Introduction to stochastic processes

Mark Scheuerell

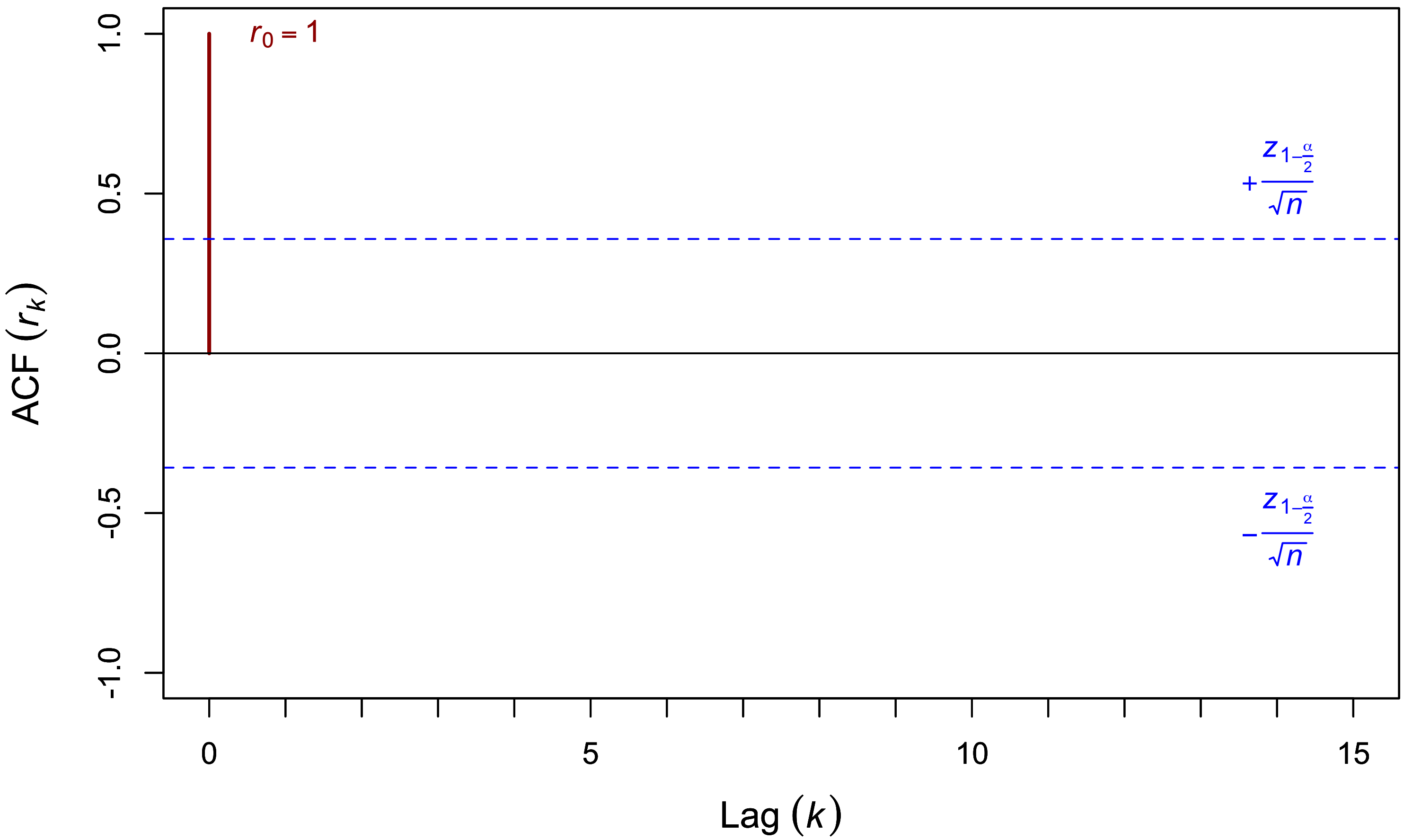
FISH 507 – Applied Time Series Analysis

10 January 2017

Topics for today

- Quick review of
  - Correlograms
  - White noise
  - Random walks
- Linear stationary models
- Autoregressive (AR)
- Moving average (MA)
- Autoregressive moving average (ARMA)
- Using ACF & PACF for model ID

The correlogram



White noise (WN)

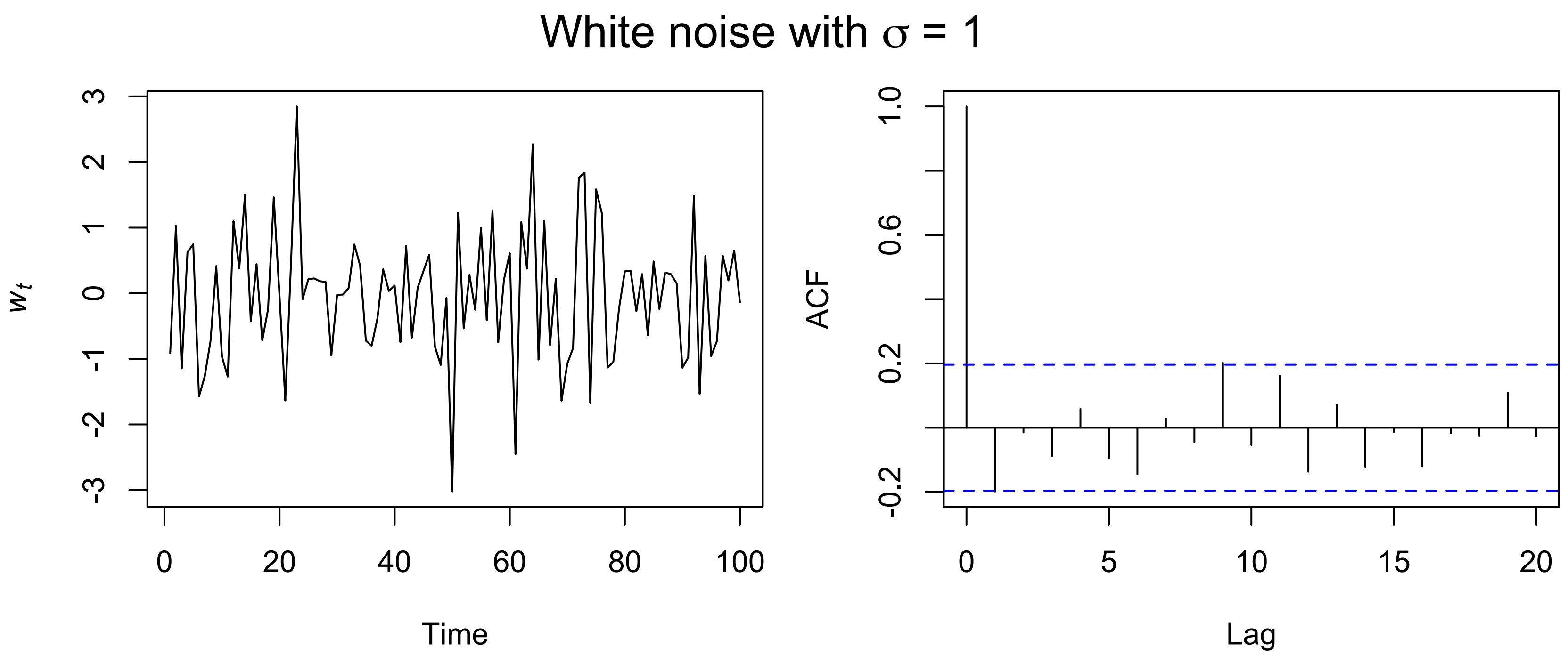
A time series $\{w_t : t = 1,2,3,\ldots,n\}$ is *discrete white noise* if the variables $w_1, w_2, w_3, \ldots, w_n$ are

1) *independent*, and
2) *identically distributed* with a mean of zero

Gaussian WN has the following 2nd-order properties:

$$\mu_w = 0 \qquad \gamma_k = \begin{cases} \sigma^2 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases} \qquad \rho_k = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$$

White noise



White noise with σ = 1
w_t
3
2
1
0
-1
-2
-3
0
20
40
60
80
100
Time
ACF
1.0
0.6
0.2
-0.2
0
5
10
15
20
Lag

Random walk (RW)

A time series $\{x_t : t = 1,2,3,\ldots,n\}$ is a *random walk* if

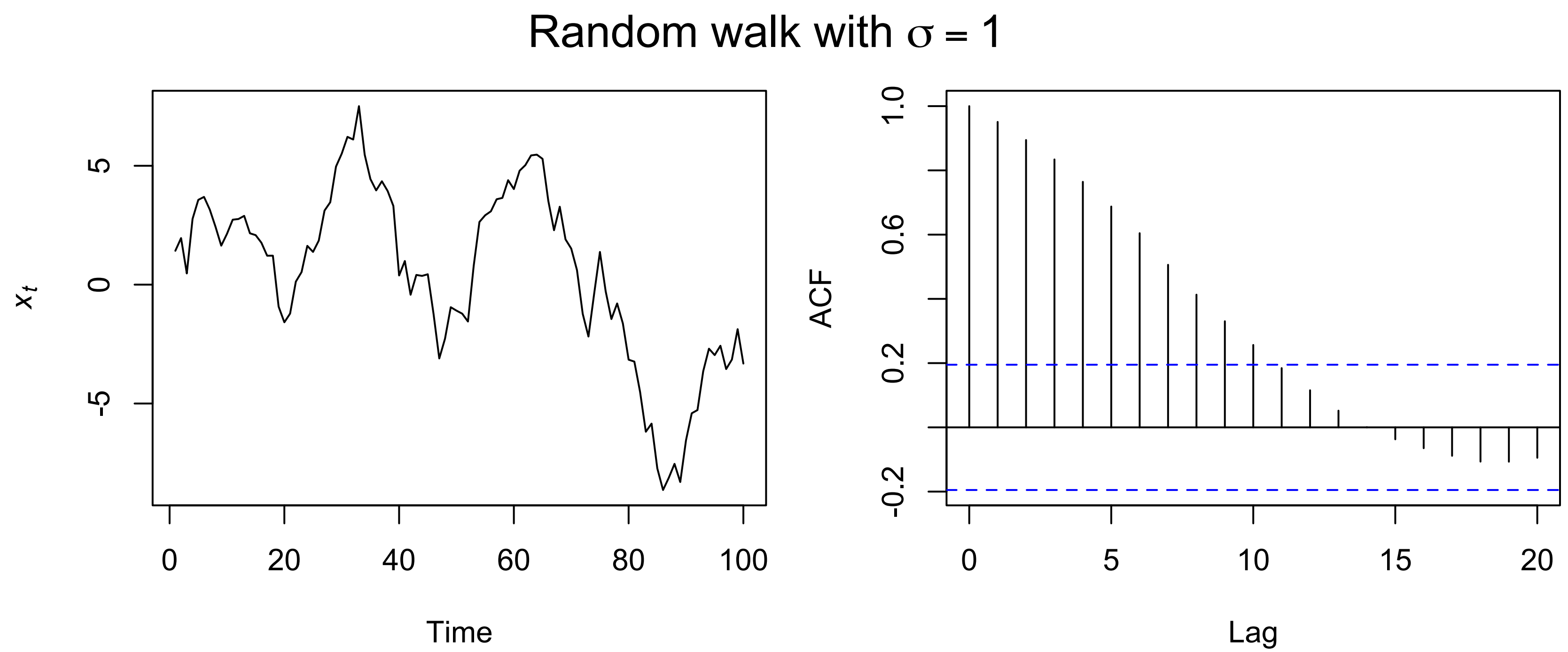
1) $x_t = x_{t-1} + w_t$, and
2) w_t is white noise

RW has the following 2nd-order properties:

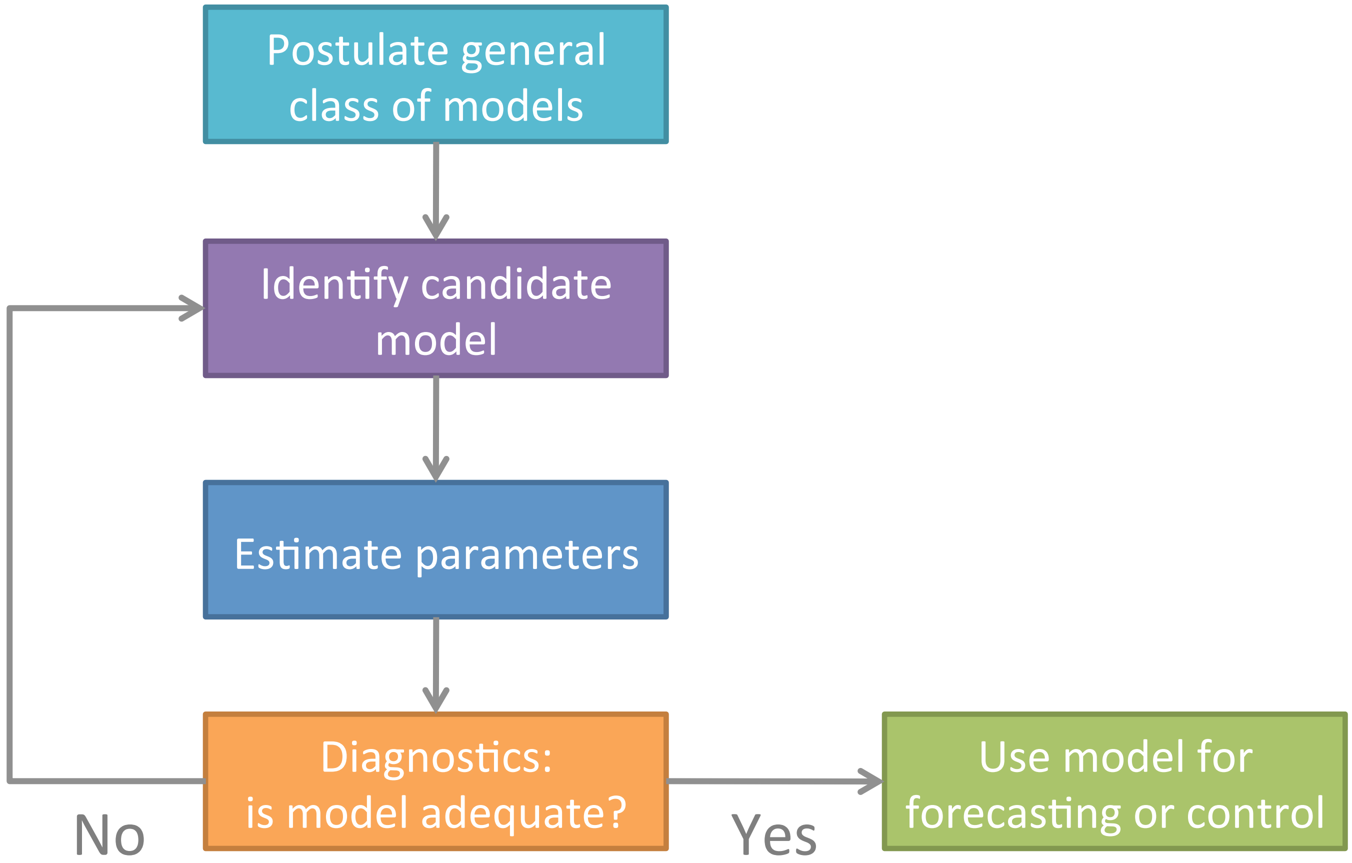
$$\mu_w = 0 \qquad \gamma_k(t) = t\sigma^2 \qquad \rho_k(t) = \frac{t\sigma^2}{\sqrt{t\sigma^2 (t+k)\sigma^2}} = \frac{1}{\sqrt{1 + k/t}}$$

Random walks are NOT stationary!

Random walk (RW)



Iterative approach to model building



Linear stationary models

- We saw last week that linear filters are a useful way of modeling time series
- Here we extend those ideas to a general class of models call autoregressive moving average (ARMA)

Autoregressive (AR) models

- An *autoregressive* model of order *p*, or AR(*p*), is defined as

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t$$

where we assume

1) w_t is WN, and

2) $\phi_p \neq 0$ for order-*p* process

- *Note*: RW model is special case of AR(1) with $\phi_1 = 1$

Stationary & nonstationary AR models

- We can write out an AR(*p*) model using the backward shift notation, such that

$$\phi_p(\mathbf{B})x_t = \left(1 - \phi_1\mathbf{B} - \phi_2\mathbf{B}^2 - \ldots - \phi_p\mathbf{B}^p\right)x_t = w_t$$

- If we treat **B** as a number, we can out write the *characteristic equation* as

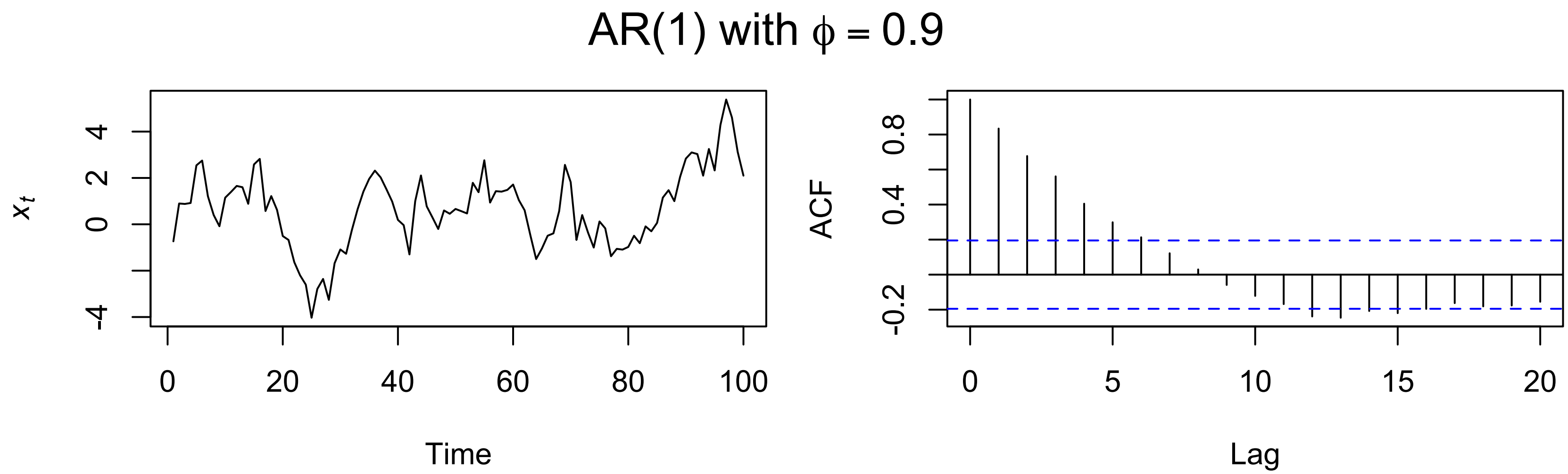
$$\phi_p(\mathbf{B}) = 0$$

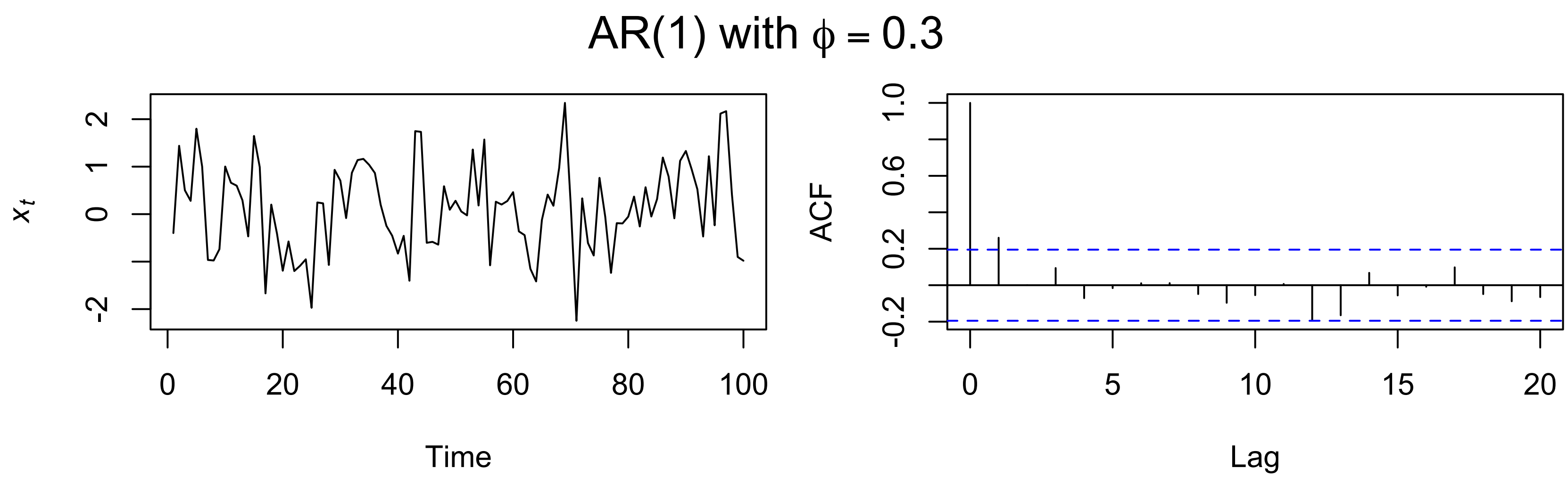
- In order to be stationary, *all roots of char eqn must exceed 1 in absolute value*

Stationary & nonstationary AR models

- For example, a RW model is not stationary because $\phi = 1 - \mathbf{B}$, and hence, $\mathbf{B} = 1$
- However, the AR(1) model $x_t = 0.5x_{t-1} + w_t$ is because $\phi = 1 - 0.5\mathbf{B}$, and hence, $\mathbf{B} = 2 > 1$

Examples of AR(1) processes





Partial autocorrelation function

- The partial *autocorrelation function* (PACF) measures the linear correlation of a series x_t and x_{t+k} with the linear dependence of $\{x_{t-1}, x_{t-2}, \ldots, x_{t-(k-1)}\}$ removed
- It is defined as

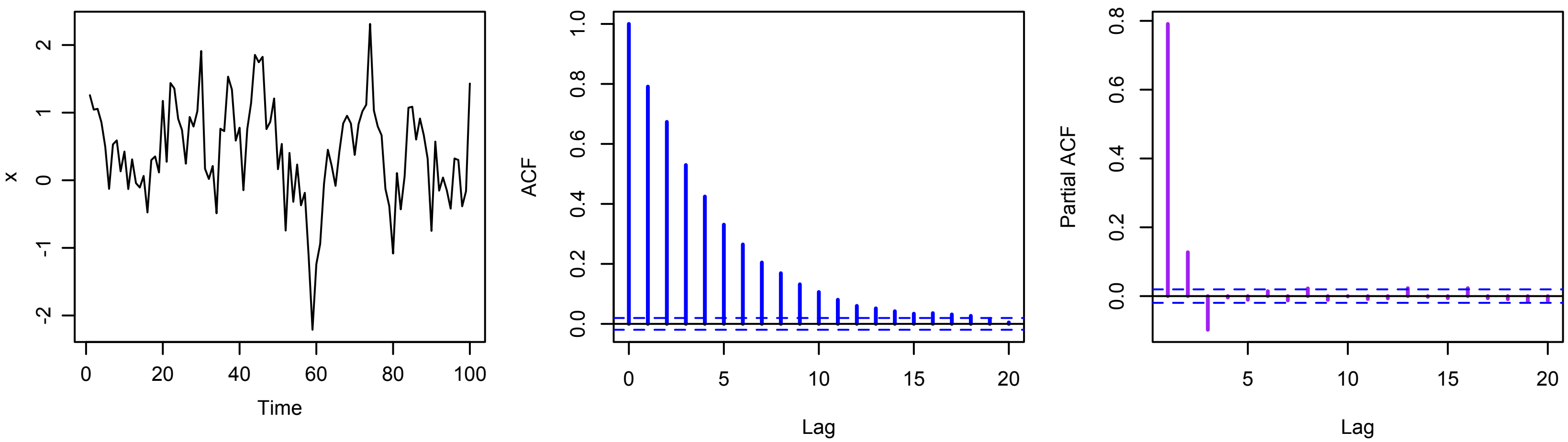
$$\phi_{kk} = \begin{cases} \mathrm{Cor}\left(x_1, x_0\right) = \rho\left(1\right) & \text{if } k = 1 \\ \mathrm{Cor}\left(x_k - x_k^{k-1}, x_0 - x_0^{k-1}\right) & \text{if } k \geq 2 \end{cases}$$

$$x_k^{k-1} = \beta_1 x_{k-1} + \beta_2 x_{k-2} + \cdots + \beta_{k-1} x_1$$

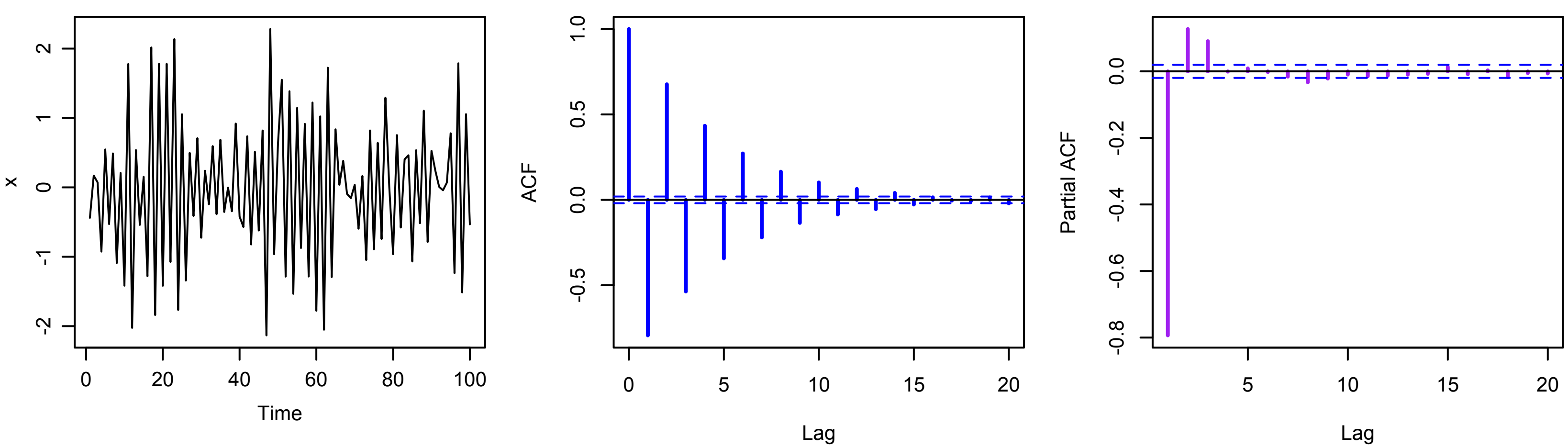
$$x_0^{k-1} = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k-1} x_{k-1}$$

ACF & PACF for AR(3) processes

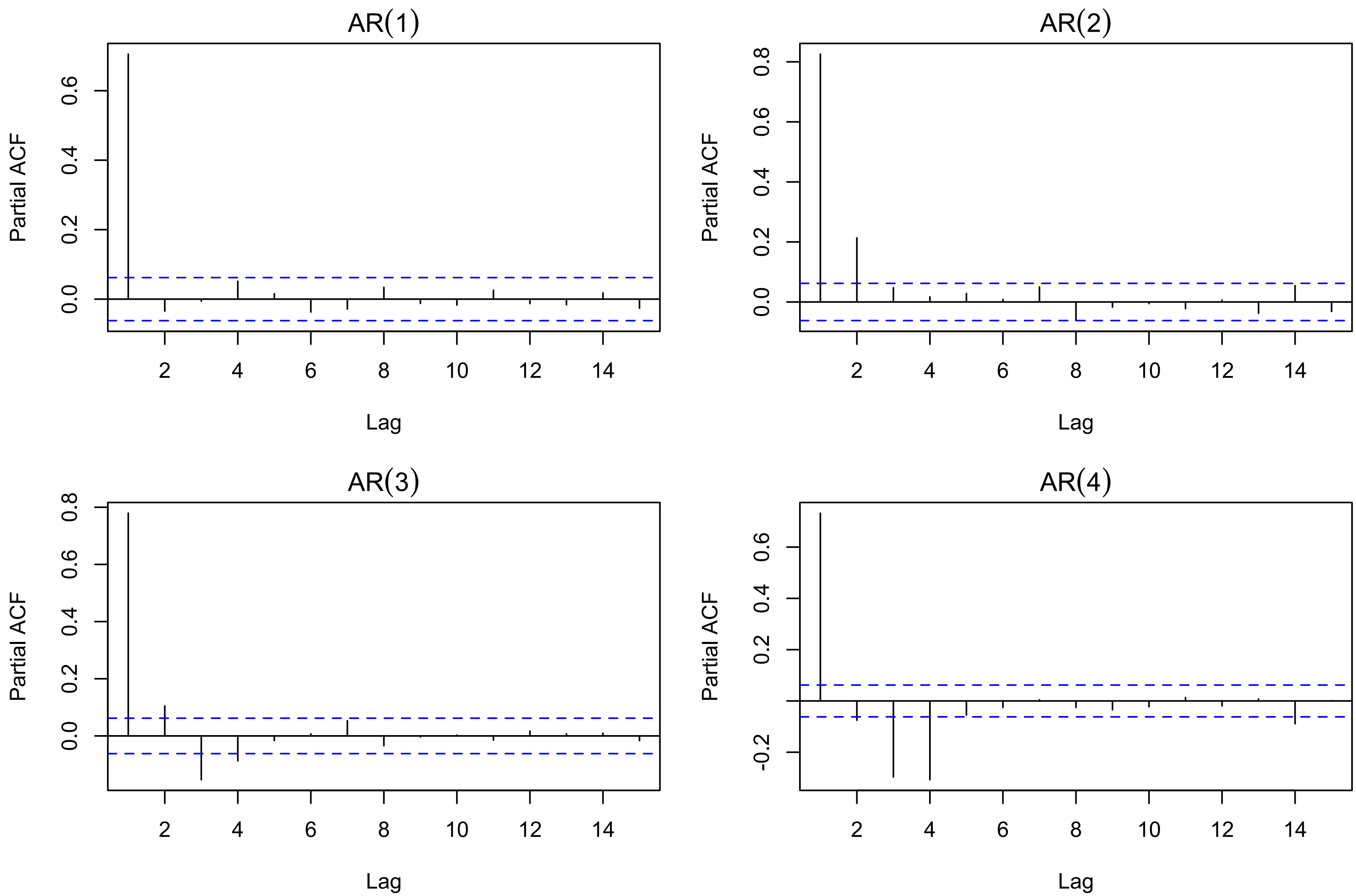
AR(3) with $\phi_1 = 0.7$, $\phi_2 = 0.2$, $\phi_3 = -0.1$



AR(3) with $\phi_1 = -0.7$, $\phi_2 = 0.2$, $\phi_3 = -0.1$



PACF for AR(*p*) processes



Using ACF & PACF for model ID

	ACF	PACF
AR(p)	Tails off slowly	Cuts off after lag-p

Moving average (MA) models

- A *moving average* model of order q, or MA(q), is defined as

$$x_t = w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}$$

  where w_t is WN (with 0 mean)

- It is simply the current error term plus a weighted sum of the q most recent error terms

- Because MA processes are finite sums of stationary WN processes, they are themselves stationary

Invertible MA models

- We can write out an MA(q) model using the backward shift notation, such that

$$x_t = \left(1 + \theta_1 \mathbf{B} + \theta \mathbf{B}^2 + \ldots + \theta_q \mathbf{B}^q\right) w_t = \theta_q\left(\mathbf{B}\right) w_t$$

- An MA process is *invertible* if it can be expressed as a stationary autoregressive process of infinite order without an error term

- For example, an MA(1) process with $\theta < |1|$

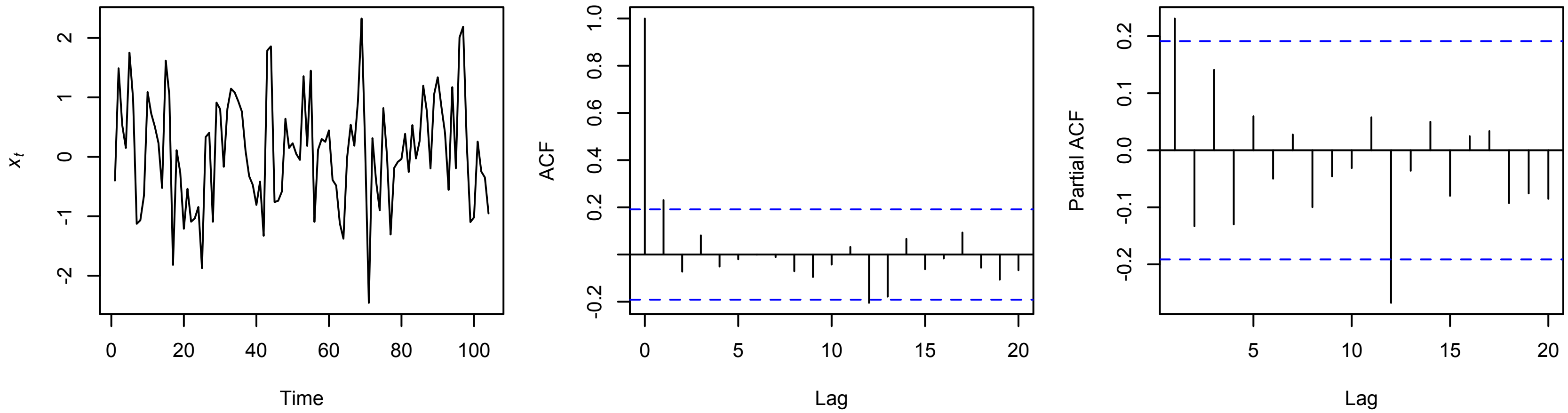
$$x_t = \left(1 - \theta \mathbf{B}\right) w_t$$

$$w_t = \left(1 - \theta \mathbf{B}\right)^{-1} x_t$$

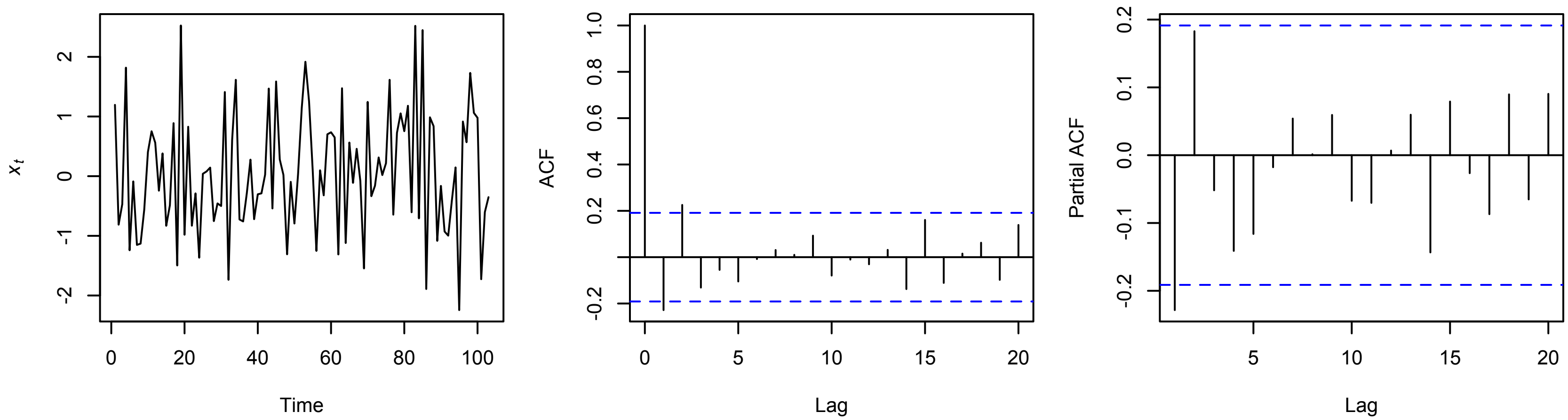
$$w_t = \left(1 + \theta \mathbf{B} + \theta^2 \mathbf{B}^2 + \ldots\right) x_t = x_t + \theta x_{t-1} + \theta^2 x_{t-2} + \ldots$$

Examples of MA(q) processes

MA(1) with $\theta = 0.3$



MA(2) with $\theta_1 = -0.3$, $\theta_2 = 0.3$



Using ACF & PACF for model ID

	ACF	PACF
AR(p)	Tails off slowly	Cuts off after lag-p
MA(q)	Cuts off after lag-q	Tails off slowly

Autoregressive moving average models

- A time series is *autoregressive moving average*, or ARMA(p,q), if it is stationary and

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q}$$

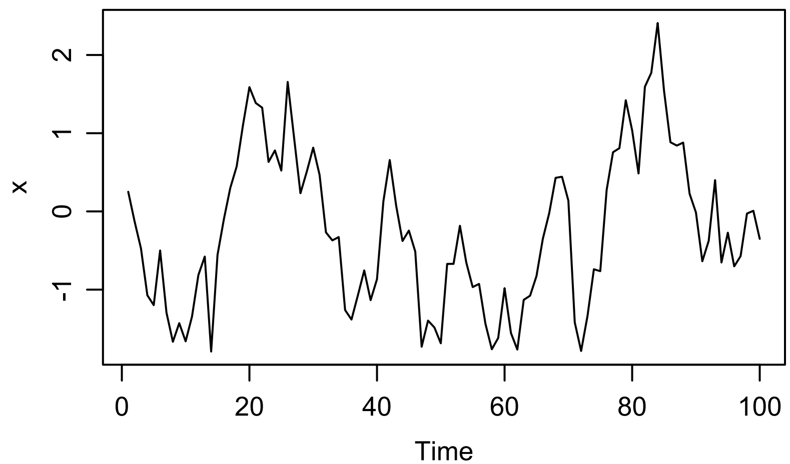
- We can write out an ARMA(p,q) model using the backward shift notation, such that

$$\phi_p(\mathbf{B}) x_t = \theta_q(\mathbf{B}) w_t$$

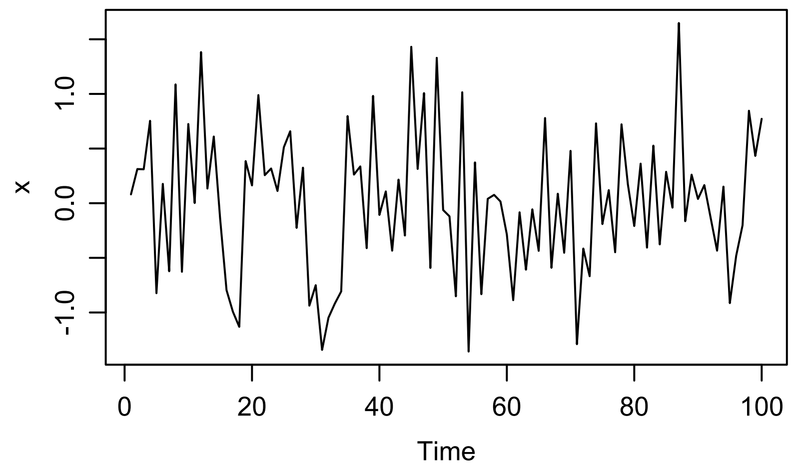
- ARMA models are *stationary* if all roots of $\phi(\mathbf{B}) > 1$
- ARMA models are *invertible* if all roots of $\theta(\mathbf{B}) > 1$

Examples of ARMA(p,q) processes

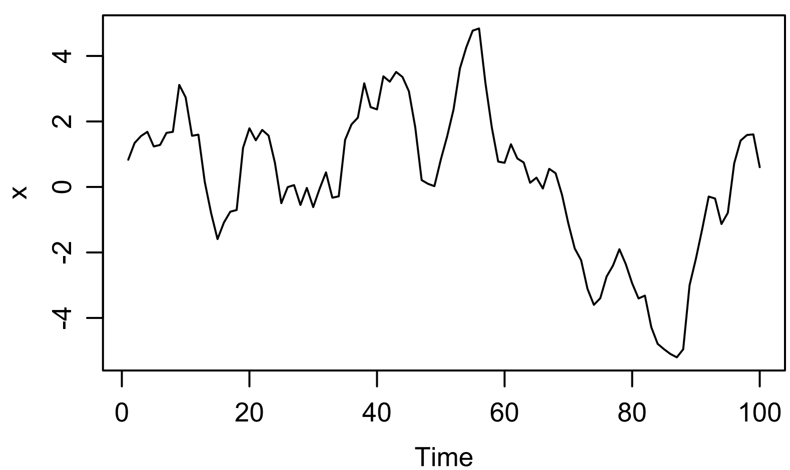
ARMA(3,1): $\phi_1 = 0.7$, $\phi_2 = 0.2$, $\phi_3 = -0.1$, $\theta_1 = 0.5$



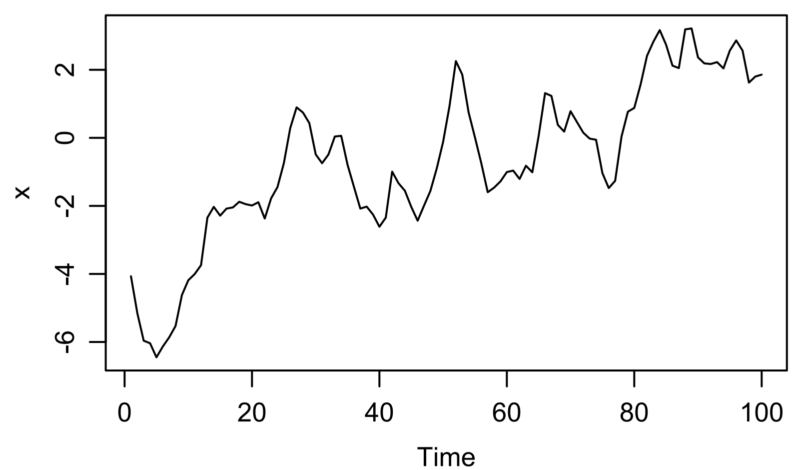
ARMA(2,2): $\phi_1 = -0.7$, $\phi_2 = 0.2$, $\theta_1 = 0.7$, $\theta_2 = 0.2$



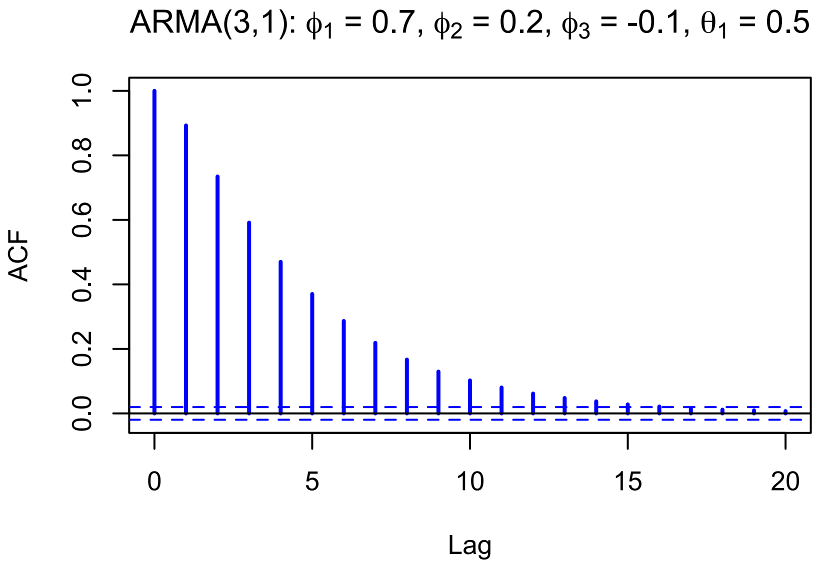
ARMA(1,3): $\phi_1 = 0.7$, $\theta_1 = 0.7$, $\theta_2 = 0.2$, $\theta_3 = 0.5$

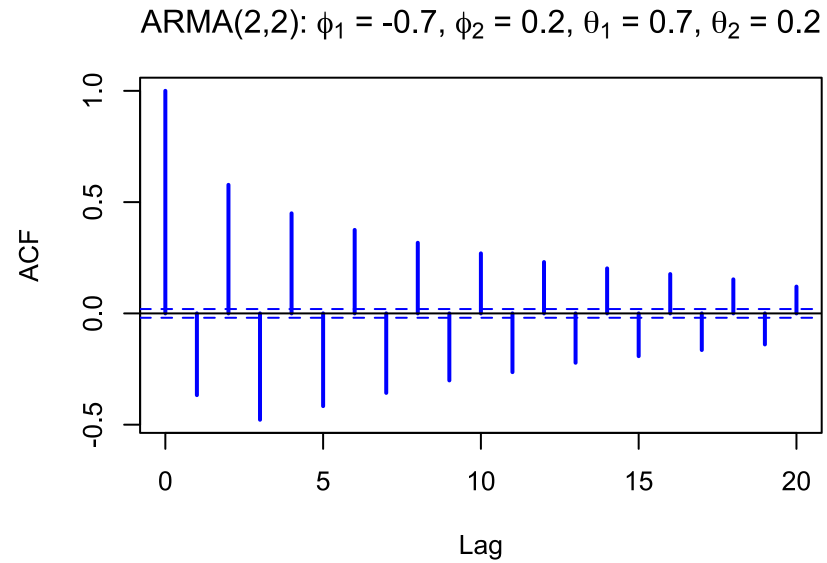


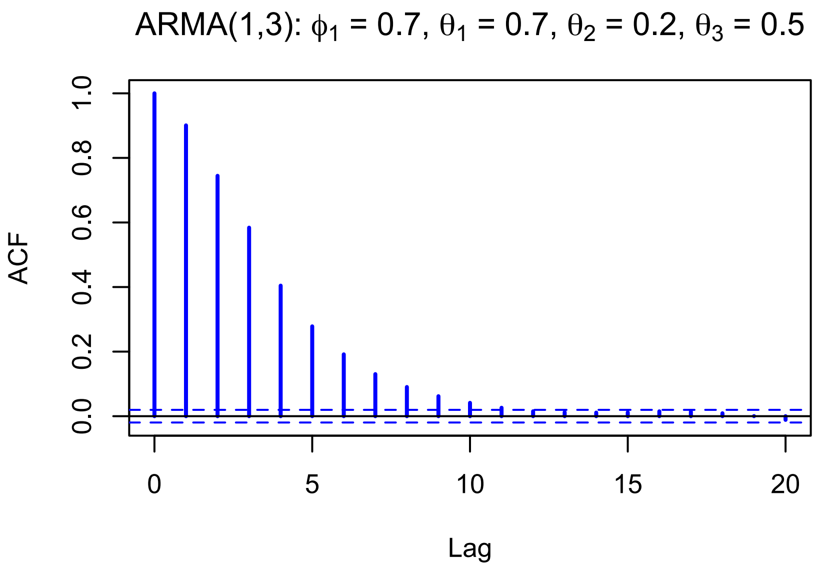
ARMA(2,2): $\phi_1 = 0.7$, $\phi_2 = 0.2$, $\theta_1 = 0.7$, $\theta_2 = 0.2$

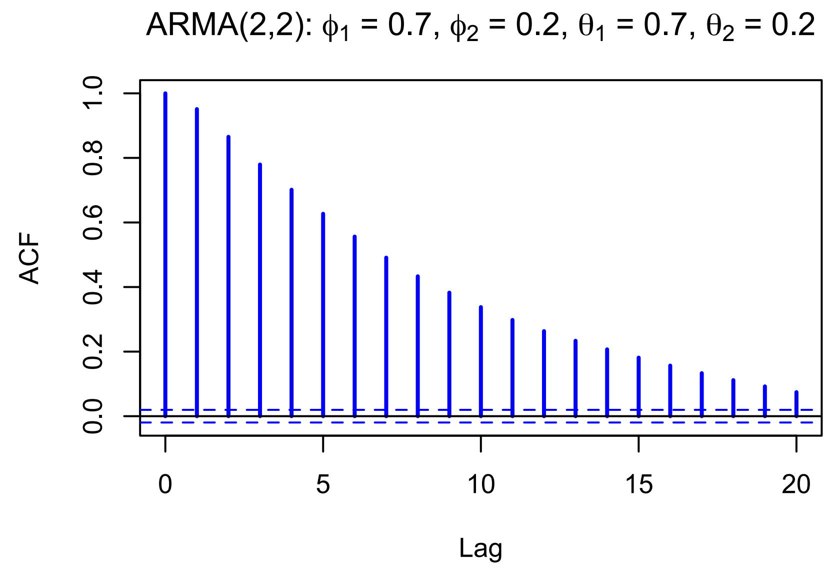


ACF for ARMA(*p*,*q*) processes

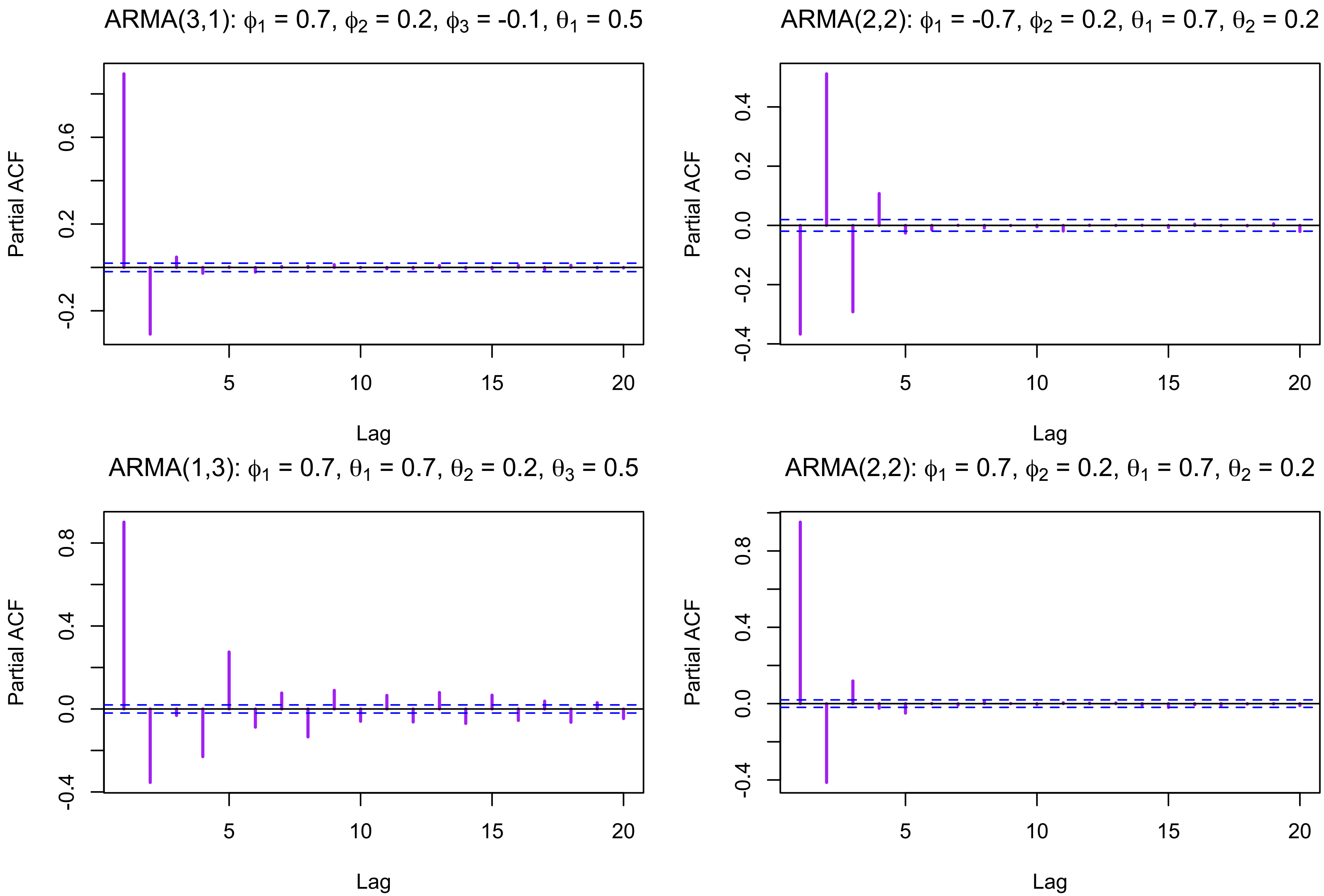








PACF for ARMA(*p*,*q*) processes



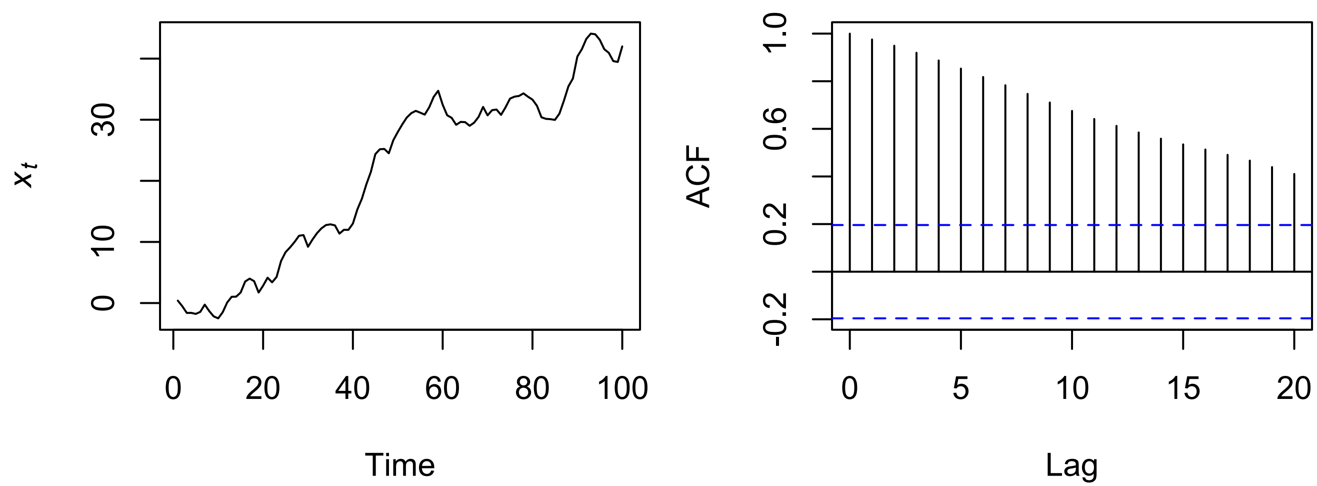
Using ACF & PACF for model ID

	ACF	PACF
AR(*p*)	Tails off slowly	Cuts off after lag-*p*
MA(*q*)	Cuts off after lag-*q*	Tails off slowly
ARMA(*p*,*q*)	Tails off (after lag [*q*-*p*])	Tails off (after lag [*p*-*q*])

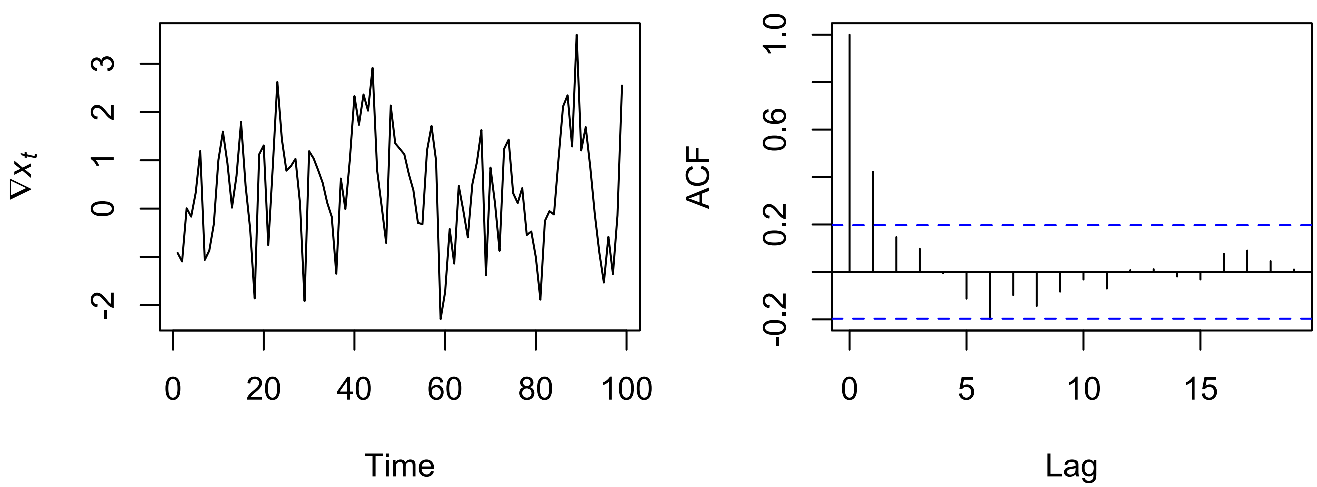
Nonstationary time series models

- If the data appear stationary, we can try various forms of ARMA models
- If not, differencing can often make them stationary
- This leads to the class of *autoregressive integrated moving average* (ARIMA) models
- ARIMA models are indexed with orders (p,d,q); the d indicates the order of differencing
- For $d > 0$, $\{x_t\}$ is an ARIMA(p,d,q) process if $(1-\mathbf{B})^d x_t$ is a causal ARMA(p,q) process

Example of an ARIMA model

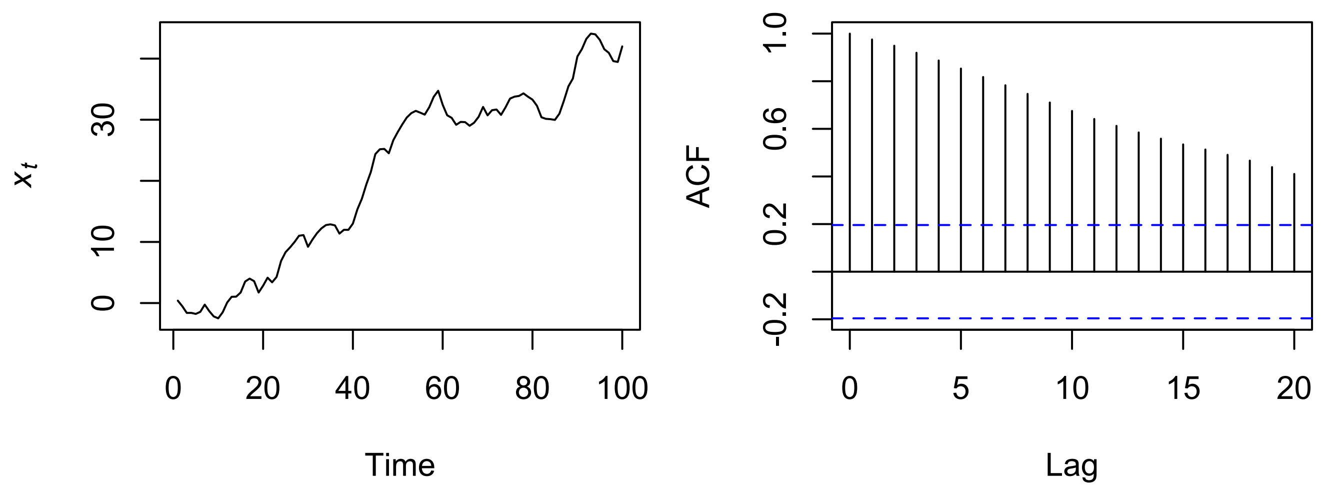


Data do not appear stationary!

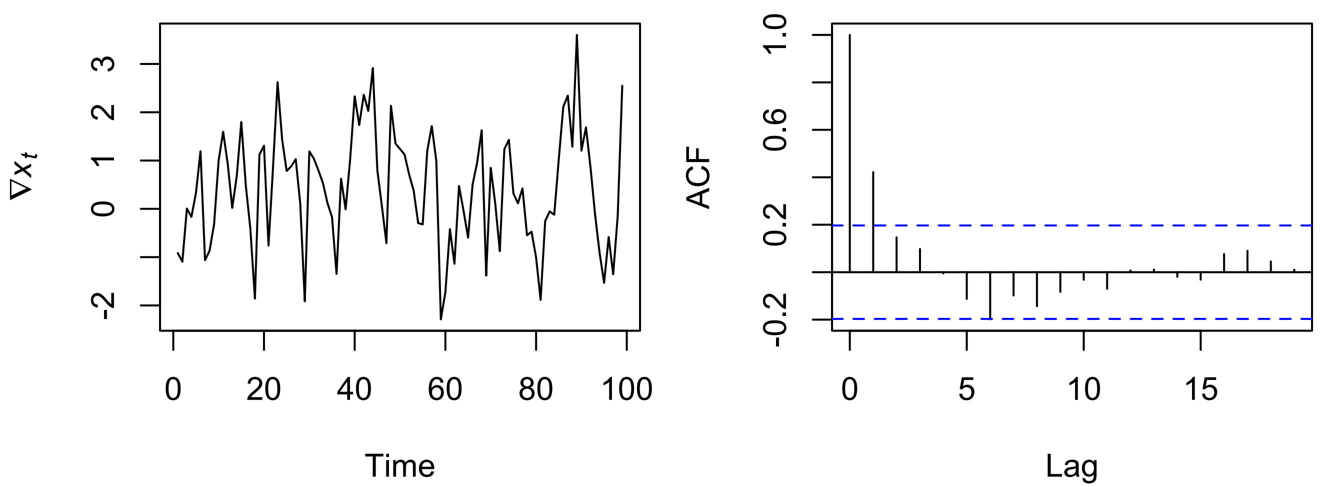


Differenced data look much better

Example of an ARIMA model



x_t is ARIMA(1,1,0)



∇x_t is AR(1)

Topics for today

- Quick review of
  - Correlograms
  - White noise
  - Random walks
- Linear stationary models
- Autoregressive (AR)
- Moving average (MA)
- Autoregressive moving average (ARMA)
- Using ACF & PACF for model ID