TIME-DOMAIN ASTRONOMY

Lectures 5: Time Domain Analysis

Stefano Covino
INAF / Brera Astronomical Observatory

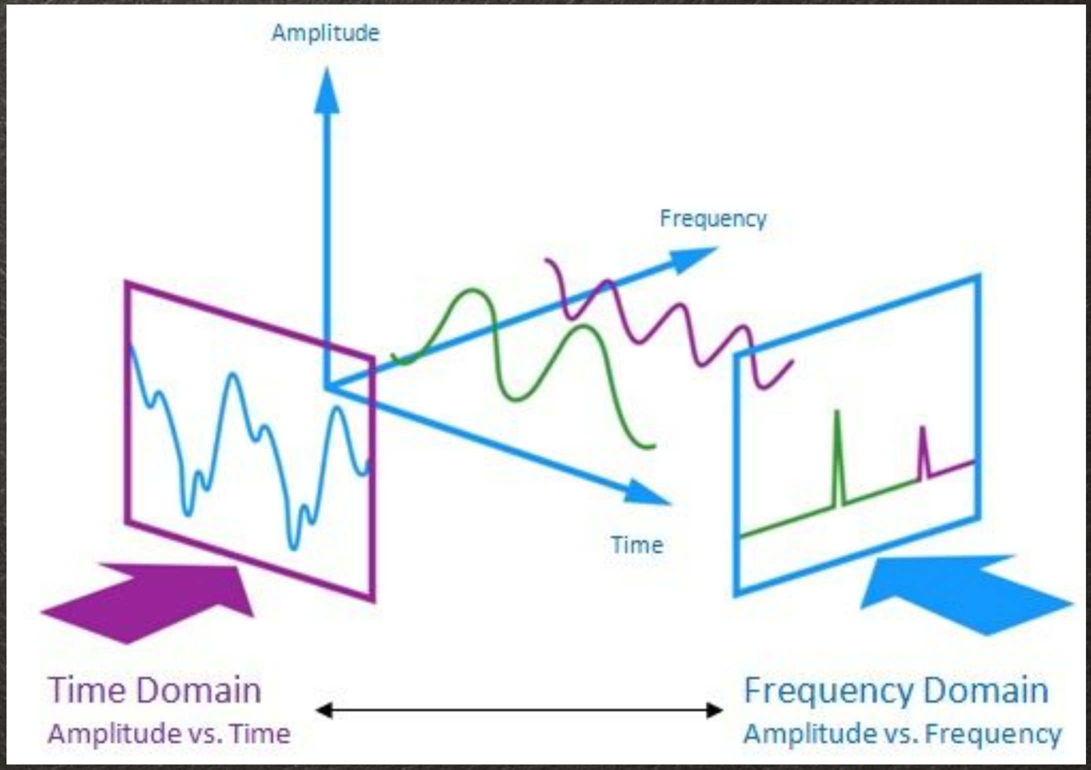
Time Domain Analysis

- One can derive inferences about the spectral content of a time series by a spectral analysis.

- An analogous (i.e. complementary) analysis can be carried out in the time domain.

- Traditionally, analyses in the temporal domain are often aimed at deriving predictions, while in the spectral domain one often looks for periodicity, etc.

- There are indeed no fundamental technical reasons for that.



Time Domain Analysis

	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)

The Correlation Function

- One of the main statistical tools for the analysis of stochastic variability is the autocorrelation function. It represents a specialized case of the correlation function of two functions, f(t) and g(t), scaled by their standard deviations, and defined at time lag as:

$$\mathrm{CF}(\Delta t) = \frac{\lim_{T\to\infty} \frac{1}{T} \int_{(T)} f(t)\, g(t+\Delta t)\, dt}{\sigma_f\, \sigma_g},$$

- where σ_f and σ_g are standard deviations of f(t) and g(t), respectively. With this normalization, the correlation function is unity for $\Delta t = 0$ (without normalization by standard deviation, the above expression is equal to the covariance function)

The Auto-Correlation Function

- It is assumed that both f and g are statistically weakly stationary functions.
- With f(t)=g(t)=y(t), the autocorrelation of y(t) defined at time lag is:

$$\mathrm{ACF}(\Delta t) = \frac{\lim_{T\to\infty} \frac{1}{T} \int_{(T)} y(t)\, y(t+\Delta t)\, dt}{\sigma_y^2}.$$

- The autocorrelation function and the PSD of function $y(t)$ are Fourier pairs; this fact is known as the Wiener–Khinchin theorem and applies to stationary random processes.

The Auto-Correlation Function

- The sample auto-correlation function is defined as:

$$ACF(k) = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n} (X_t - \bar{X})^2}$$

- where the numerator is just the sample auto-covariance function and the denominator the sample variance.

- ACF(k) is often called a correlogram.

The Auto-Correlation Function

- The ACF is fundamental measure of the serial correlation in a time series.
- It is defined for evenly spaced time-series, yet there are generalization to irregularly sampled data.
- Under the null hypothesis that the time series has no correlated structure and the population ACF is zero for all lags [except for ACF(0) which is always unity], the distribution of the sample is asymptotically normal with mean $-1/n$ and variance $1/n$;
  - that is, the distribution of the null case ACF is $ACF(k) = N(-1/n, 1/n)$.

The Auto-Correlation Function

- Simple quantities like the sample mean are valid estimates of the population mean for stationary processes, but its uncertainty is not the standard value when autocorrelation is present:

$$\widehat{Var}(\bar{X}_n) = \frac{\sigma^2}{n}\left[1 + 2\sum_{k=1}^{n-1}(1 - k/n)ACF(k)\right].$$

- Qualitatively, this can be understood as due to the decrease of independent measurements.

The partial ACF (PACF)

- The PACF at lag k gives the autocorrelation at value k removing the effects of correlations at shorter lags.
- The value of the p-th coefficient $PACF(p)$ is found by successively fitting autoregressive models with order 1, 2, . . . , p and setting the last coefficient of each model to the PACF parameter (*this will be clearer later*).

- For instance, for a stationary AR(2) process:

$$PACF(2) = \frac{ACF(2) - ACF(1)^2}{1 - ACF(1)^2}.$$

The partial ACF (PACF)

- The theoretical partial autocorrelation function of a stationary time series can be calculated by using the Durbin–Levinson Algorithm:

$$\phi_{n,n} = \frac{\rho(n) - \sum_{k=1}^{n-1} \phi_{n-1,k}\rho(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k}\rho(k)}$$

where $\phi_{n,k} = \phi_{n-1,k} - \phi_{n,n}\phi_{n-1,n-k}$ for $1 \leq k \leq n-1$ and $\rho(n)$ is the autocorrelation function.

Discrete Correlation Function

- The discrete correlation function is a procedure for computing the autocorrelation function that avoids interpolating an unevenly spaced dataset onto a regular grid.

- The method can treat both autocorrelation within one time series or cross-correlation between two unevenly spaced time series.

- Consider two datasets (x_i, t_{xi}) and (z_j, t_{zj}) with $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,m$ points, respectively. For autocorrelation within a single dataset, let $z = x$.

Discrete Correlation Function

- Construct two matrices, the **unbinned discrete correlation function (UDCF)** and its associated time lags:

$$UDCF_{ij} = \frac{(x_i - \bar{x})(z_j - \bar{z})}{\sigma_x \sigma_z}$$
$$\Delta t_{ij} = t_j - t_i$$

- The UDCF is then grouped into a univariate function of lag-time τ by collecting the M (τ) data pairs with lags falling within the interval $\tau - \Delta\tau/2 \leq \Delta t_{ij} < \tau + \Delta\tau/2$.
- The resulting discrete correlation function (DCF) and its variance are:

$$DCF(\tau) = \frac{1}{M(\tau)} \sum_{k=1}^{M(\tau)} UDCF_{ij}$$
$$Var(\tau) = \frac{1}{(M(\tau)-1)^2} \sum_{k=1}^{M(\tau)} [UDCF_{ij} - DCF(\tau)]^2.$$

Z-transformed DCF

- If bins have an equal number of points (i.e. bins can have different length), The bin distribution becomes approximately binomial and if a Fisher z-transform is applied:

$$z(\tau) = \frac{1}{2}\ln\left(\frac{1 + DCF(\tau)}{1 - DCF(\tau)}\right).$$

- The $z(\tau)$ values are now normally distributed with known mean and variance.

The Structure Function

- The structure function (sometimes called the Kolmogorov structure function) is again a measure of autocorrelation.

- The *q*-th order structure function is:

$$D^q(\tau) = \langle |x(t) - x(t + \tau)|^q \rangle$$

  - where the angular brackets $\langle \ \rangle$ indicate an average over the time series and q is the order of the structure function (not necessarily integer).

- The $q = 2$ structure function is also called the variogram.

The Structure Function

- When a dataset has a characteristic time-scale of variation τ_c, the structure function is small at τ shorter than τ_c, rises rapidly to a high level around τ_c, and stays at this plateau for longer τ.

- If a structure function exhibits a power-law dependence on lag, $D_q(\tau) \propto \tau^{\alpha}$, the time series has a multi-fractal behavior. In this case, the Wiener–Khinchin theorem links the power-law dependence of the $q = 2$ structure function to the power-law dependence of the Fourier power spectrum.

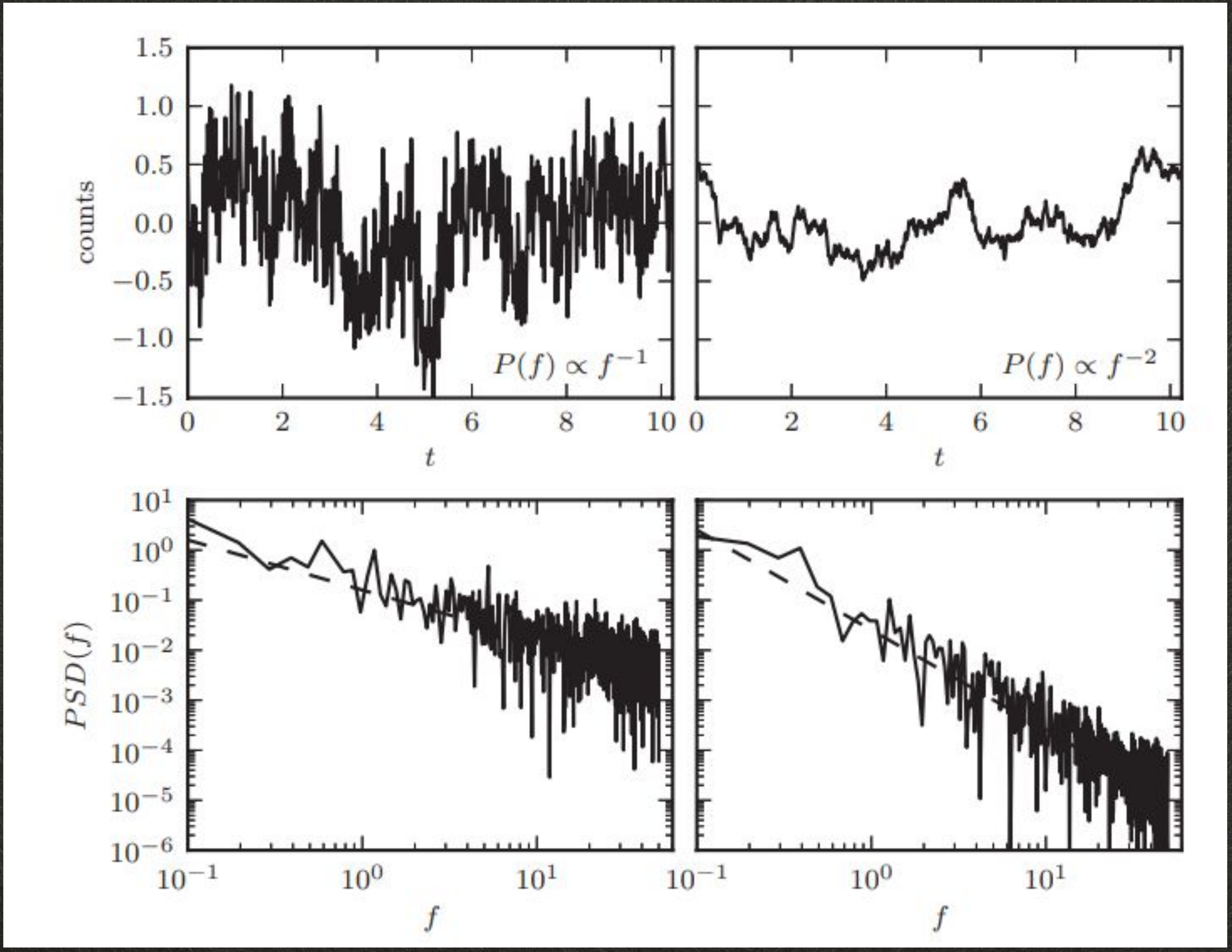
The Structure Function

- The structure function with q = 2 can also be seen as:

$$\mathrm{SF}(\Delta t) = \mathrm{SF}_\infty \left[1 - \mathrm{ACF}(\Delta t)\right]^{1/2},$$

- where SF_∞ is the standard deviation of the time series evaluated over an infinitely large time interval (or at least much longer than any characteristic timescale τ).
- The structure function with q=2 is equal to the standard deviation of the distribution of the difference of $y(t_2)$-$y(t_1)$ evaluated at many different t_1 and t_2.
- As mentioned before, when the structure function SF $\propto t^\alpha$, then PSD $\propto 1/f^{1+2\alpha}$.

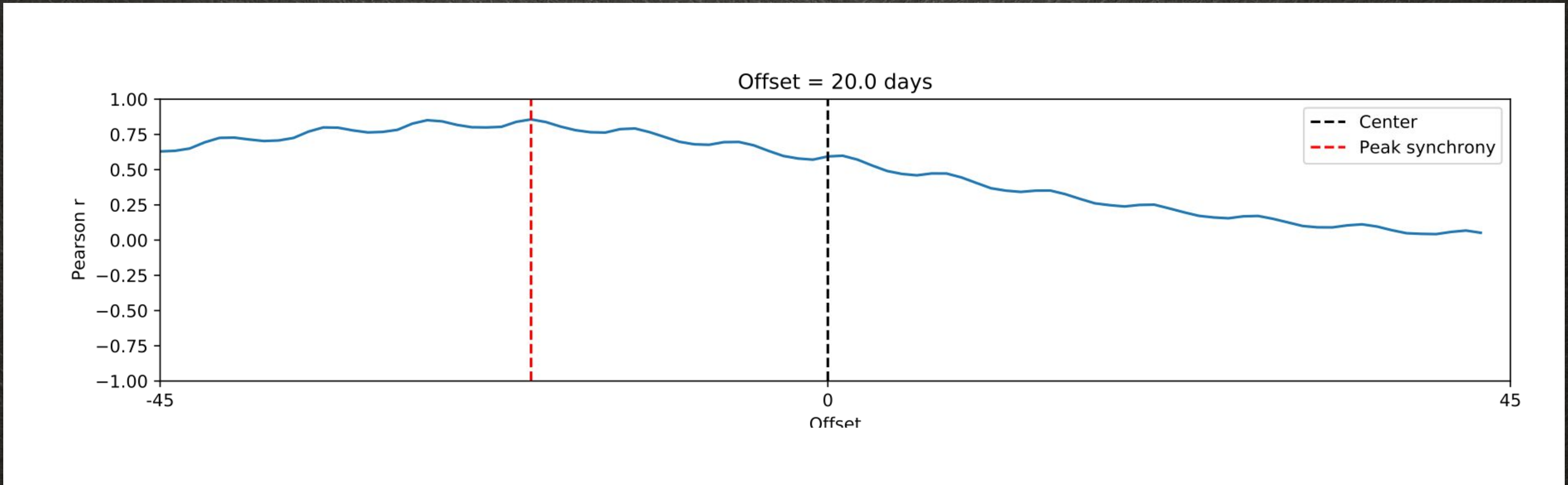
Long memory processes



Exercise

- Useful notebook:

1. Covid19CrossCorrelation



Stationarity

- This is a fundamental concept. A stationary time-series is a dataset where it is "meaningful" to draw predictions.

- A stochastic process X_t; $t = 0, \pm 1, \ldots$ is stationary if a joint distribution of $X_t, \cdots, X_{t+k}$ is same as a joint distribution of $X_0, \cdots, X_k$ for all t and all k.

    - This is typically called strict stationarity.

Stationarity

- A stochastic process X_t; $t = 0, \pm 1, \ldots$ is weakly (or second order) stationary if, for all t:

$$E(Xt) \equiv \mu$$

- and $\mathrm{Cov}(X_t, X_{t+h}) = \mathrm{Cov}(X_0, X_h) = C_X(h)$ is a function of h only (it does not depend on t). C_X is known as autocovariance of X. $C_X(h)/C_x(0)$ is known as autocorrelation.

- Let's recall that:

$$\mathrm{Cov}(X, Y) \equiv E[(X - E[X])(Y - E[Y])]$$

Stationarity

- It can be shown that:
  - If X_t has finite variance, strictly stationary implies X_t is weakly stationary.
  - If X_t is second order stationary and Gaussian it implies X_t is also strictly stationary.
- X_t is Gaussian if for each $t_1,\ldots,t_k$ the vector $(X_{t1}, \ldots, X_{tk})^T$ has a Multivariate Normal Distribution.
- The process X_t is said to have a stationary covariance if:
  - $Cov(X_t, X_s) = Cov(X_{t+1}, X_{s+1}) = Cov(X_{t+2}, X_{s+2}) \cdots$

Covariance Matrix

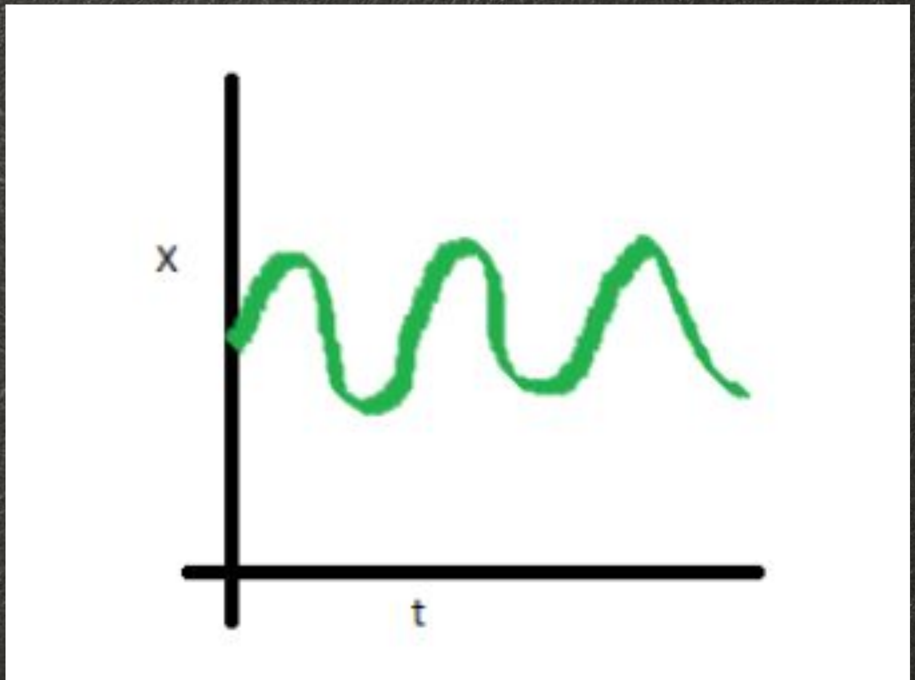
- One can create the covariance matrix Σ.
- The element of Σ, $S_{i,j} = C(i-j)$
- This kind of matrix is also a "Toeplitz" matrix (it has useful properties for massive computations)

$$A = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & \cdots & a_{-n+1} \\ a_1 & a_0 & a_{-1} & \ddots & & \vdots \\ a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & a_{-1} & a_{-2} \\ \vdots & & \ddots & a_1 & a_0 & a_{-1} \\ a_{n-1} & \cdots & \cdots & a_2 & a_1 & a_0 \end{bmatrix}$$

Exercise

- Useful notebook:

1. Non-stationary Time-Series



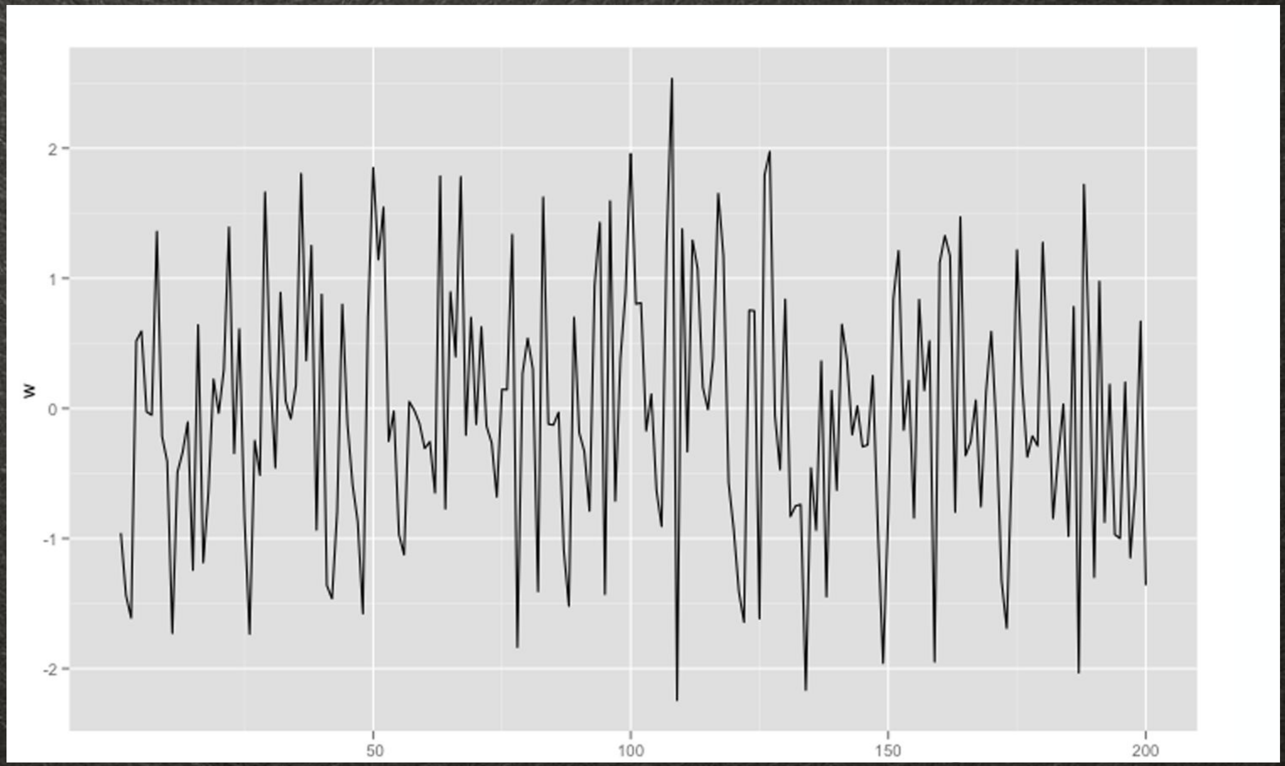
White Noise

- $\{w_t\}$ is a white noise process if w_t are uncorrelated and identically distributed random variables with:
  - $E[w_t] = 0$ and $Var[w_t] = \sigma^2$, for all t.

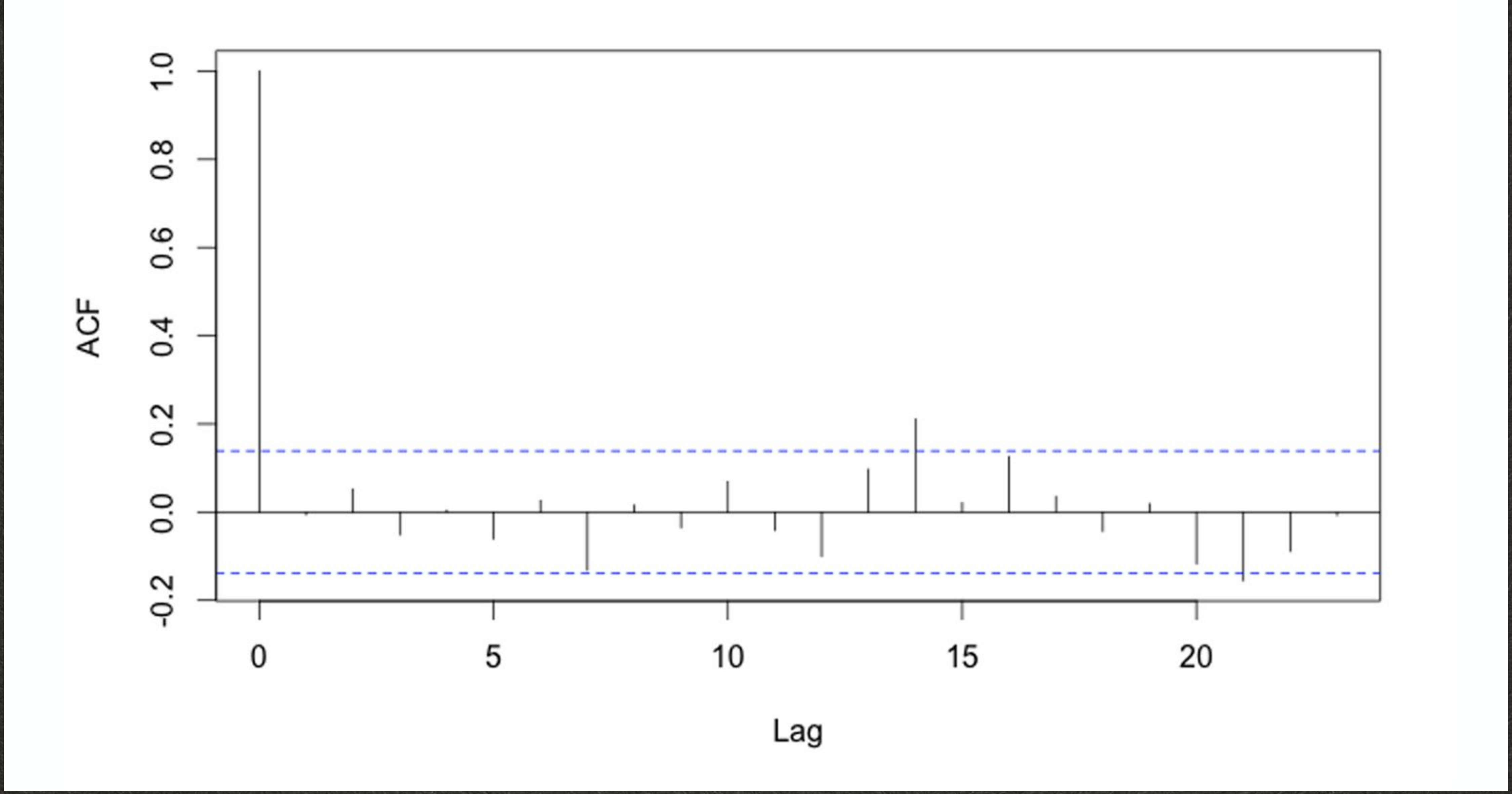
- If $\{w_t\}$ are Normally (Gaussian) distributed, we call this Gaussian white noise.

- A white noise process is stationary.

- Example with $\sigma^2 = 1$



White Noise ACF



Random walk with drift

$$x_t = \delta + x_{t-1} + w_t$$

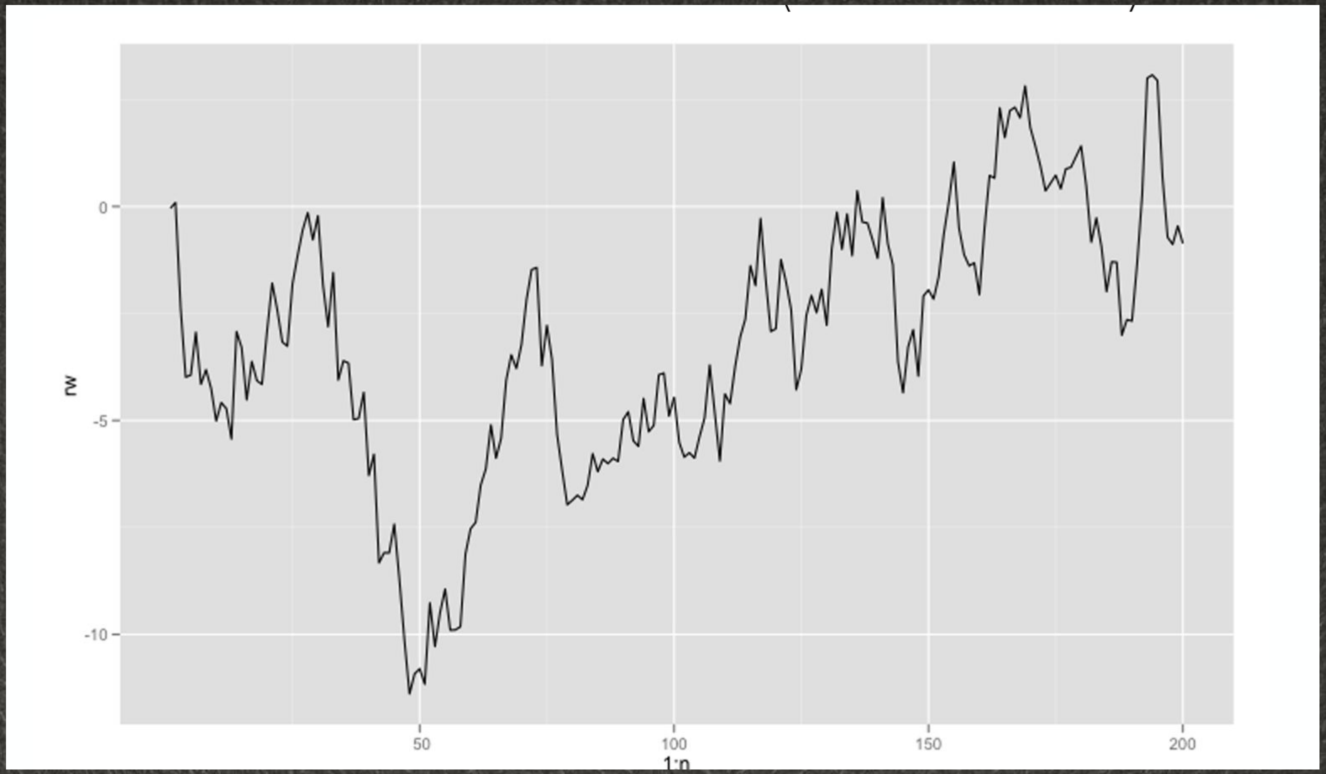
- δ is a constant, w_t is a white noise process, $x_0 = 0$.
- The equation can be rewritten as:

$$x_t = t\delta + \sum_{j=1}^{t} w_j$$

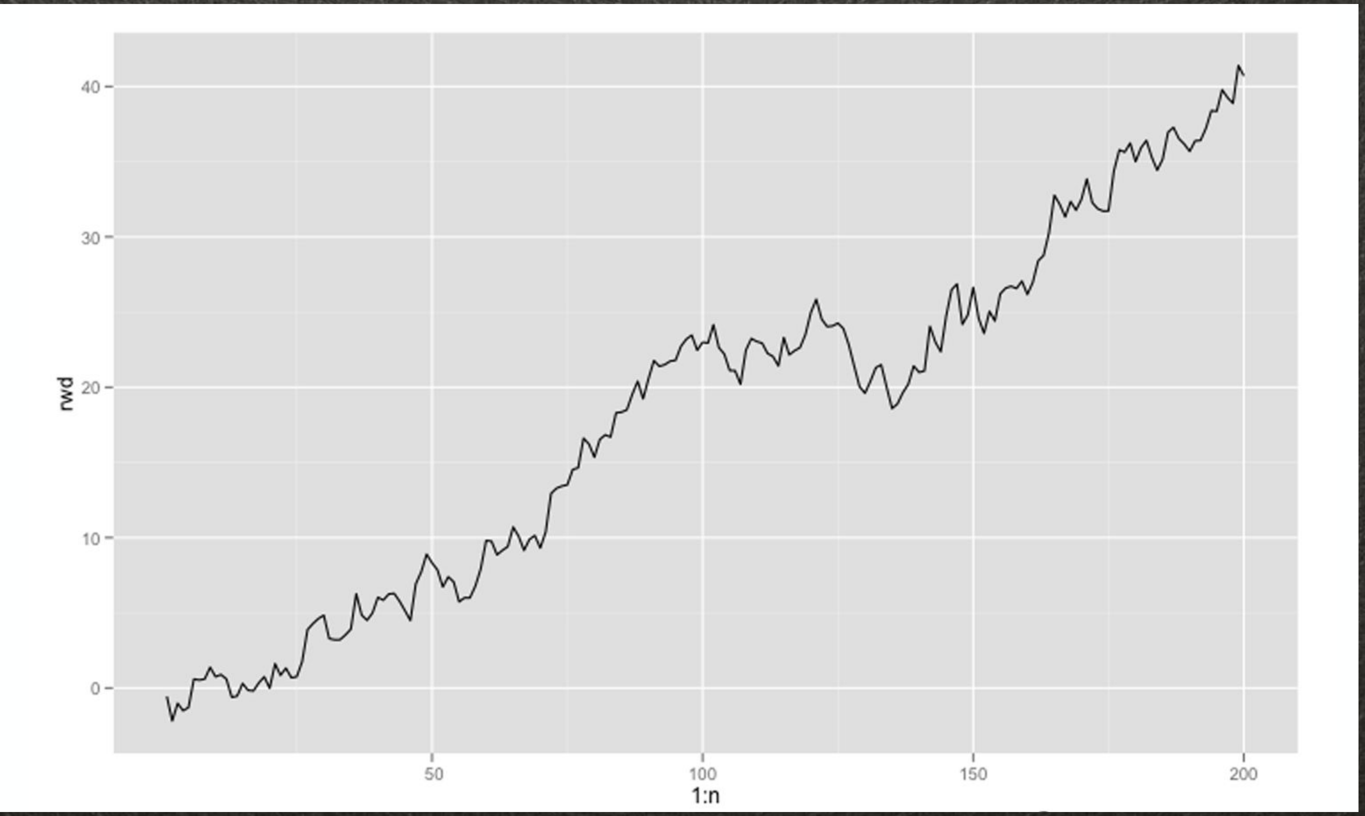
  - simply writing x_1, x_2, etc.
- Also known as “cammino dell’ubriaco”…

Random walk with drift

- If δ = 0:



- If δ = 0.1:



Random walk with drift

- Is a random walk (with drift) stationary?

$$E[x_t] = E[t\delta + \sum_{j=1}^{t} w_j] = t\delta$$

  - unless δ=0 the expectation value depends on t.

- For the covariance we need some more steps:

$$Cov(x_t, x_{t+h}) = Cov(t\delta + \sum_{j=1}^{t} w_j, (t+h)\delta + \sum_{j=1}^{t+h} w_j)$$

Random walk with drift

- Let's recall some useful relation for covariances :

$$Cov(\sum_i x_i, \sum_j y_j) = \sum_i \sum_j Cov(x_i, y_j)$$

$$Cov(X, X) = Var(X)$$

- So that:

$$Cov(x_t, x_{t+h}) = tVar(w_t) = t\sigma^2$$

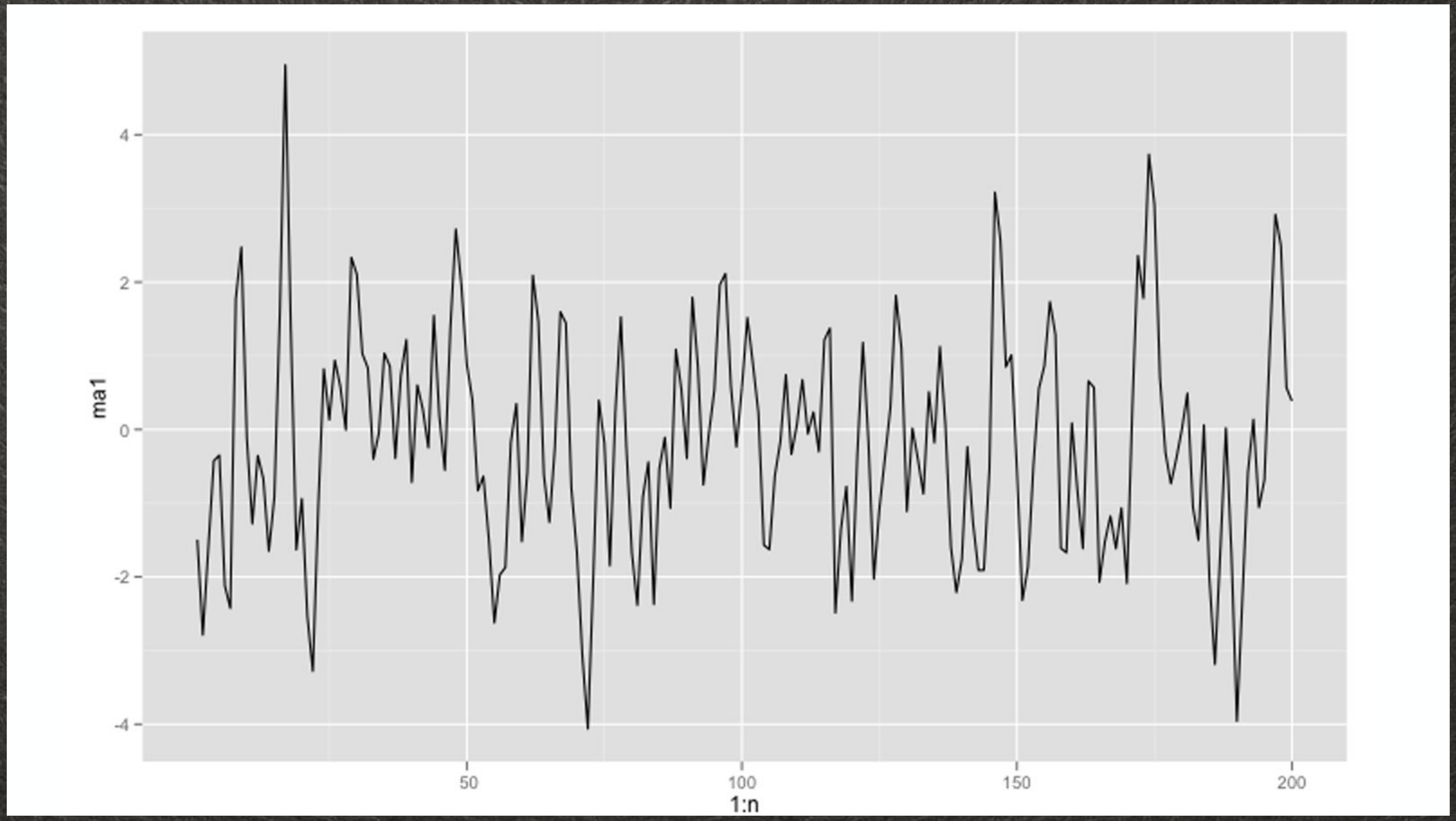
- which is clearly non-stationary (for any δ).

Moving Average MA(1)

$$x_t = \beta_1 w_{t-1} + w_t$$

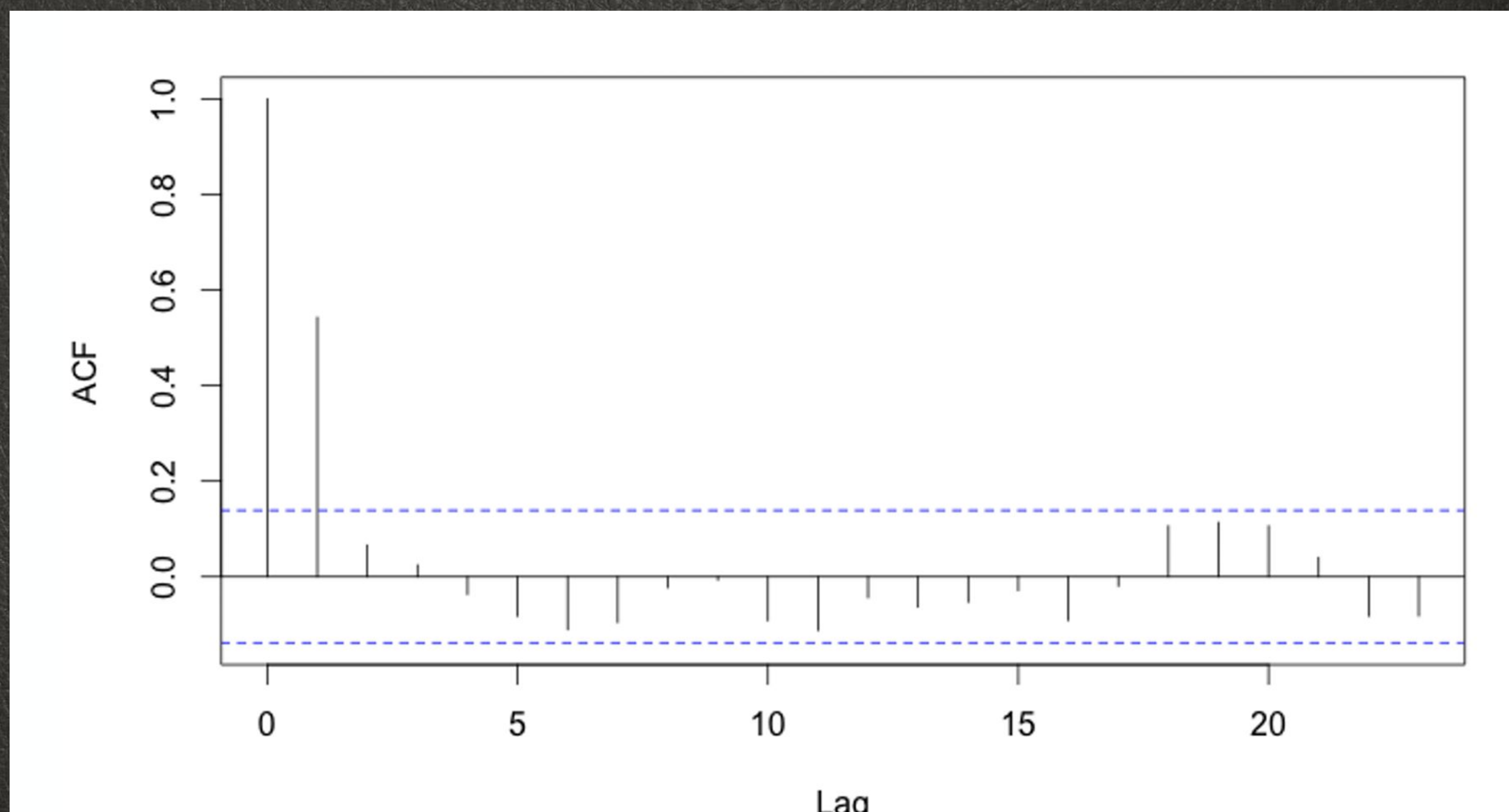
- where again w_t is a white noise process.

- $\beta_1 = 1$



MA(1) ACF

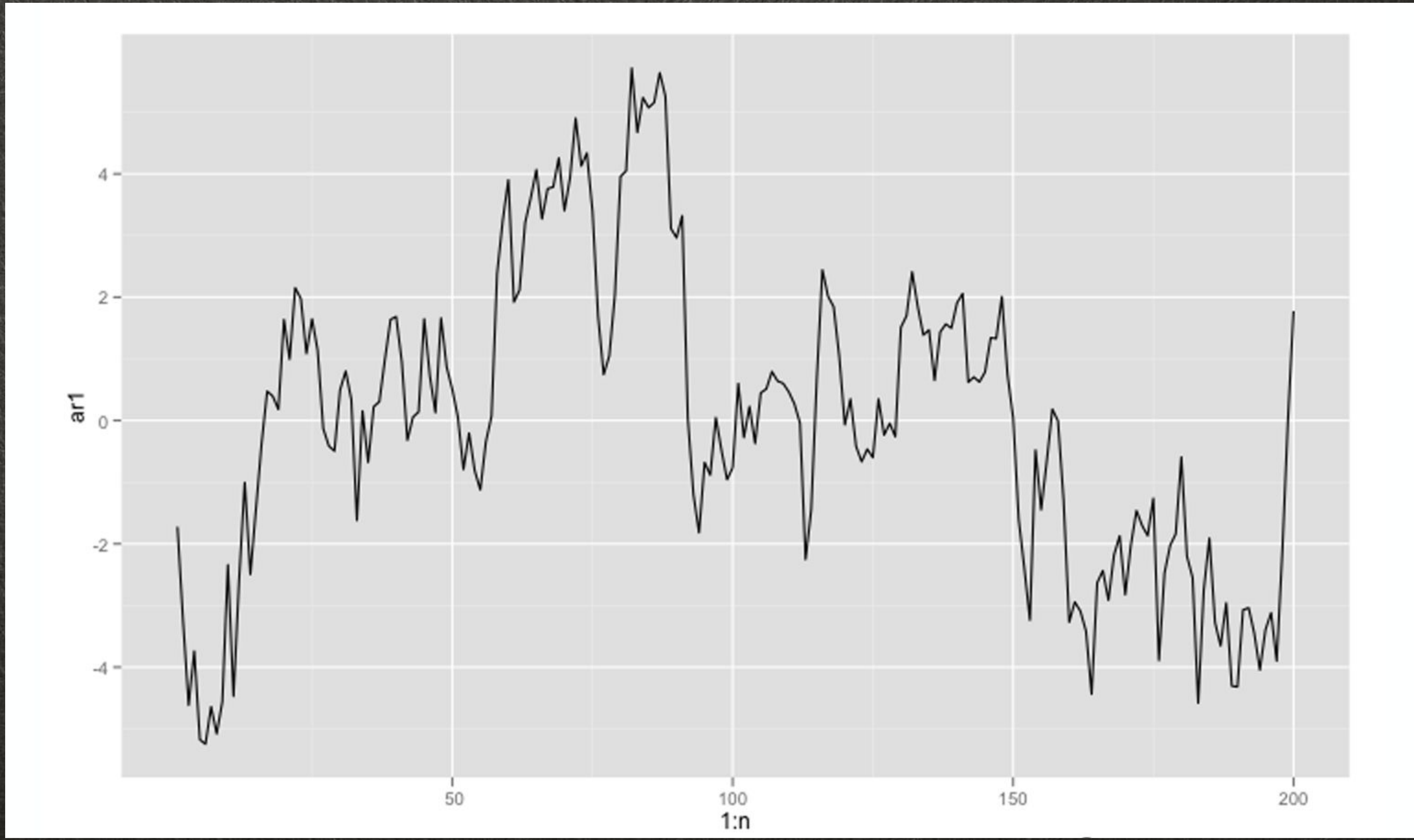
- $\beta_1 = 1$



Autoregressive AR(1)

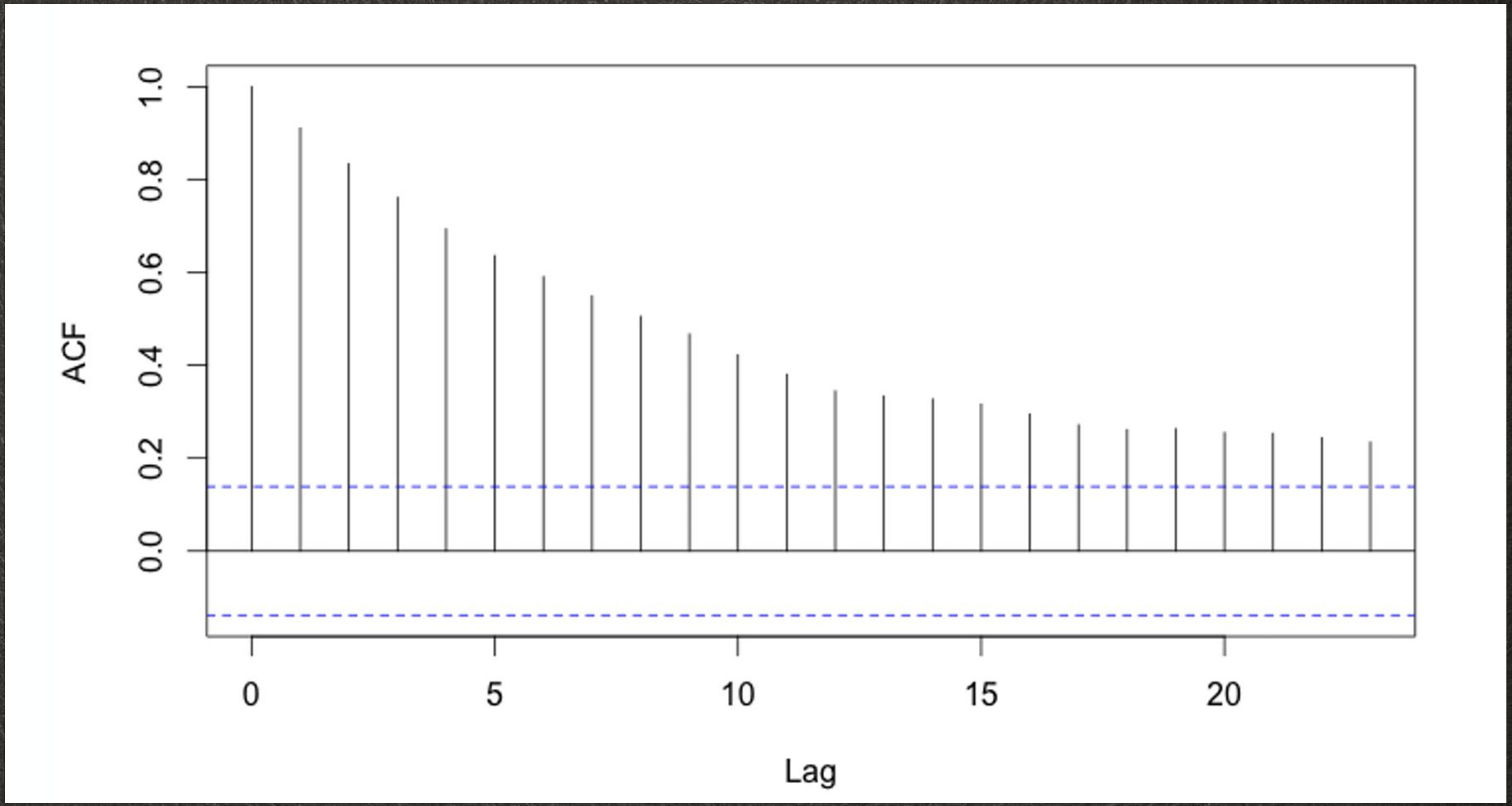
$$x_t = \alpha_1 x_{t-1} + w_t$$

- and again w_t is a white noise process.
- $\alpha_1 = 0.9$



AR(1) ACF

- $\alpha_1 = 0.9$

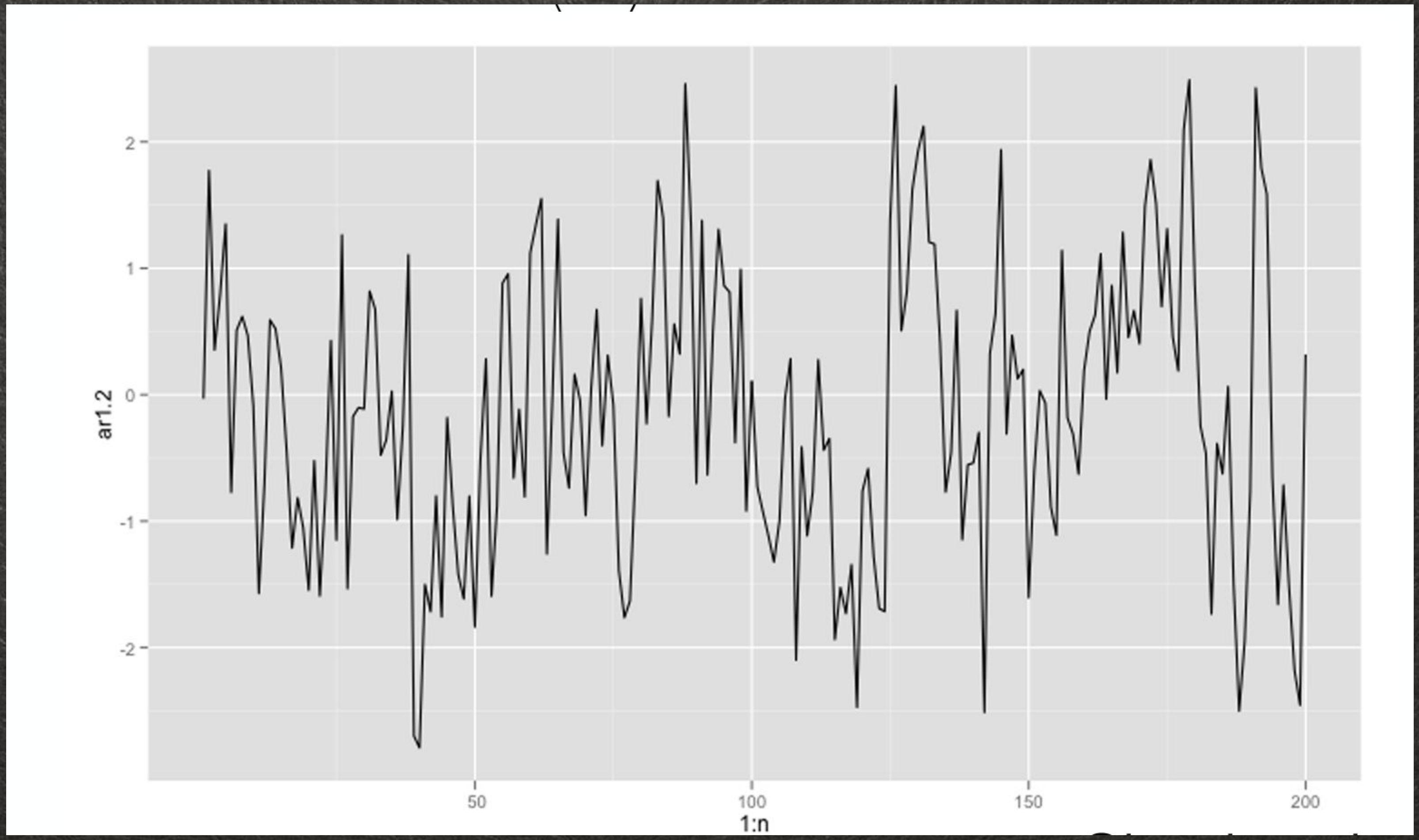


Autoregressive AR(1)

$$x_t = \alpha_1 x_{t-1} + w_t$$

- and again w_t is a white noise process.

- $\alpha_1 = 0.5$

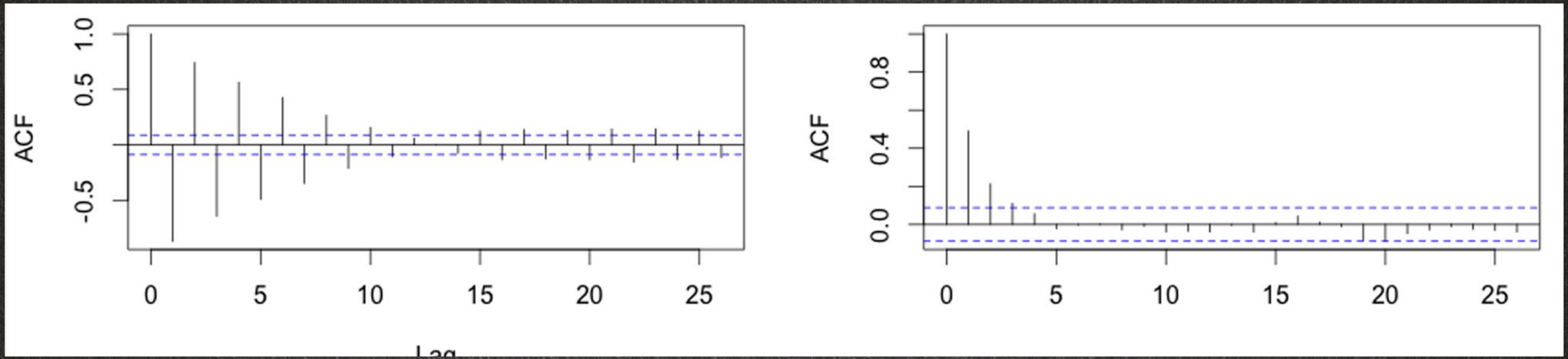


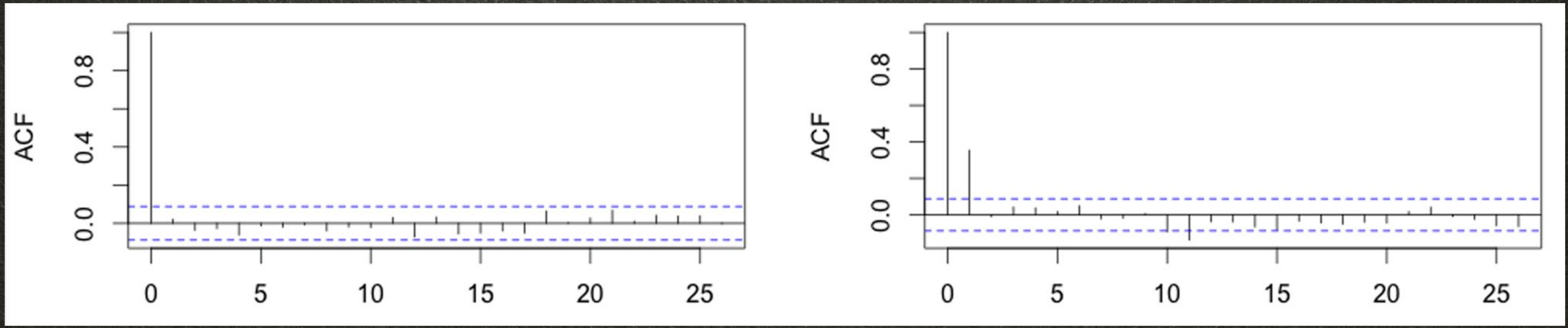
Stationary processes

White noise	MA(1), any β_1	AR(1), $\lvert\alpha_1\rvert < 1$
$\rho(h) = 1$, when $h = 0$ $= 0$, otherwise	$\rho(h) = 1$, when $h = 0$ $= \beta_1/(1 + \beta_1^2)$, $h = 1$ $= 0$, $h \geq 2$	$\rho(h) = 1$, when $h = 0$ $= \alpha_1^h$, $h > 0$
Only lag 0 shows non-zero ACF.	Only lag 0 and 1 show non-zero ACF.	Decreasing ACF

Exercise

1. Guess what these processes are basing on their ACF

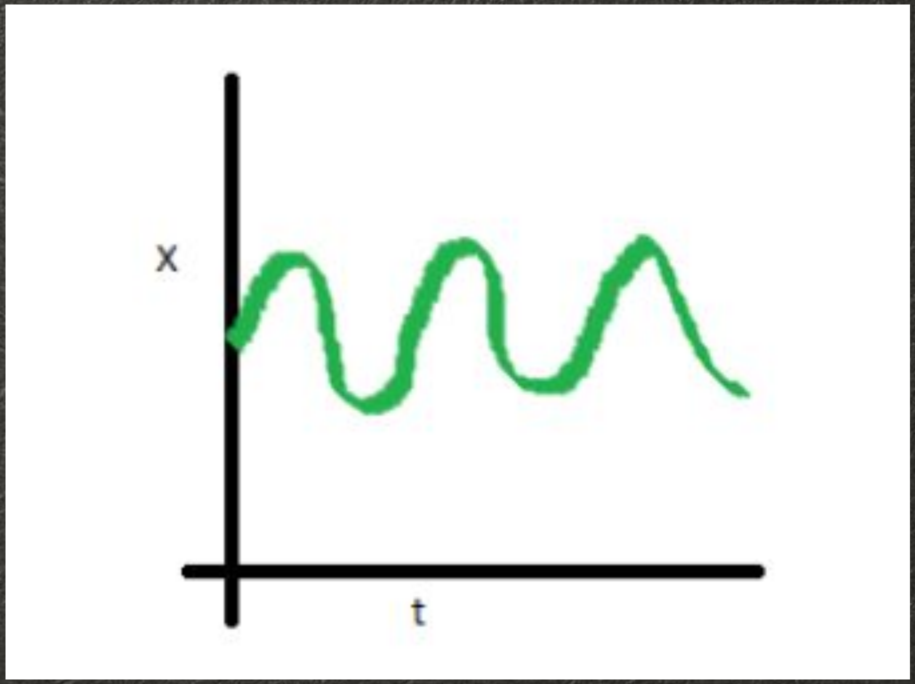




Exercise

- Useful notebook:

1. SingleTimeSeriesVariable



General Linear Processes

- We call "linear process" a linear combination of noise variates Z_t:

$$x_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i}$$

with

$$\sum_{i=0}^{\infty} |\psi_i| < \infty$$

This is enough to ensure stationarity

- and it can be shown that the autocovariance of a linear process is:

$$\gamma(h) = \sigma^2 \sum_{i=0}^{\infty} \psi_{i+h} \psi_i$$

Backshift Operator

- The backshift operator, B, is defined as:

$$Bx_t = x_{t-1}$$

- and it can be extended to more powers as:

$$B^2 x_t = (BB)x_t = B(Bx_t) = Bx_{t-1} = x_{t-2}$$

- so that:

$$B^k x_t = x_{t-k}$$

Difference Operator

- The difference operator, ∇, is defined as:

$$\nabla^d x_t = (1 - B)^d x_t$$

  - e.g. $\nabla^1 x_t = (1 - B)^1 x_t = x_t - x_{t-1}$

- The different operator is often used to convert non-stationary time-series to stationary.

Exercise

- MA(1): $$x_t = \beta_1 Z_{t-1} + Z_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i}$$
  - with: $\psi_0 = 1, \quad \psi_1 = \beta_1, \quad \psi_i = 0 \quad i \geq 2$
- AR(1): $$x_t = \alpha_1 x_{t-1} + Z_t = Z_t + \alpha_1 Z_{t-1} + \alpha_1^2 Z_{t-2} + \ldots$$
  - since $x_{t-1} = \alpha_1 x_{t-2} + Z_{t-1}$, etc. then: $$x_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i}$$
  - with: $\psi_i = \alpha_1^i$

MA(q) and AR(p)

- MA(q): $x_t = Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2} + \ldots + \beta_q Z_{t-q}$
- AR(p): $x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \ldots + \alpha_p x_{t-p} + Z_t$

$$\theta(B) = 1 + \beta_1 B + \beta_2 B^2 + \ldots + \beta_q B^q$$
$$\phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p$$

MA(q): $x_t = \theta(B) Z_t$

AR(p): $\phi(B) x_t = Z_t$

ARMA(p,q) processes

- A process, x_t, is ARMA(p,q) if it has the form:

$$\phi(B)\, x_t = \theta(B)\, Z_t,$$

- And let's remember that Z_t is white noise and:

$$\theta(B) = 1 + \beta_1 B + \beta_2 B^2 + \ldots + \beta_q B^q$$
$$\phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \ldots - \alpha_p B^p$$

- ARMA(p,q) is stationary if and only if the roots of $\Phi(Z)$ lie outside the unit circle.
- ARMA(p,q) is invertible if and only if the roots of $\Theta(Z)$ lie outside the unit circle.

Just to make things even more chaotic...

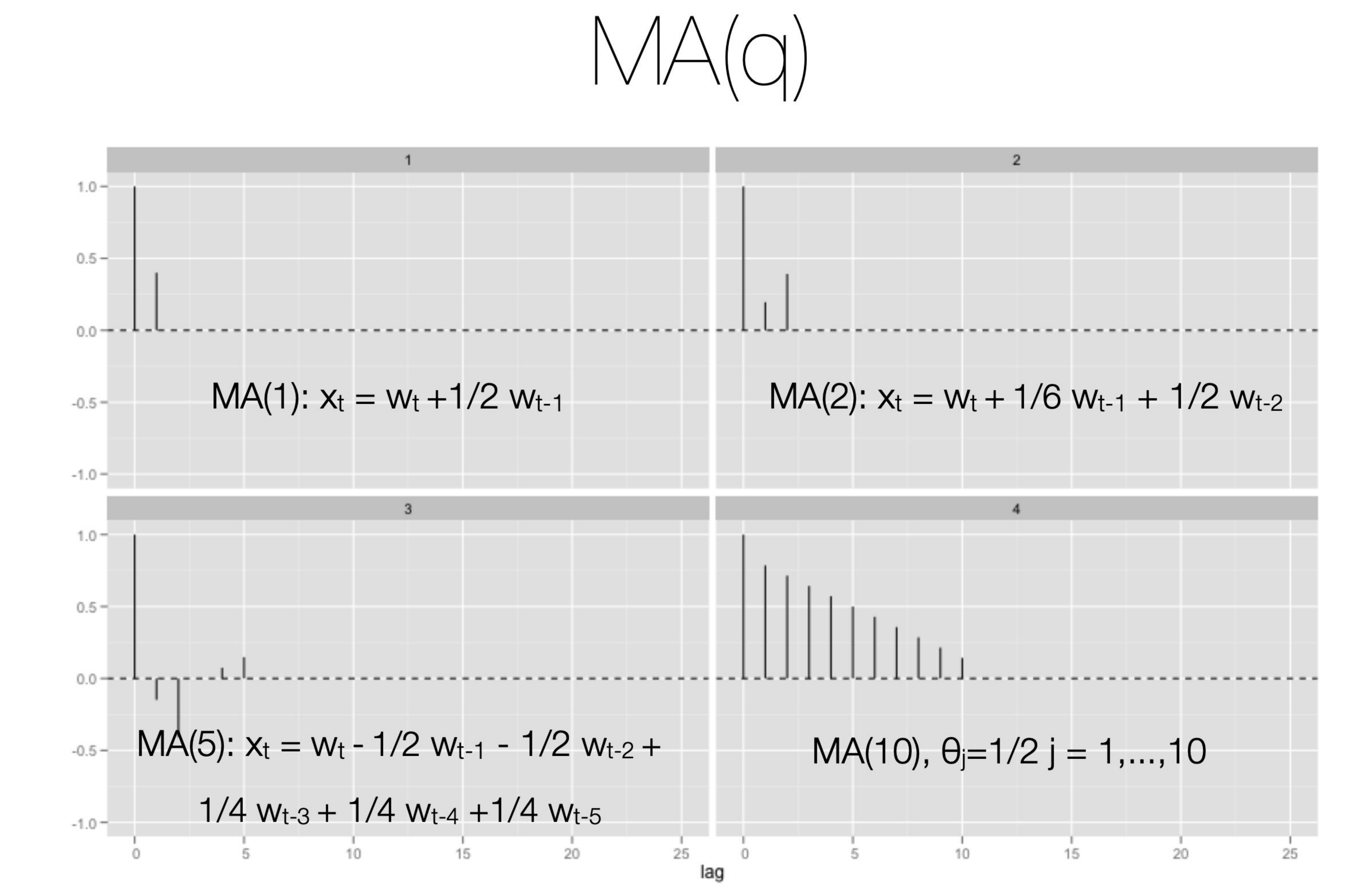
- It is possible to take an ARMA process $\phi(B)x_t = \theta(B)Z_t$
  - where Φ(B) and φ(B) are finite order polynomials,
- and convert it to an infinite order MA process:

$$x_t = \psi(B)Z_t \qquad \psi(B) = \theta(B)/\phi(B)$$

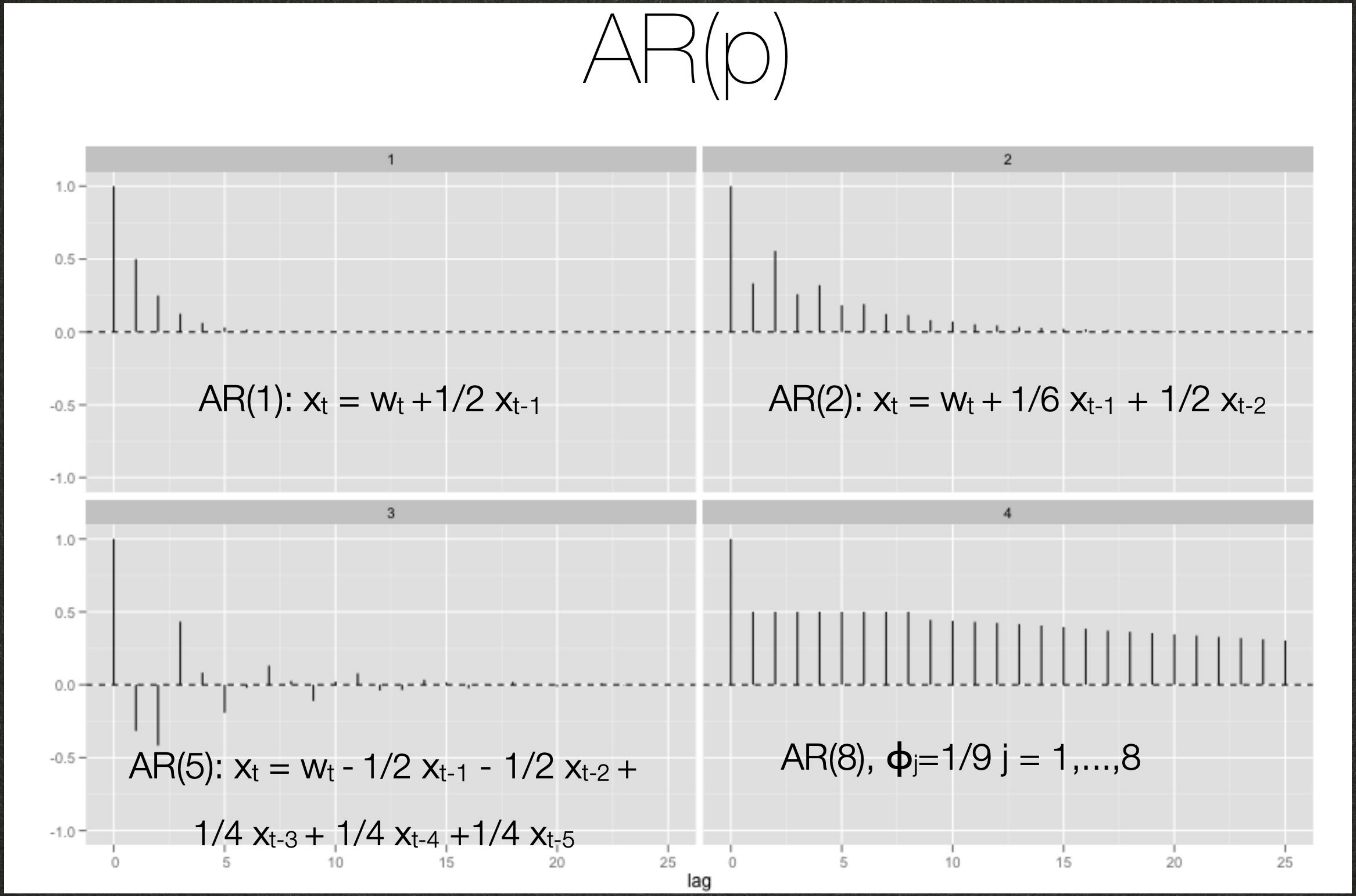
- or an infinite order AR process:

$$\pi(B)x_t = Z_t \qquad \pi(B) = \phi(B)/\theta(B)$$

ACF


MA(q)
1
2
3
4
MA(1): $x_t = w_t + 1/2\, w_{t-1}$
MA(2): $x_t = w_t + 1/6\, w_{t-1} + 1/2\, w_{t-2}$
MA(5): $x_t = w_t - 1/2\, w_{t-1} - 1/2\, w_{t-2} +$
$1/4\, w_{t-3} + 1/4\, w_{t-4} + 1/4\, w_{t-5}$
MA(10), θ_j=1/2 j = 1,...,10
lag

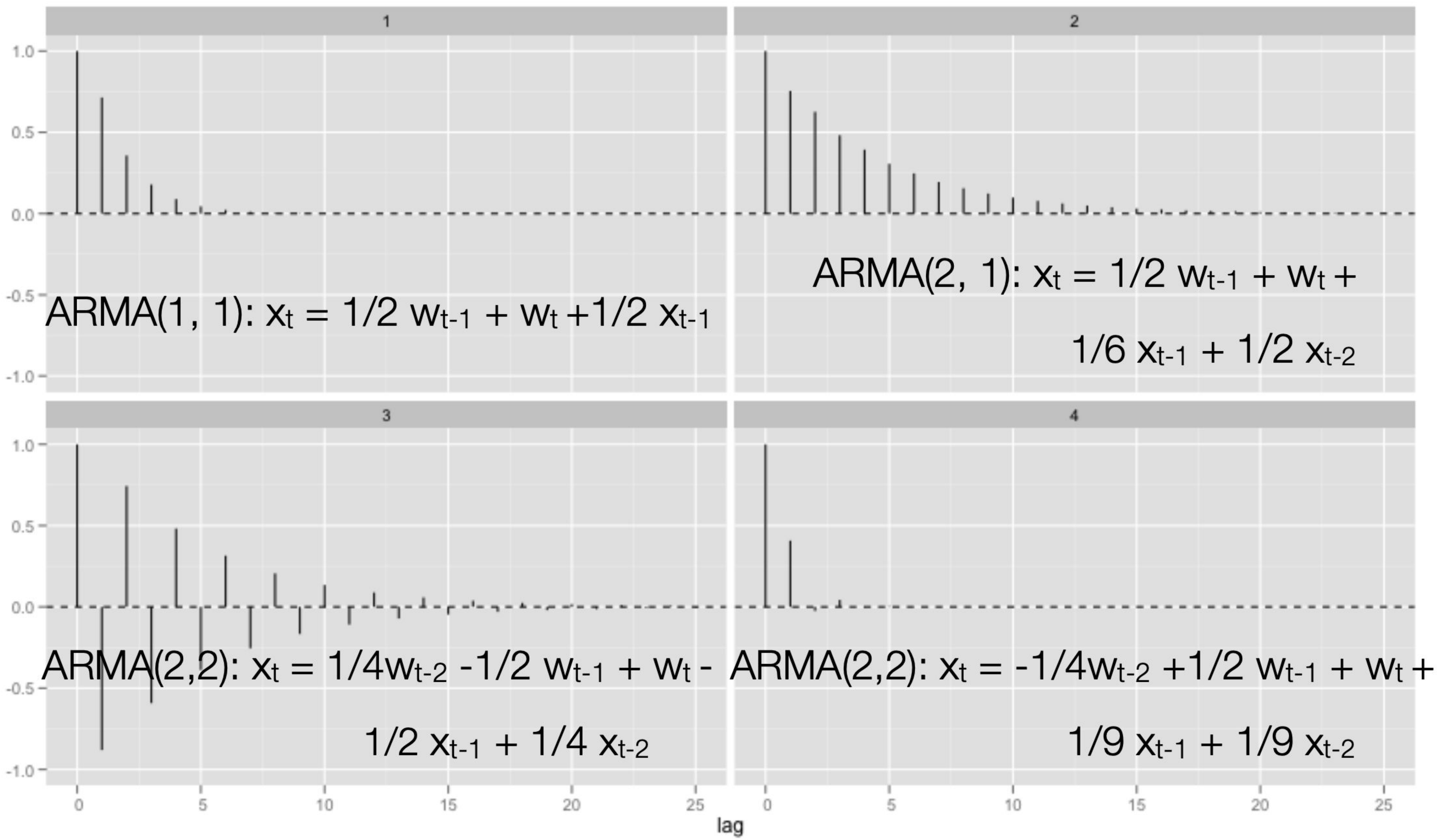
ACF



AR(p)
1
2
3
4
1.0
0.5
0.0
-0.5
-1.0
AR(1): x_t = w_t +1/2 x_t-1
AR(2): x_t = w_t + 1/6 x_t-1 + 1/2 x_t-2
AR(5): x_t = w_t - 1/2 x_t-1 - 1/2 x_t-2 +
1/4 x_t-3 + 1/4 x_t-4 +1/4 x_t-5
AR(8), φ_j=1/9 j = 1,...,8
0
5
10
15
20
25
lag

ACF

ARMA(p, q)



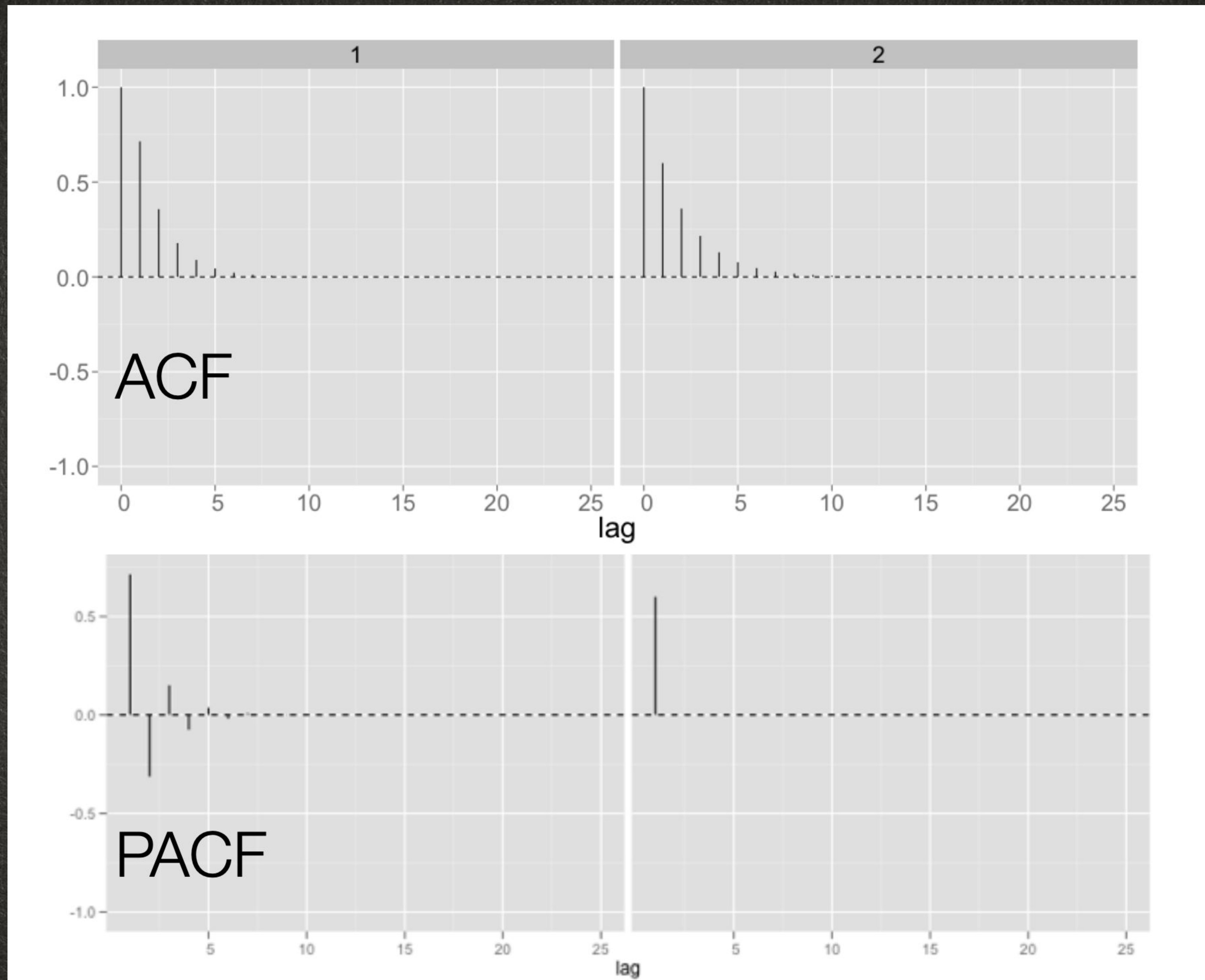
1
2
3
4
1.0
0.5
0.0
-0.5
-1.0
0
5
10
15
20
25
lag
ARMA(1, 1): $x_t = 1/2\ w_{t-1} + w_t + 1/2\ x_{t-1}$
ARMA(2, 1): $x_t = 1/2\ w_{t-1} + w_t +$
$1/6\ x_{t-1} + 1/2\ x_{t-2}$
ARMA(2,2): $x_t = 1/4 w_{t-2} - 1/2\ w_{t-1} + w_t -$
$1/2\ x_{t-1} + 1/4\ x_{t-2}$
ARMA(2,2): $x_t = -1/4 w_{t-2} + 1/2\ w_{t-1} + w_t +$
$1/9\ x_{t-1} + 1/9\ x_{t-2}$

ACF & PACF for MA, AR and ARMA processes

	MA(q)	AR(p)	ARMA(p, q)
ACF	zero lags > q	tails off	tails off
PACF	tails off	zero lags > p	tails off

One is AR(1), $\alpha_1 = 0.6$

The other is ARMA(1, 1), $\beta_1 = 0.5$, $\alpha_1 = 0.5$



ARIMA(p,d,q)
Autoregressive Integrated Moving Average

- A process x_t is ARIMA(p,d,q) if x_t, differenced “d times ($\nabla^d x_t$), is an ARMA(p,q) process!

$$\phi(B)\,\nabla^d x_t = \theta(B)\,w_t$$

$$\phi(B)\,(1 - B)^d x_t = \theta(B)\,w_t$$

ARIMA(p,d,q) fitting

- Having chosen the right orders of the moving average autoregressive processes there are different strategies for determining the ARIMA parameters (Yule-Walker equations, maximum-likelihood, etc.).

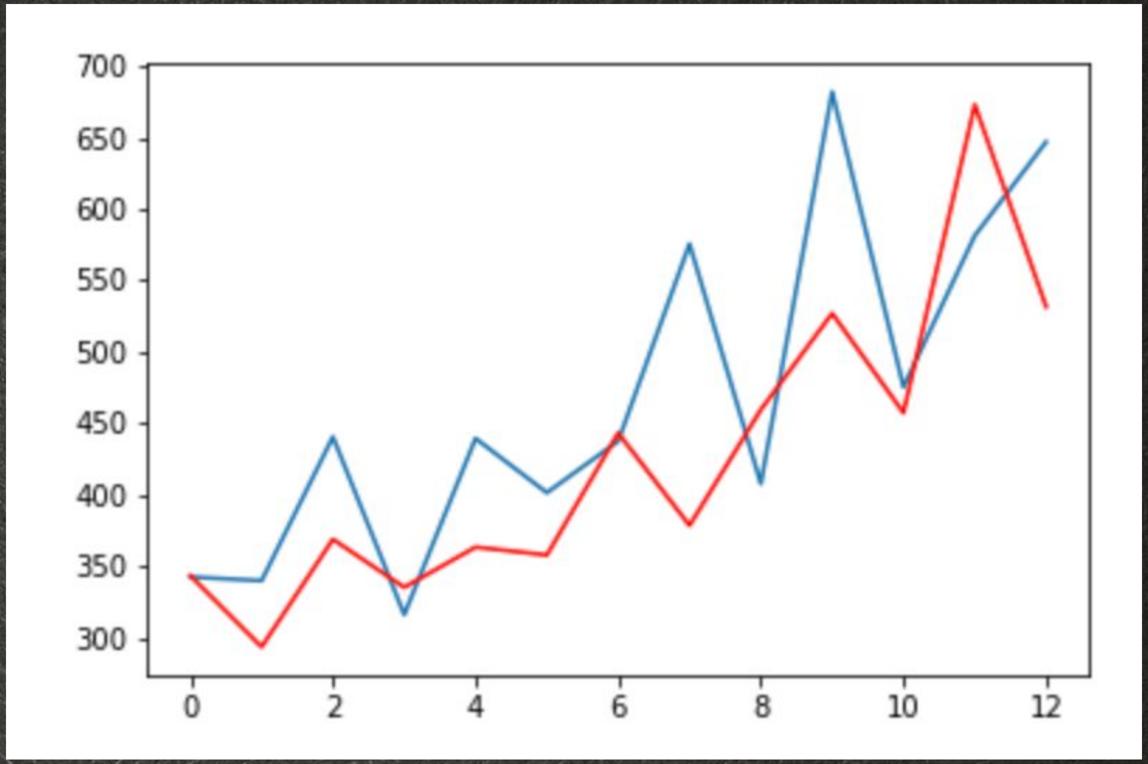
- Typically, one could follow this procedure:

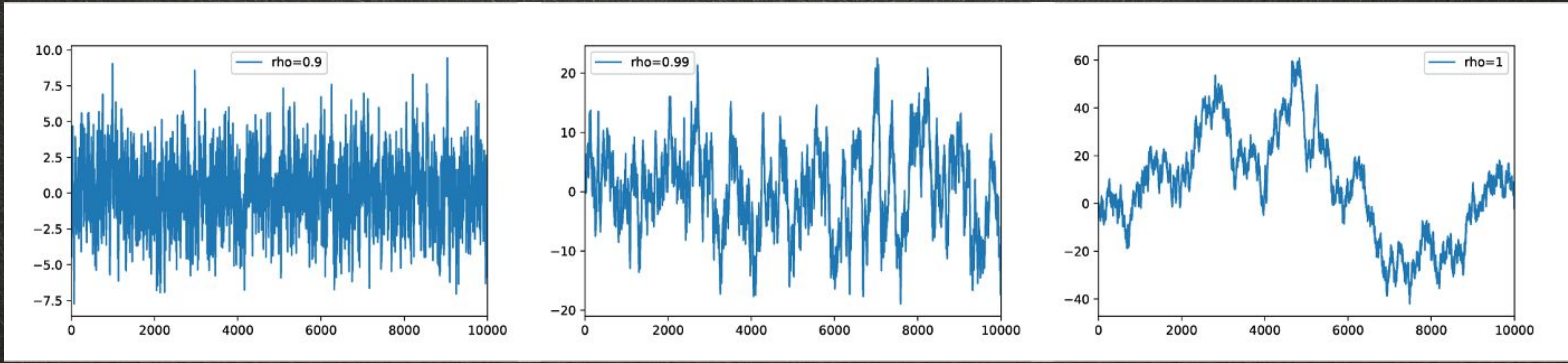
  1. Plot the data
  2. Difference until series is stationary (find d)
  3. Examine ACF and PACF to guess p and q
  4. Fit ARIMA(p,d,q) to the original data
  5. Check model diagnostics

Exercise

- Useful notebooks:

1. ARIMAModel
2. ARIMAForecasting
3. ARIMAModelParameters





Just to mention...

- Another form of nonstationarity occurs when the variance, rather than the local mean level, of the time series changes during the observation.
- This is commonly called **volatility** in econometric contexts.
- These kind of problems can be addressed by the **autoregressive conditional heteroscedastic (ARCH) models**. Here the variance is assumed to be a stochastic autoregressive process depending on previous values.
- For instance, in a ARCH(1) model:

$$Var_{ARCH(1)}(\epsilon_i) = \alpha_0 + \alpha_1 Var(\epsilon_{i-1}).$$

The zoo is ways richer than you think...

- ARFIMA: autoregressive moving average fractionally integrated
- SARIMA: multiplicative seasonal autoregressive integrated moving average model
- CARFIMA: designed for irregularly sampled time series
- ARMAX: where X stands for "exogenous covariate"

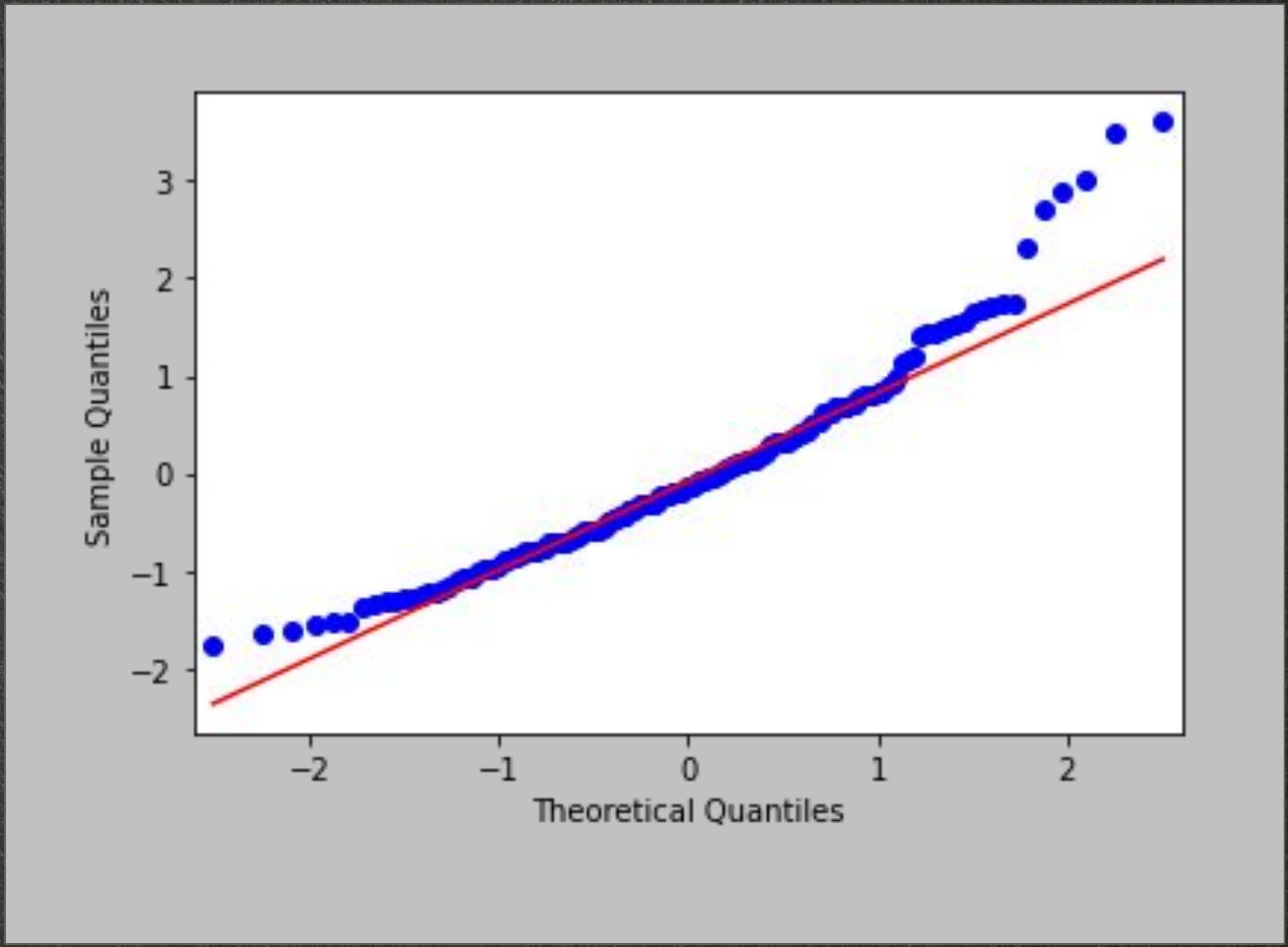
$$x_t = \sum_{i=1}^{p} a_i x_{t-i} + \sum_{j=1}^{q} b_j \epsilon_{t-j} + \sum_{k-1}^{r} c_k y_{tk}$$

- The exogenous covariates can be any function linear in the parameters: periodic, or some deterministic trend (e.g., polynomial), or some tabulated variables without a simple mathematical form.

Exercise

- Useful notebook:

1. PG_ARIMA



State Space Models

- It is a broader approach to parametric modeling of complicated time series, where autoregressive and other stochastic behaviors can be combined with linear or nonlinear deterministic trends or periodic components.

- Functional behaviors can be linear or nonlinear, and can involve derivatives of state variables. Noise can be Gaussian or non-Gaussian, white or autoregressive.

- State space models are defined hierarchically with one level describing relationships between "state variables" defining the temporal behavior of the system and other levels describing how the underlying system relates to the observables.

State Space Models

- A state-space model for a (possibly multivariate) time series $\{Y_t, t = 1,2,...\}$ consists of two equations. The first, known as the **observation equation**, expresses the w-dimensional observation Y_t as a linear function of a v-dimensional state variable X_t plus noise.

$$Y_t = G_t X_t + W_t, \quad t = 1.2....$$

- The second equation is called the state equation, and determines the state X_{t+1} at time t+1 in terms of the previous state X_t and a noise term.

$$X_{t+1} = F_t X_t + V_t, \quad t = 1.2, \ldots$$

- The noise terms are often (but not necessarily) supposed to be uncorrelated.

State Space Models

- It is possible to find a state-space representation for a large number of time-series.

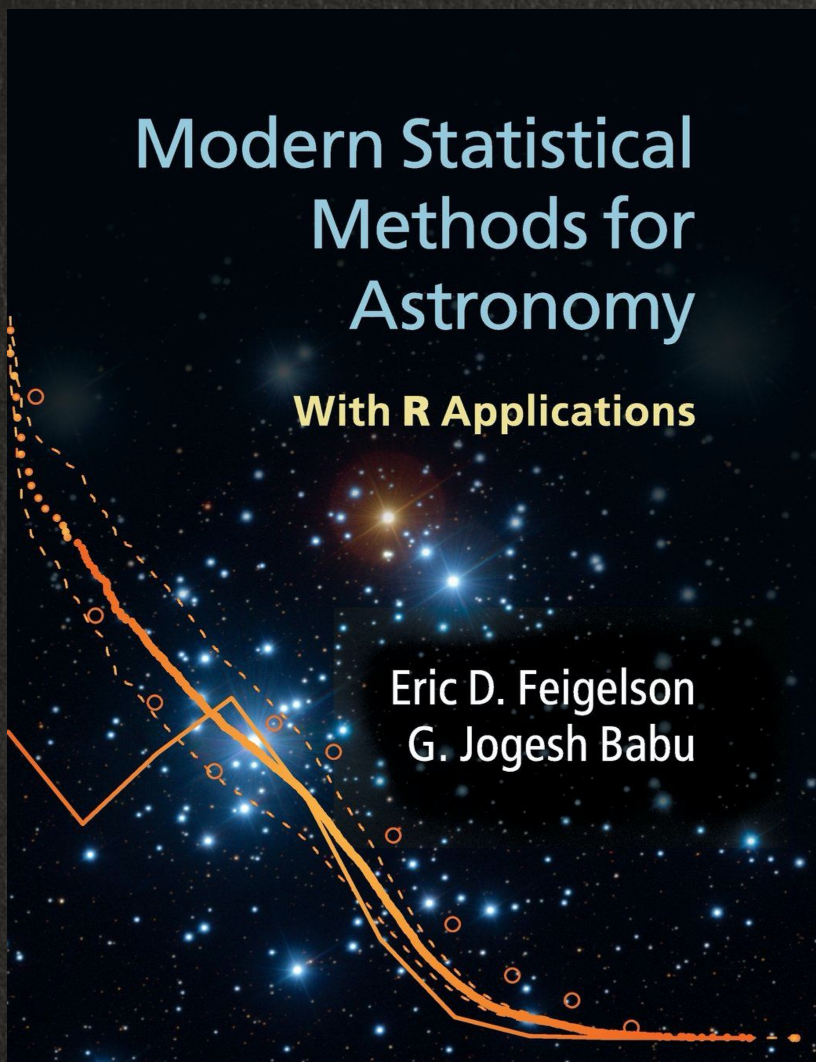
- Neither $\{X_t\}$ nor $\{Y_t\}$ is necessarily stationary.

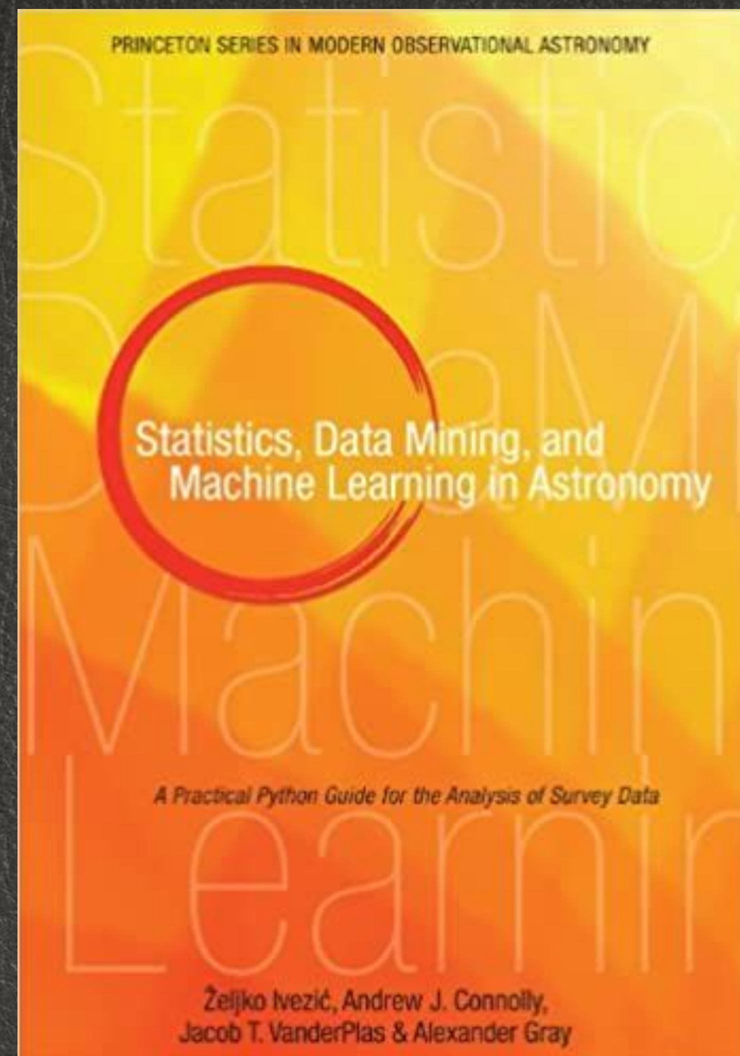
- The beauty of a state-space representation lies in the simple structure of the state equation, which permits relatively simple analysis of the process $\{X_t\}$.

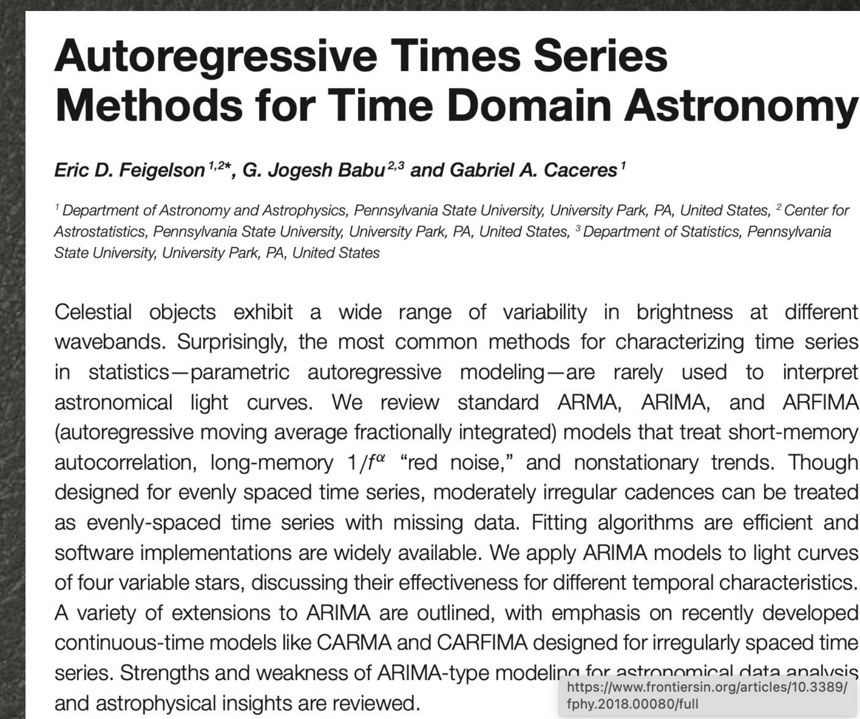
- The behavior of $\{Y_t\}$ is then easy to determine from that of $\{X_t\}$ using the observation equation.

- The coefficients of the model are calculated by least squares or maximum likelihood estimation, often through a recursive algorithm such as, e.g., the Kalman filter.

REFERENCES AND DEEPENING







Autoregressive Times Series Methods for Time Domain Astronomy

Eric D. Feigelson[1,2], G. Jogesh Babu[2,3] and Gabriel A. Caceres[1]*

[1] Department of Astronomy and Astrophysics, Pennsylvania State University, University Park, PA, United States, [2] Center for Astrostatistics, Pennsylvania State University, University Park, PA, United States, [3] Department of Statistics, Pennsylvania State University, University Park, PA, United States

Celestial objects exhibit a wide range of variability in brightness at different wavebands. Surprisingly, the most common methods for characterizing time series in statistics—parametric autoregressive modeling—are rarely used to interpret astronomical light curves. We review standard ARMA, ARIMA, and ARFIMA (autoregressive moving average fractionally integrated) models that treat short-memory autocorrelation, long-memory $1/f^{\alpha}$ "red noise," and nonstationary trends. Though designed for evenly spaced time series, moderately irregular cadences can be treated as evenly-spaced time series with missing data. Fitting algorithms are efficient and software implementations are widely available. We apply ARIMA models to light curves of four variable stars, discussing their effectiveness for different temporal characteristics. A variety of extensions to ARIMA are outlined, with emphasis on recently developed continuous-time models like CARMA and CARFIMA designed for irregularly spaced time series. Strengths and weakness of ARIMA-type modeling for astronomical data analysis and astrophysical insights are reviewed.

https://www.frontiersin.org/articles/10.3389/fphy.2018.00080/full

Eric Feigelson et al.

Eric Feigelson & Jogesh Babu


Zeljko Izevic et al.