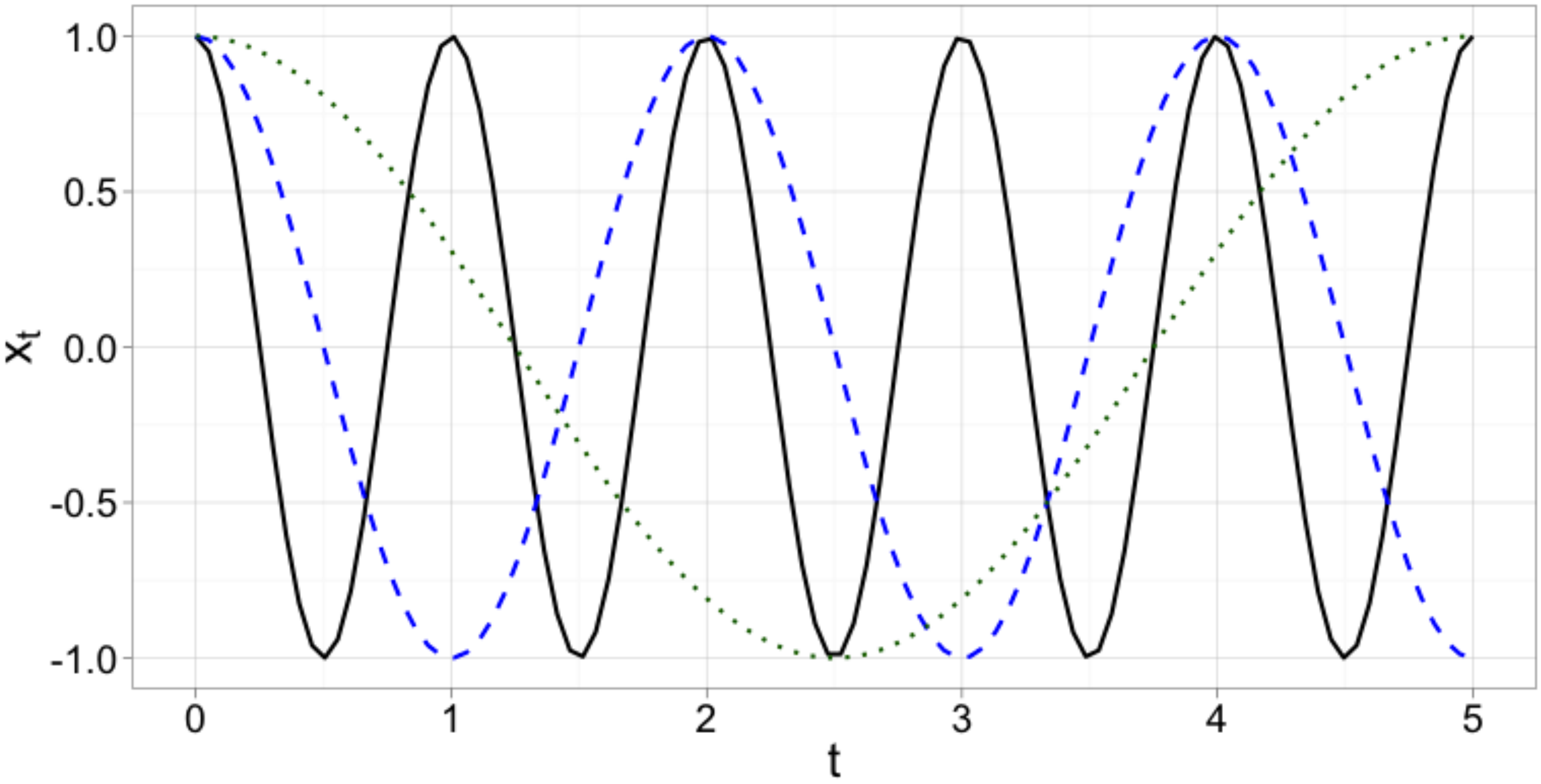
Imagine a series,

$$x_t = \underset{\overset{\uparrow}{\text{Amplitude}}}{R} \cos\left(\omega\, t + \overset{\overset{\text{Phase}}{\downarrow}}{\phi} \right)$$

$$\omega \,/\, (2\pi) = f \underset{\text{(cycles per unit time)}}{\text{Frequency}}$$

$$x_t = R \cos(t\,\omega + \boldsymbol{\phi})$$

Frequency, f



f = 1, cycle/unit time

f = 1/2, cycle/unit time

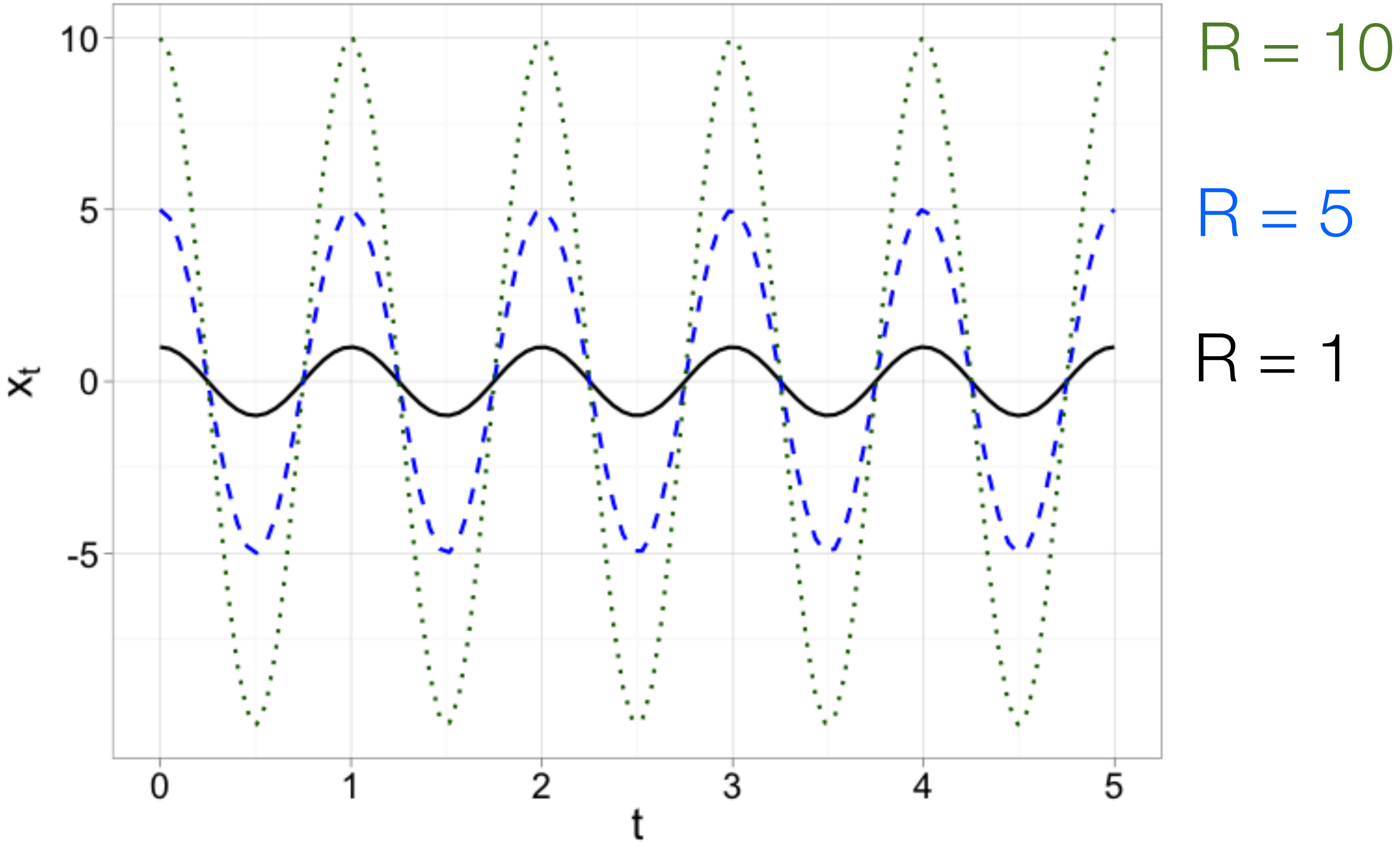
f = 1/5, cycle/unit time

Period = 1/f

time to complete one cycle

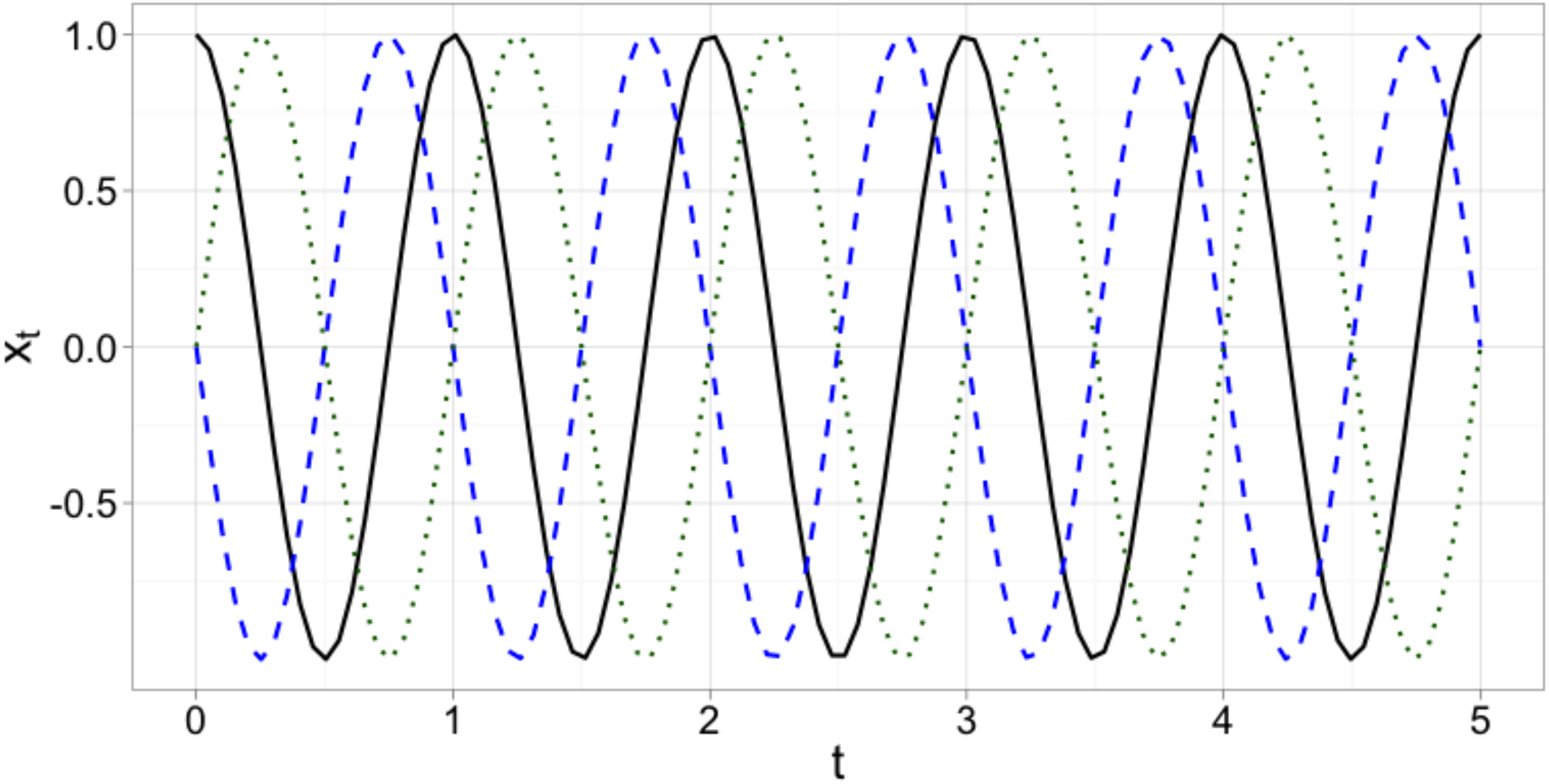
$$x_t = A \cos (2\pi\, t\, \omega + \boldsymbol{\phi})$$

Amplitude, R



$$x_t = A \cos(2\pi t \omega + \boldsymbol{\phi})$$

Phase, $\boldsymbol{\phi}$



$\phi = \pi/2$ $\phi = 0$ $\phi = -\pi/2$

$$x_t = R \cos(\omega t + \phi) + \underset{\substack{\uparrow \\ \text{white noise}}}{Z_t}$$

This isn’t stationary, (why?), but it is if we assume R is a zero mean random variable and ϕ is Uniform(0, 2π).

Generally we rewrite this as:

$$x_t = a \cos(\omega t) + b \sin(\omega t) + Z_t$$

where $a = R \cos(\phi)$, $b = - R \sin(\phi)$

and $a^2 + b^2 = R^2$

(using the identity: $\cos(u + v) = \cos(u)\cos(v) - \sin(u)\sin(v)$)

Extend to

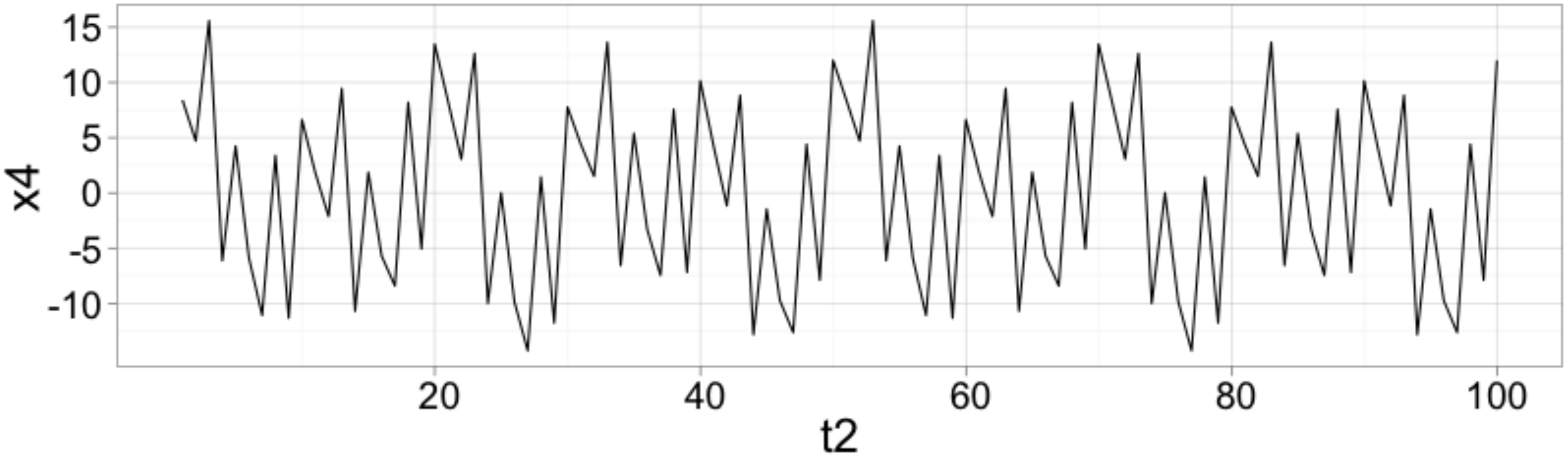
$$x_t = \sum_j a_j \cos(\omega_j t) + b_j \sin(\omega_j t) + Z_t \quad j = 1, ..., k$$

a sum of k periodic components

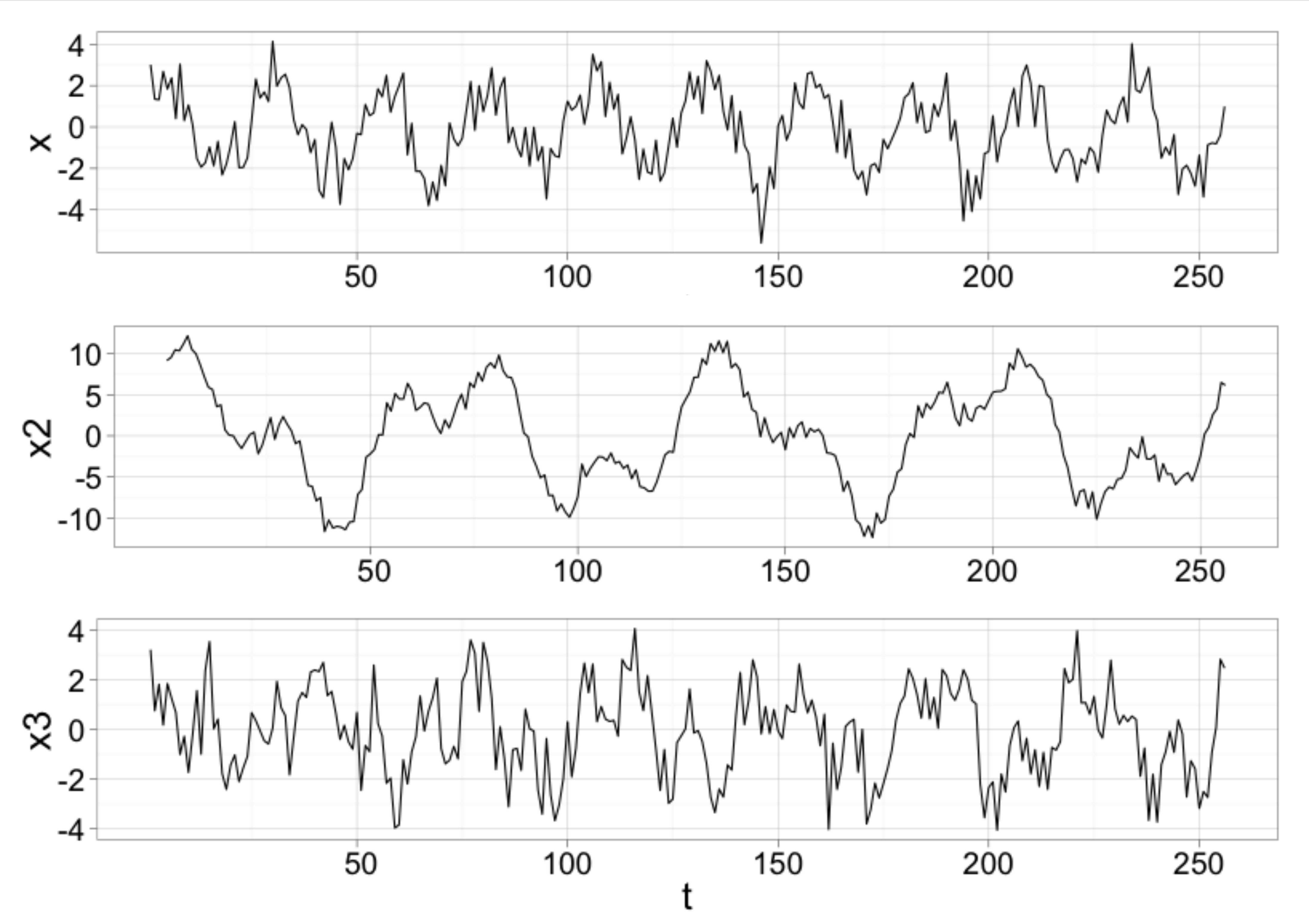
Deterministic example n=100

$$x_t = 2 \cos(2\pi t\ 6/n) + 3 \sin(2\pi\ 6/n) +$$
$$4 \cos(2\pi t\ 10/n) +5 \sin(2\pi t\ 10/n) +$$
$$6 \cos(2\pi t\ 40/n) + 7 \sin(2\pi t\ 40/n)$$

k=3



Can you guess the dominant frequencies in these (simulated) series?



Spectral Representation Theorem

Any stationary time series may be thought of as the random superposition of sines and cosines oscillating at various frequencies.

stochastic integrals

$$x_t = \int_0^\pi \cos \omega t \, du(\omega) + \int_0^\pi \sin \omega t \, dv(\omega)$$

where u(ω) and v(ω) are uncorrelated continuous processes with orthogonal increments.

Every frequency, ω, between 0 and π contributes to the variation in the process.

Spectral distribution function

For any real valued stationary series, there exists a monotonically increasing function, F, such that

$$\gamma(k) = \int_0^\pi \cos(\omega k)\, dF(\omega)$$

F(0) = 0

Riemann–Stieltjes integral

F is called the power spectrum distribution function.

F(ω) has a direct interpretation, it is the contribution to the variance from the frequencies in 0 to ω.

Riemann–Stieltjes integral

If F(x) is continuous, which it will be under certain conditions

$$\int_a^b g(x)dF(x) = \int_a^b g(x)f(x)dx, \quad \text{where } f(x) = F'(x)$$

$$\gamma(k) = \int_0^\pi \cos(\omega k)\, dF(\omega)$$

Show that the variance of x_t is F(π).

Spectral density function

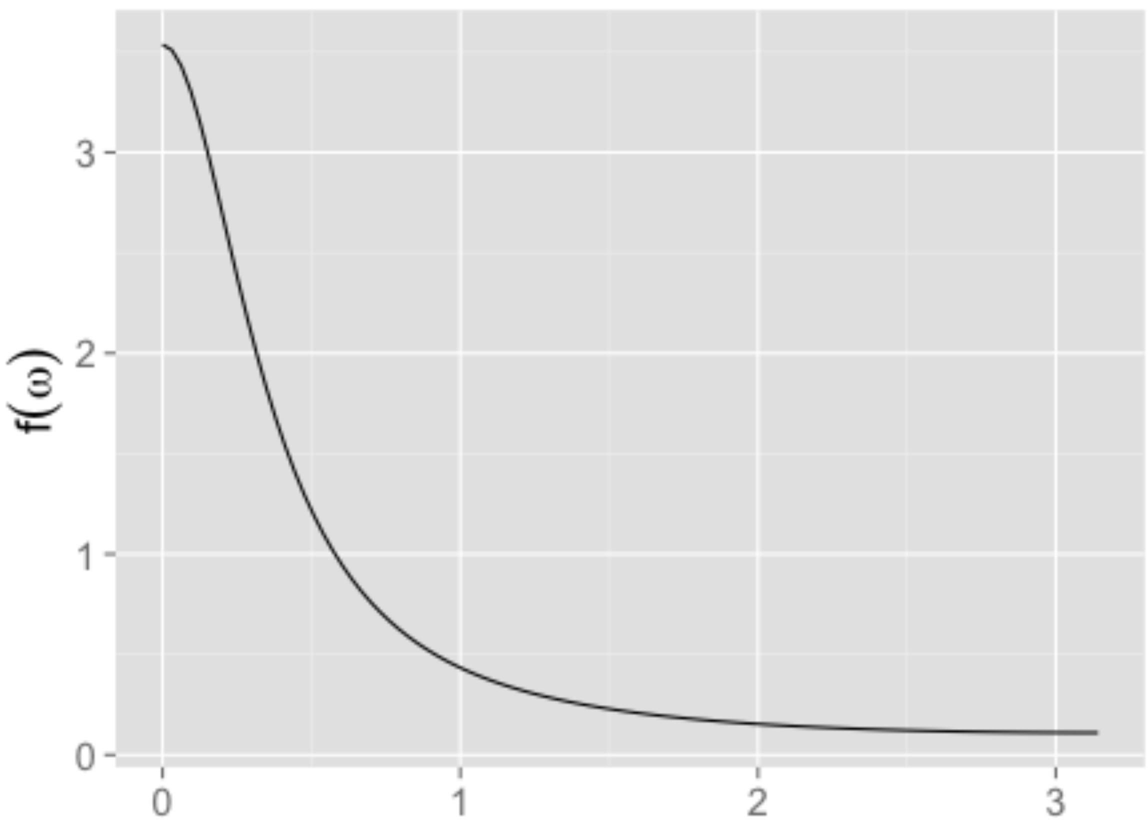
The derivative of the spectral distribution function is called the spectral density function, $f(\omega)$ (it exists for purely indeterministic discrete time stationary processes).

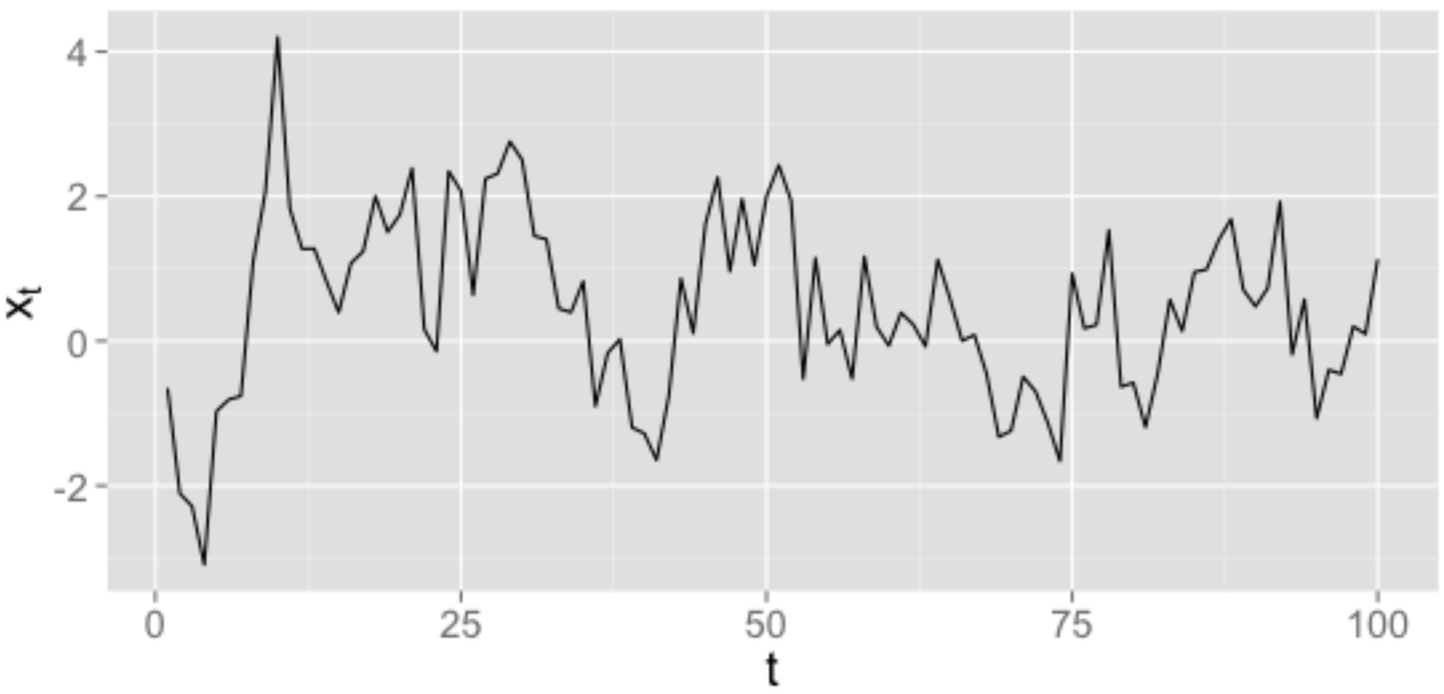
A.k.a. **the spectrum**

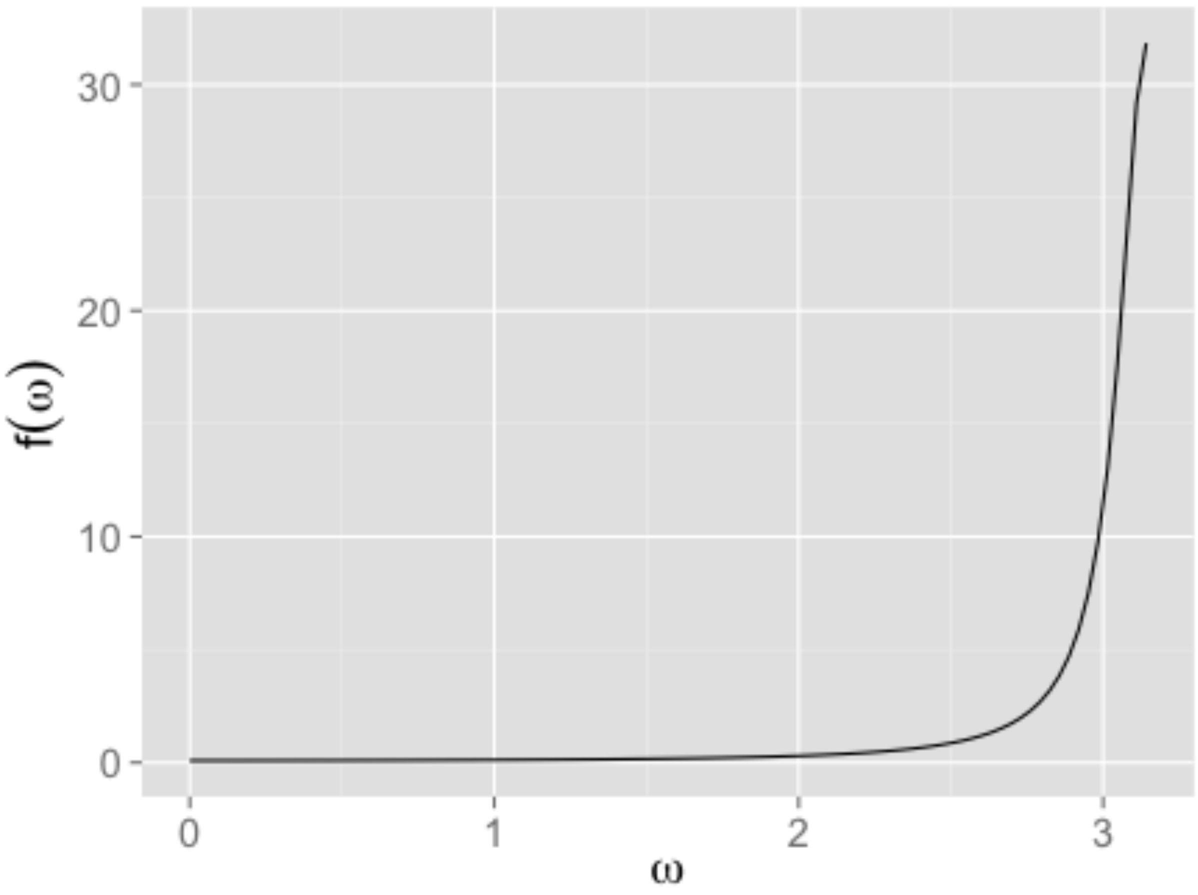
$f(\omega)\, d\omega$ is the contribution to the variance of the frequencies $(\omega, \omega + d\omega)$.

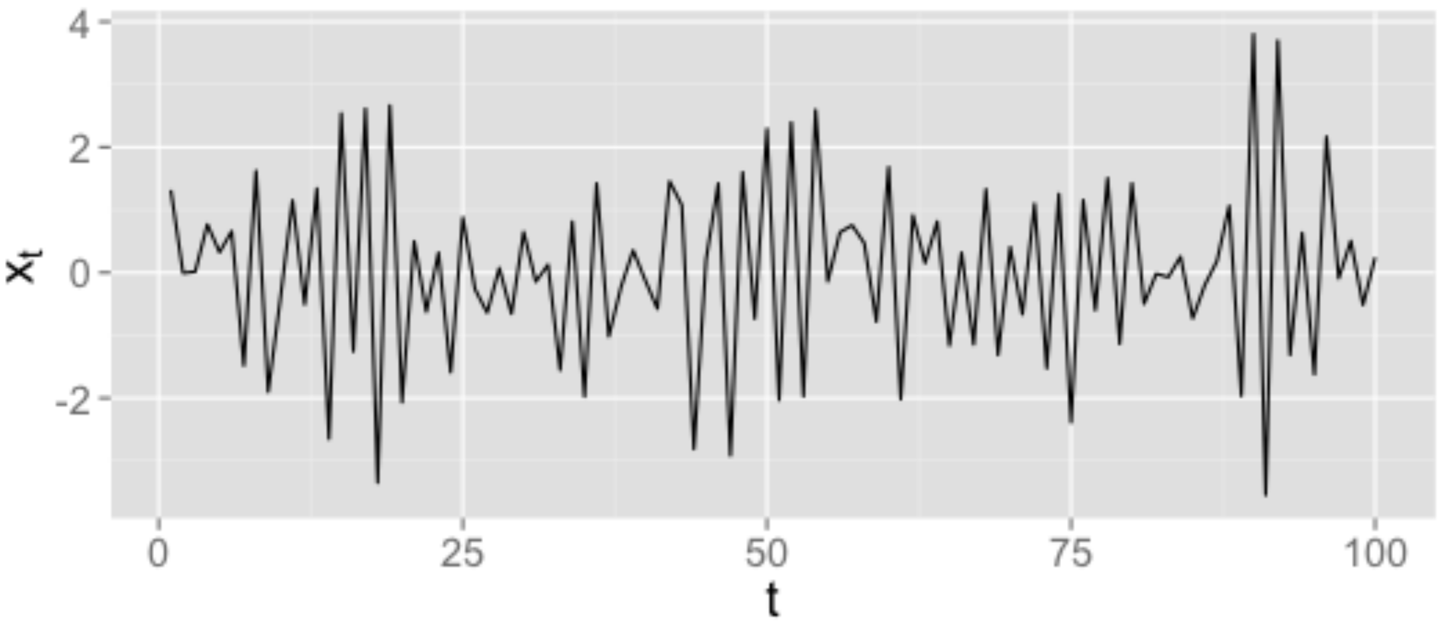
We draw $f(\omega)$ against ω and examine it.

Some examples









1.
Joel Gorder
Dane Skinner
Christopher Skypeck

2.
Jeremiah Cloud
Daniel Garmat
Merih Wahid

3.
Kyle Hirata
Kai Li
Senal Weerasooriya

4.
Daniel Claborne
Yijun Yang
Songqing Ye
Jonathon Valente

5.
Caley Johns
Johanna Doty
Casey Stevens
Song Hoa Choi

6.
Meng Mei
Trevor Ruiz
Paul Logan
Spencer Ledoux

7.
Yun Tang
Casey Bausell
Si Liu
Xiaoxi Gu

8.
Samuel Engle
Chenxiao Hu
Arpita Mukherjee

9.
Camden Lopez
Michael Dumelle
Tadesse Zemicheal
Amir Azarbakht

10.
Jeremiah Groom
Chuan Tian
Laura Gamble
Yiran Wang

11.
Shivani Patel
Karin Kralicek
Paulo Jose Murillo Sandoval
Peter Rise

12.
Matthew Higham
Mai Nguyen
Alyssa Pedersen

11am Project Time

Agenda

1. Introduce yourselves
2. Exchange emails and agree on how you will collaborate (dropbox/google drive/github latex/word etc.)
3. Discuss possible topics
4. Agree on your next meeting time
5. Agree (and repeat back to your team) your tasks before the next meeting time

#4 and #5 should happen every time you meet!

Projects

Proposal due in one week. One page.

I guarantee at least 15 mins of class time each lecture for the remaining weeks.

Final report due last day of class.

Deliverable is negotiable.