

Time-Series Analysis: Stochastic Models

Evgeny Burnaev

Skoltech, Moscow, Russia

- 1 Motivation
- 2 Time series models
- 3 Stationarity
- 4 Auto Regression Moving Average (ARMA)

1 Motivation

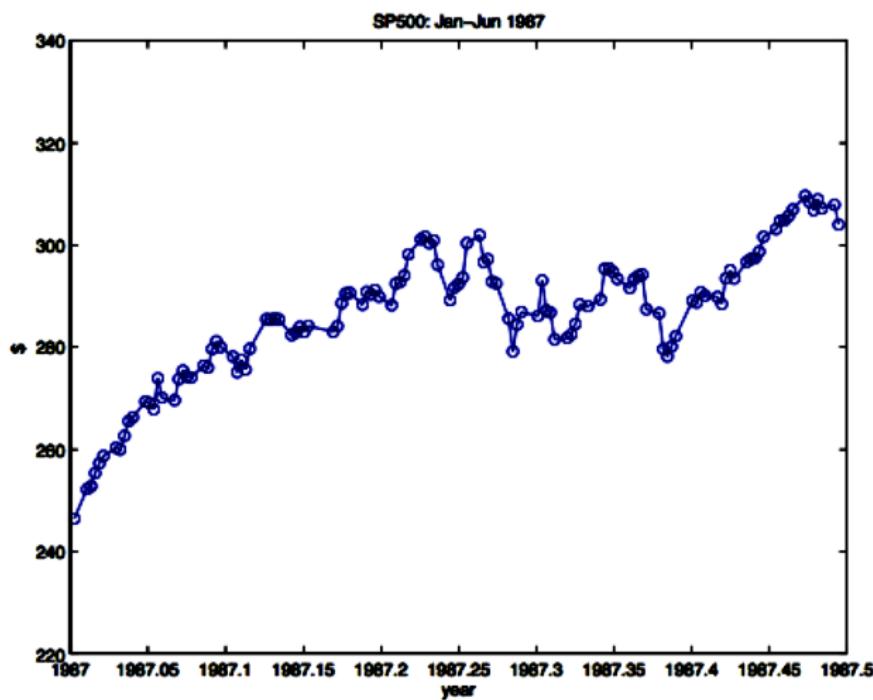
2 Time series models

3 Stationarity

4 Auto Regression Moving Average (ARMA)

Example

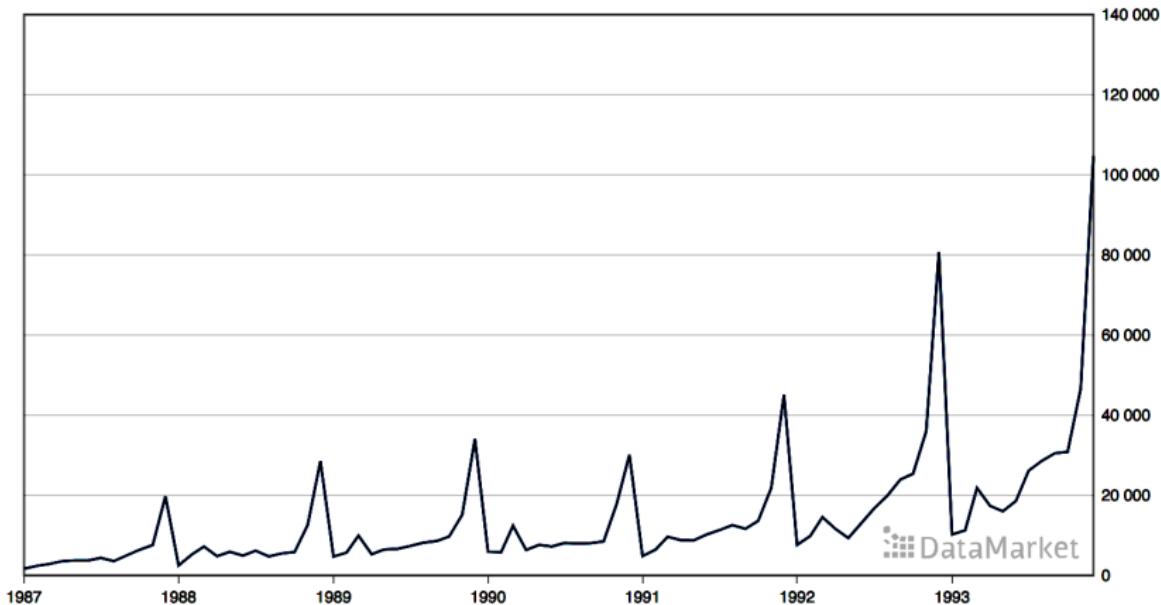
- SP500: Jan-Jun 1987

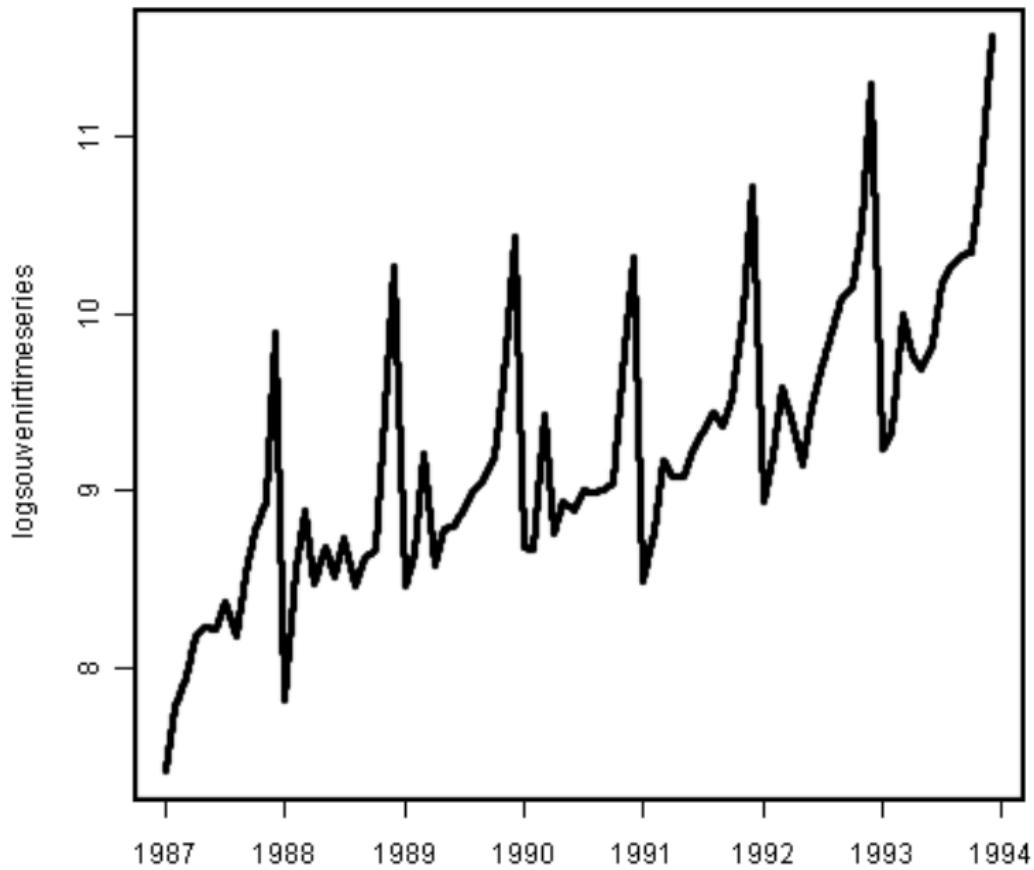


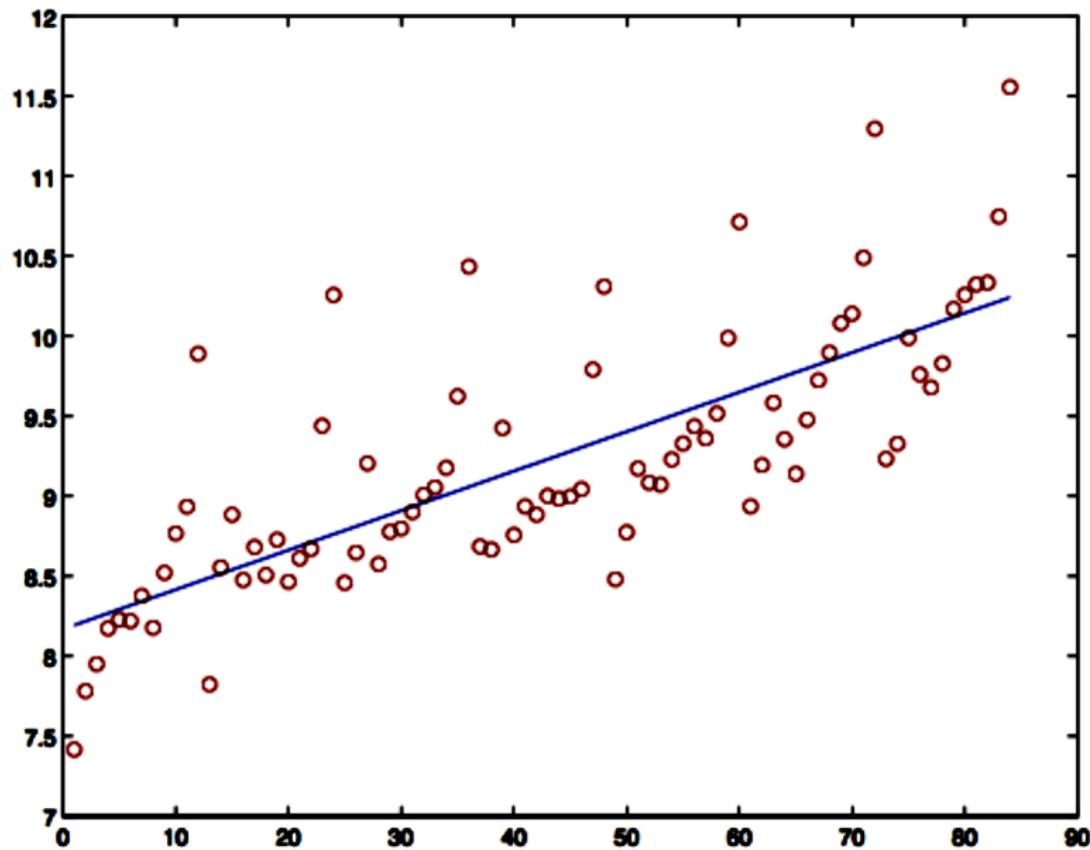
- Compact description of data
- Interpretation
- Forecasting
- Control
- Hypothesis testing
- Simulation

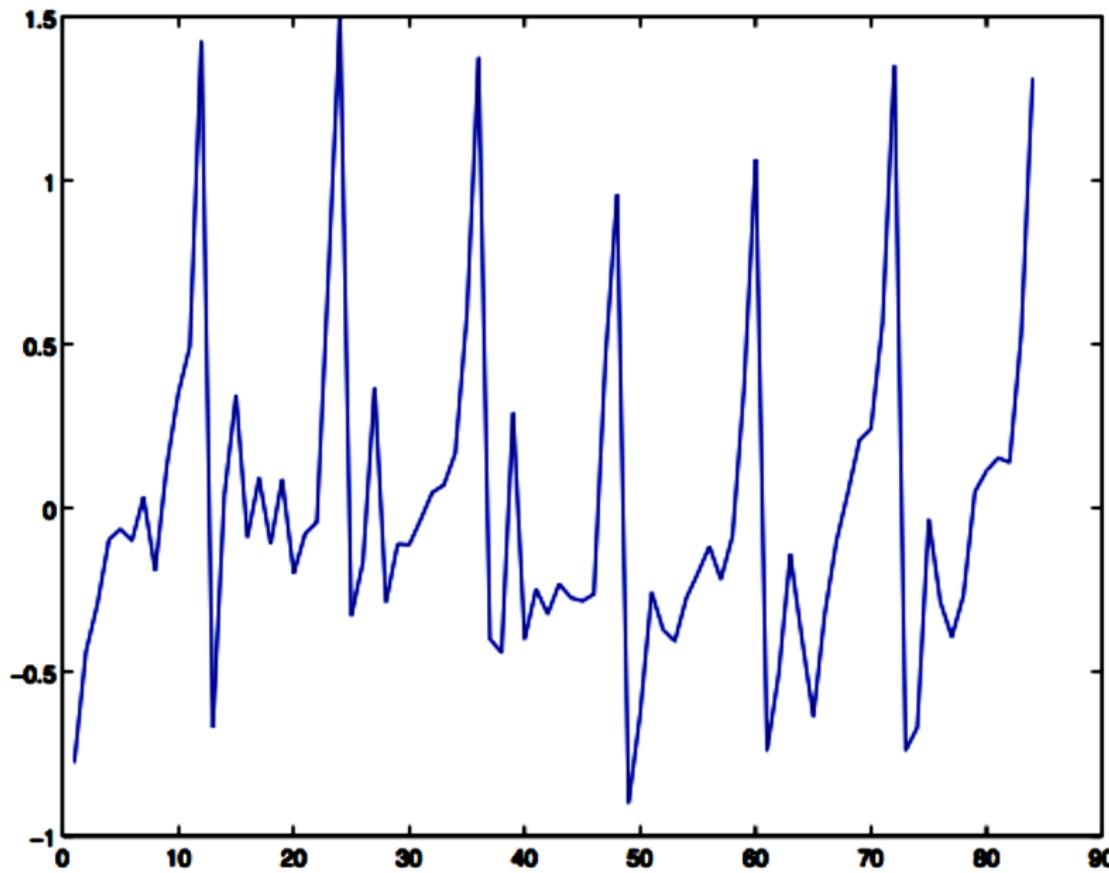
Classical decomposition: An example

- Monthly sales for a souvenir shop at a beach resort town in Queensland (Makridakis, Wheelwright and Hyndman, 1998)

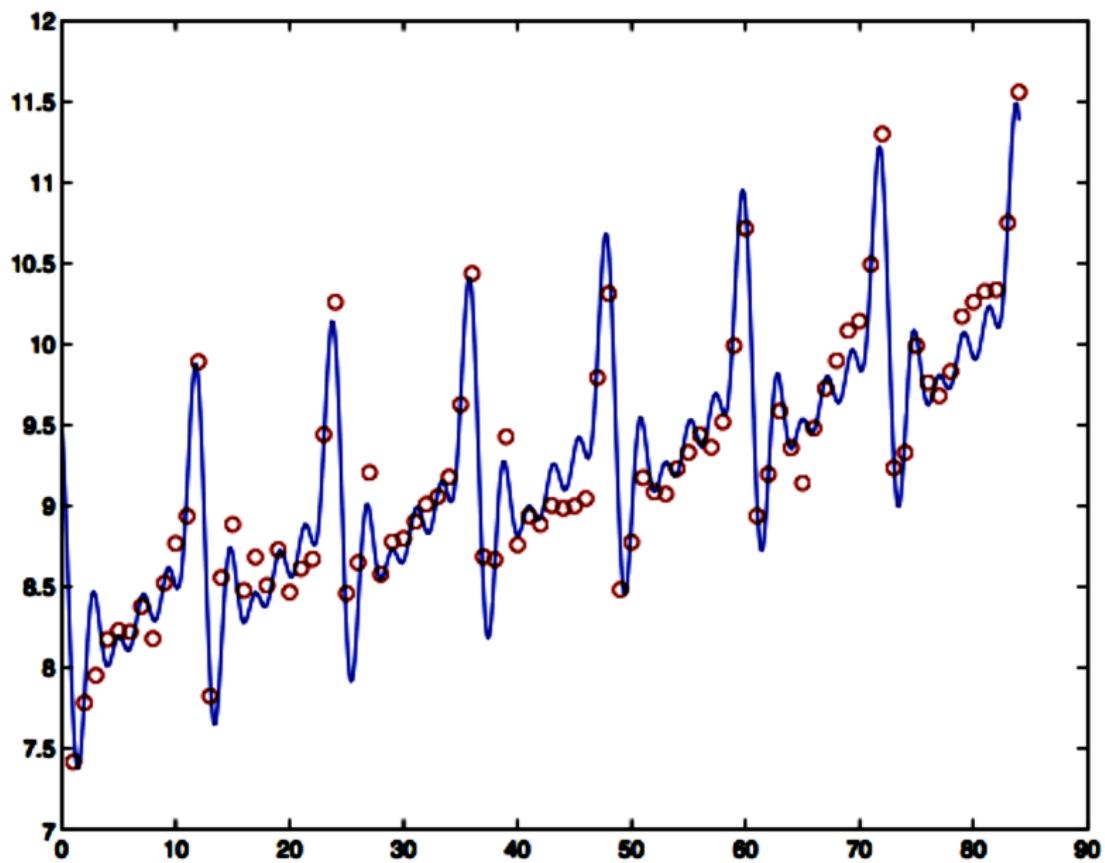






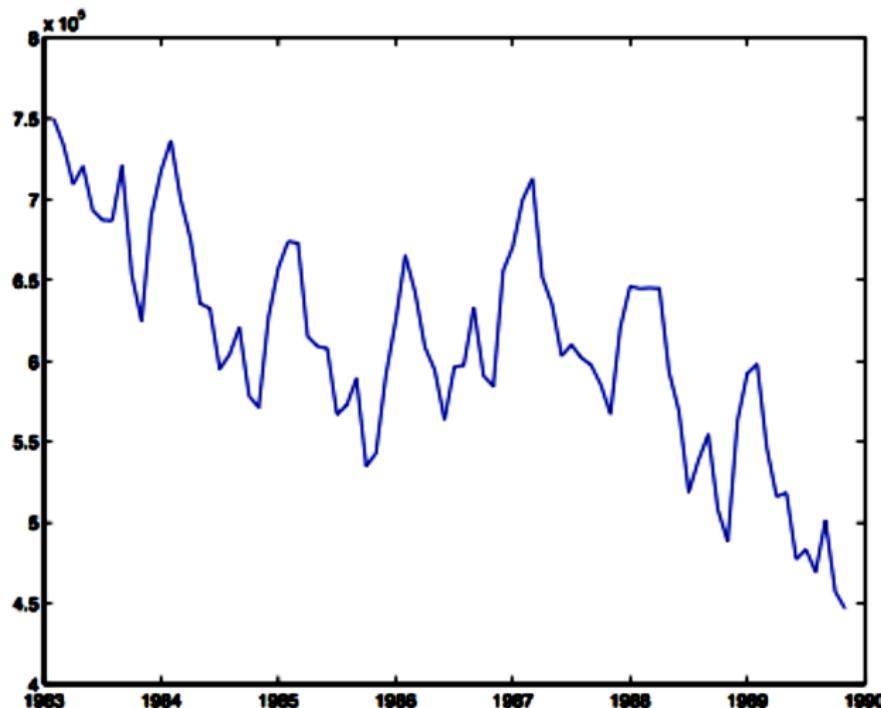


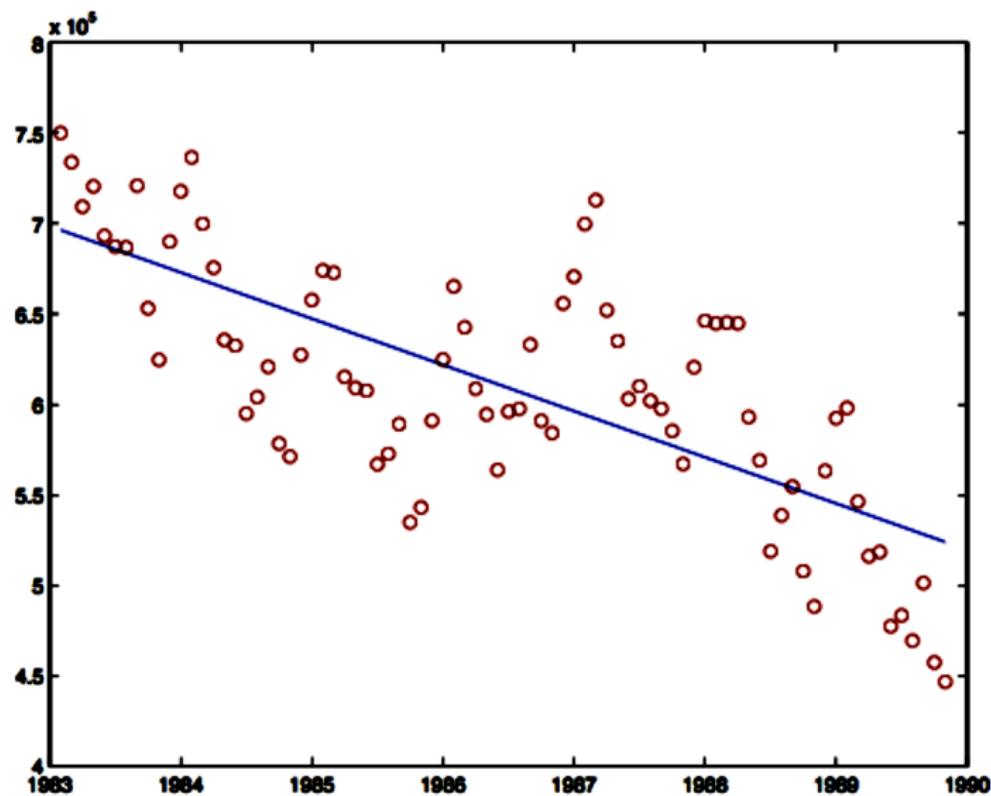
Trend and seasonal variation



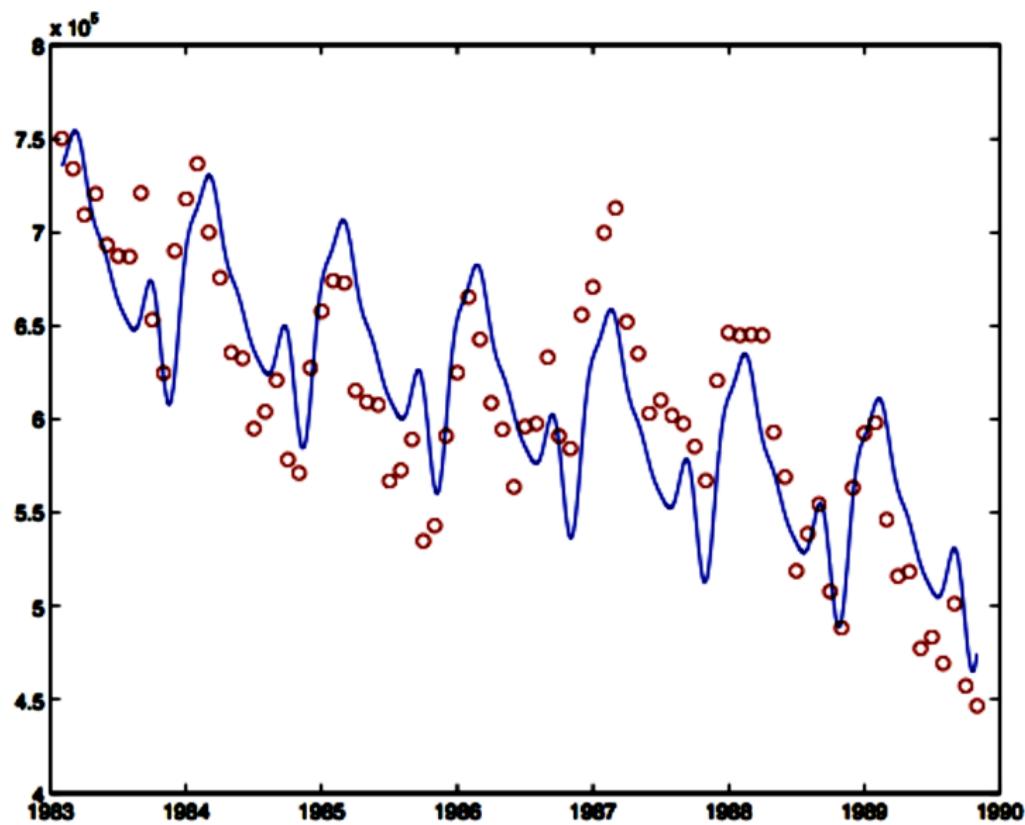
- Compact description of data \Rightarrow
Example: Classical decomposition $y_t = x_t + s_t + \varepsilon_t$
- Interpretation \Rightarrow Example: Seasonal adjustment
- Forecasting \Rightarrow Example: Predict sales
- Control
- Hypothesis testing
- Simulation

- Monthly number of unemployed people in Australia (Hipel and McLeod, 1994)





Trend + seasonal variation



- Compact description of data \Rightarrow

$$y_t = x_t + s_t + f(c_t) + \varepsilon_t$$

- Interpretation \Rightarrow Example: Seasonal adjustment
- Forecasting \Rightarrow Example: Predict unemployment
- Control \Rightarrow Example: Impact of monetary policy on unemployment
- Hypothesis testing \Rightarrow Example: Global warming
- Simulation \Rightarrow Example: Estimate probability of catastrophic events

1 Motivation

2 Time series models

3 Stationarity

4 Auto Regression Moving Average (ARMA)

- A time series model specifies the joint distribution of the sequence $\{y_t\}_{t=1,2,\dots}$ of random variables

$$\mathbb{P}[y_{t_1} \leq z_1, \dots, y_{t_k} \leq z_k] \text{ for any } k \text{ and } z_1, \dots, z_k$$

- We mostly restrict our attention to **second-order** properties only:

$$\mathbb{E}y_t, \quad \mathbb{E}(y_{t_1}y_{t_2})$$

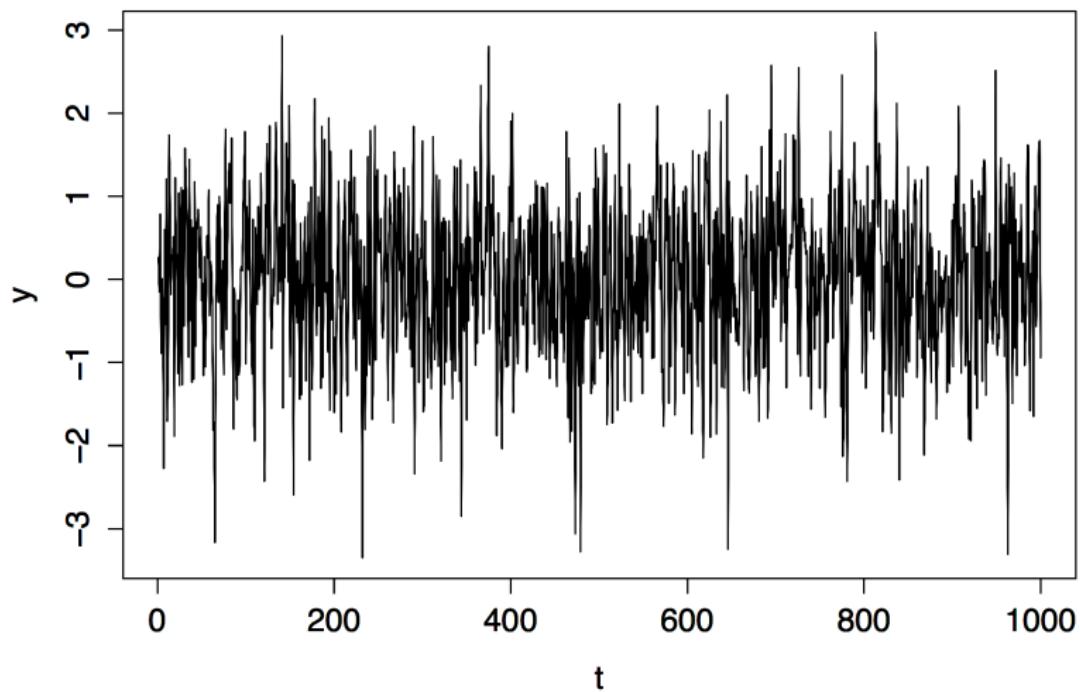
- Example: white noise $y_t \sim WN(0, \sigma^2)$, i.e. $\{y_t\}_{t \geq 1}$ are i.i.d.,
 $Ey_t = 0$, $\text{Var}(y_t) = \sigma^2$
- In this case

$$\mathbb{P}[y_{t_1} \leq z_1, \dots, y_{t_k} \leq z_k] = \prod_{s=1}^k \mathbb{P}[y_{t_s} \leq z_s]$$

- Not interesting for forecasting

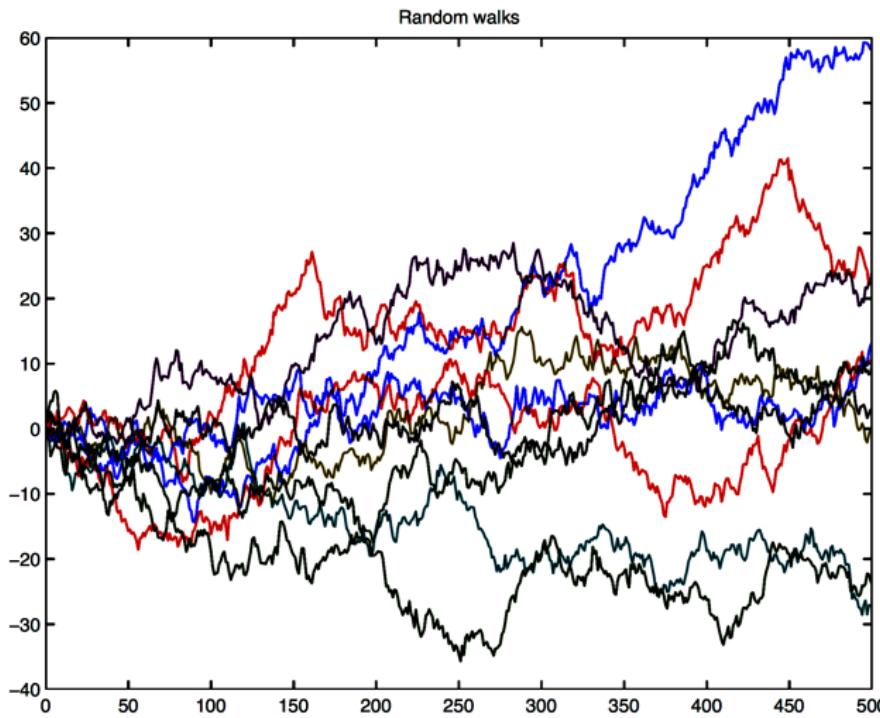
$$\mathbb{P}[y_t \leq z_t | y_1 \leq z_1, \dots, y_{t-1} \leq z_{t-1}] = \mathbb{P}[y_t \leq z_t]$$

Gaussian white noise

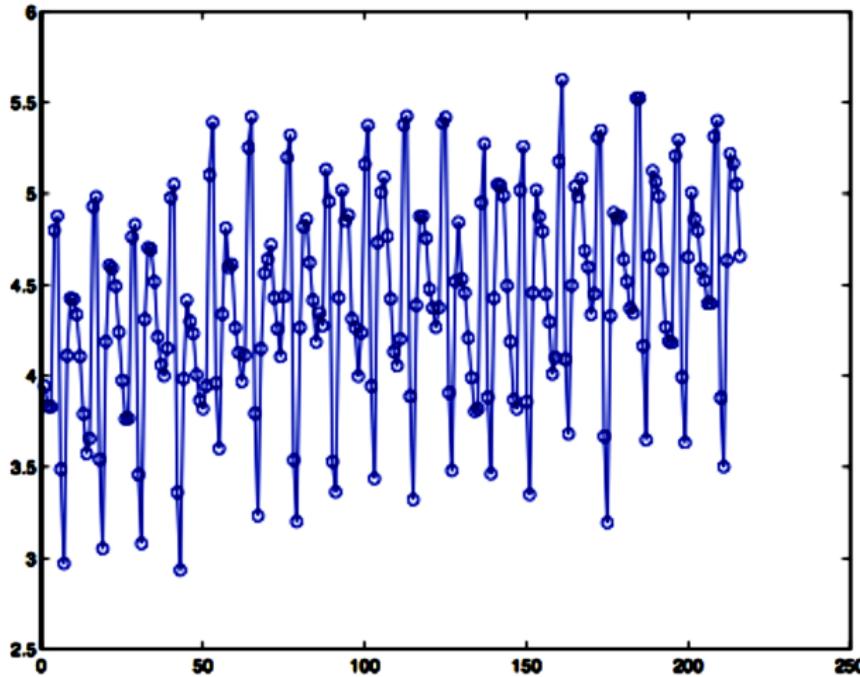


Random walk

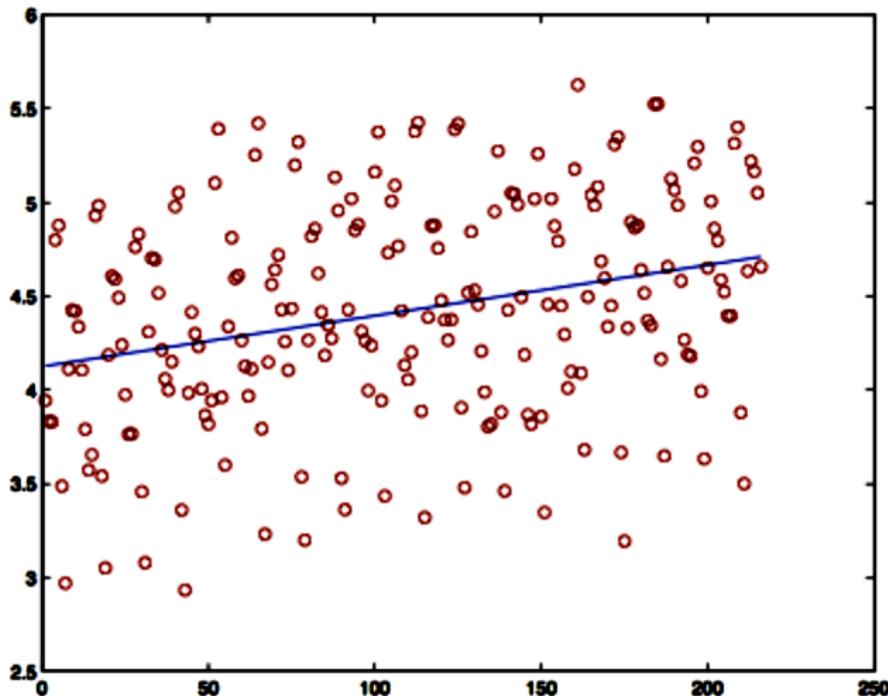
- $y_t = \sum_{i=1}^t \varepsilon_i, \varepsilon_i \sim \mathcal{N}(0, 1)$
- Stock prices on successive days; the path traced by a molecule as it travels in a liquid or a gas; the search path of a foraging animal



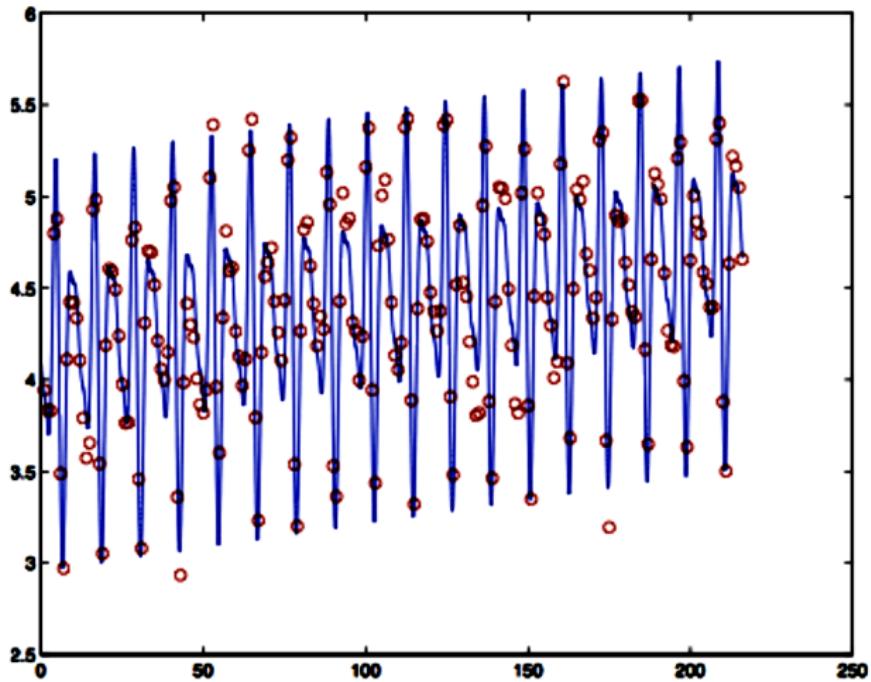
$$y_t = x_t + s_t + \varepsilon_t = \beta_0 + \beta_1 t + \sum_i (\beta_i \cos(\lambda_i t) + \gamma_i \sin(\lambda_i t)) + \varepsilon_t$$



$$y_t = x_t + \varepsilon_t = \beta_0 + \beta_1 t + \varepsilon_t$$

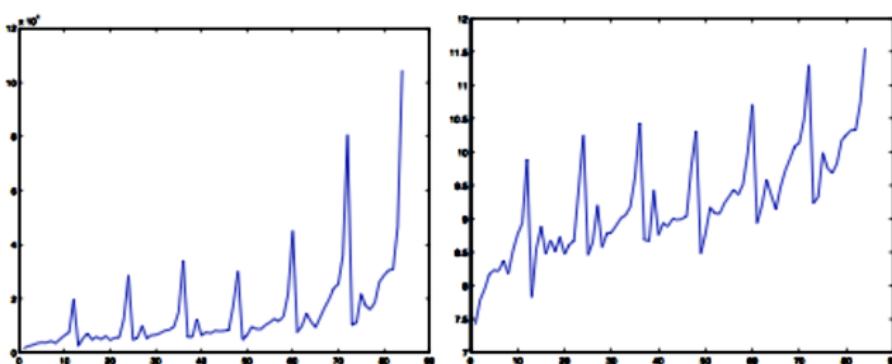


$$y_t = x_t + s_t + \varepsilon_t = \beta_0 + \beta_1 t + \sum_i (\beta_i \cos(\lambda_i t) + \gamma_i \sin(\lambda_i t)) + \varepsilon_t$$

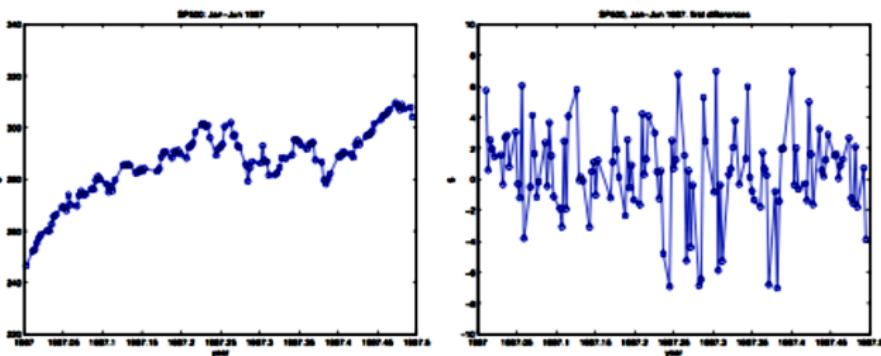


- Plot time series. Look for trends, seasonal components, step changes, outliers
- Transform data so that residuals are **stationary**
 - Estimate and subtract x_t, s_t
 - Differencing
 - Nonlinear transformations (e.g. $\log(\cdot)$, $\sqrt{\cdot}$)
- Fit model to residuals

- Recall: Monthly sales (Makridakis, Wheelwright and Hyndman, 1998), logarithm



- Recall: SP500, difference



- Let us denote $\nabla y_t = y_t - y_{t-1}$
- If $y_t = \sum_{i=0}^k \beta_i t^i + \varepsilon_t$, then

$$\nabla^k y_t = k! \beta_k + \nabla^k \varepsilon_t$$

- If $y_t = x_t + s_t + \varepsilon_t$, s_t has period d ($s_t = s_{t-d}$ for any t), then

$$\nabla_d y_t = x_t - x_{t-d} + \nabla_d \varepsilon_t$$

1 Motivation

2 Time series models

3 Stationarity

4 Auto Regression Moving Average (ARMA)

- $\{y_t\}_{t \geq 1}$ is **strictly stationary** if for all $k, t_1, \dots, t_k, z_1, \dots, z_k$ and h
$$\mathbb{P}[y_{t_1} \leq z_1, \dots, y_{t_k} \leq z_k] = \mathbb{P}[y_{t_1+h} \leq z_1, \dots, y_{t_k+h} \leq z_k],$$
i.e. shifting the time axis does not affect the distribution
- We shall consider **second-order properties** only

- Suppose that $\{y_t\}_{t \geq 1}$ is a time series with $\mathbb{E}[y_t^2] < \infty$
- Its **mean function** is

$$\mu_t = \mathbb{E}[y_t]$$

- Its **autocovariance function** is

$$\begin{aligned}\gamma_y(s, t) &= \text{Cov}(y_s, y_t) \\ &= \mathbb{E}[(y_s - \mu_s)(y_t - \mu_t)]\end{aligned}$$

- We say that $\{y_t\}_{t \geq 1}$ is **(weakly) stationary** if
 1. μ_t is independent of t , and
 2. For each h , $\gamma_y(t+h, t)$ is independent of t
- In that case, we write

$$\gamma_y(h) = \gamma_y(h, 0)$$

- The **autocorrelation function (ACF)** of $\{y_t\}_{t \geq 1}$ is defined as

$$\begin{aligned}\rho_y(h) &= \frac{\gamma_y(h)}{\gamma_y(0)} \\ &= \frac{\text{Cov}(y_{t+h}, y_t)}{\text{Cov}(y_t, y_t)} \\ &= \text{Corr}(y_{t+h}, y_t)\end{aligned}$$

- **Example:** i.i.d. noise, $\mathbb{E}[y_t] = 0$, $\mathbb{E}[y_t^2] = \sigma^2$. We get that

$$\gamma_t(t+h, t) = \begin{cases} \sigma^2, & \text{if } h = 0 \\ 0 & \text{otherwise} \end{cases}$$

- Thus
 1. $\mu_t = 0$ is independent of t
 2. $\gamma_y(t+h, t) = \gamma_y(h, 0)$ for all t

So, $\{y_t\}_{t \geq 1}$ is stationary

- **Example:** Random walk, $y_t = \sum_{i=1}^t \varepsilon_i$ for i.i.d., mean zero $\{\varepsilon_t\}_{t \geq 1}$. We get that

$$\mathbb{E}[y_t] = 0, \mathbb{E}[y_t^2] = t\sigma^2, \text{ and}$$

$$\begin{aligned}\gamma_y(t+h, t) &= \text{Cov}(y_{t+h}, y_t) \\ &= \text{Cov}\left(y_t + \sum_{s=1}^h \varepsilon_{t+s}, y_t\right) \\ &= \text{Cov}(y_t, y_t) = t\sigma^2\end{aligned}$$

- Thus
 1. $\mu_t = 0$ is independent of t
 2. $\gamma_y(t+h, t)$ is not
- So, $\{y_t\}_{t \geq 1}$ is not stationary

1 Motivation

2 Time series models

3 Stationarity

4 Auto Regression Moving Average (ARMA)

- **Example:** MA(1) process (**Moving Average**)

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}, \quad \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

We have $\mathbb{E}[y_t] = 0$ and

$$\begin{aligned}\gamma_y(t+h, t) &= \mathbb{E}(y_{t+h} y_t) \\ &= \mathbb{E}[(\varepsilon_{t+h} + \theta \varepsilon_{t+h-1})(\varepsilon_t + \theta \varepsilon_{t-1})] \\ &= \begin{cases} \sigma^2(1 + \theta^2) & \text{if } h = 0, \\ \sigma^2\theta & \text{if } h = \pm 1, \\ 0 & \text{otherwise} \end{cases}\end{aligned}$$

- Thus $\{y_t\}_{t \geq 1}$ is stationary

- **Example:** AR(1) process (**AutoRegression**)

$$y_t = \varphi y_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

Assume that y_t is stationary (i.e. $|\varphi| < 1$), then

$$\mathbb{E}[y_t] = \varphi \mathbb{E}[y_{t-1}] = 0 \quad (\text{from stationarity})$$

$$\mathbb{E}[y_t^2] = \varphi^2 \mathbb{E}[y_{t-1}^2] + \sigma^2 = \frac{\sigma^2}{1 - \varphi^2} \quad (\text{from stationarity})$$

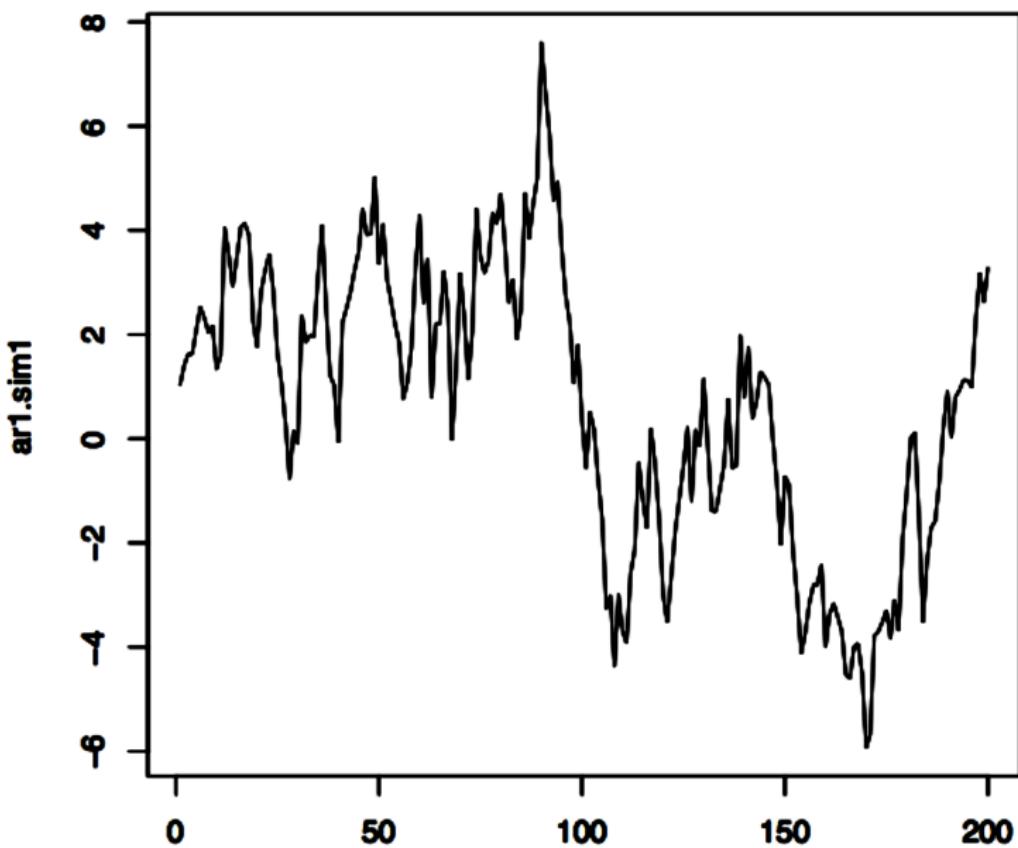
- **Example:** AR(1) process (**AutoRegression**)

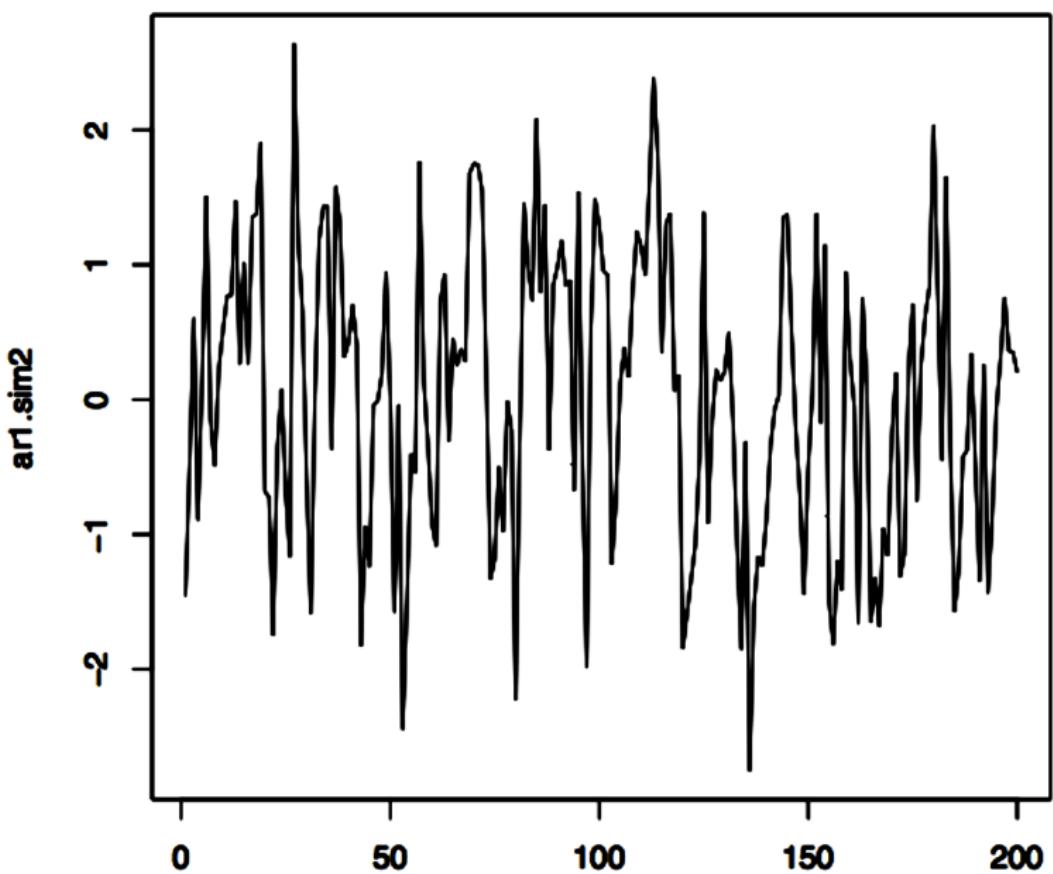
$$y_t = \varphi y_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

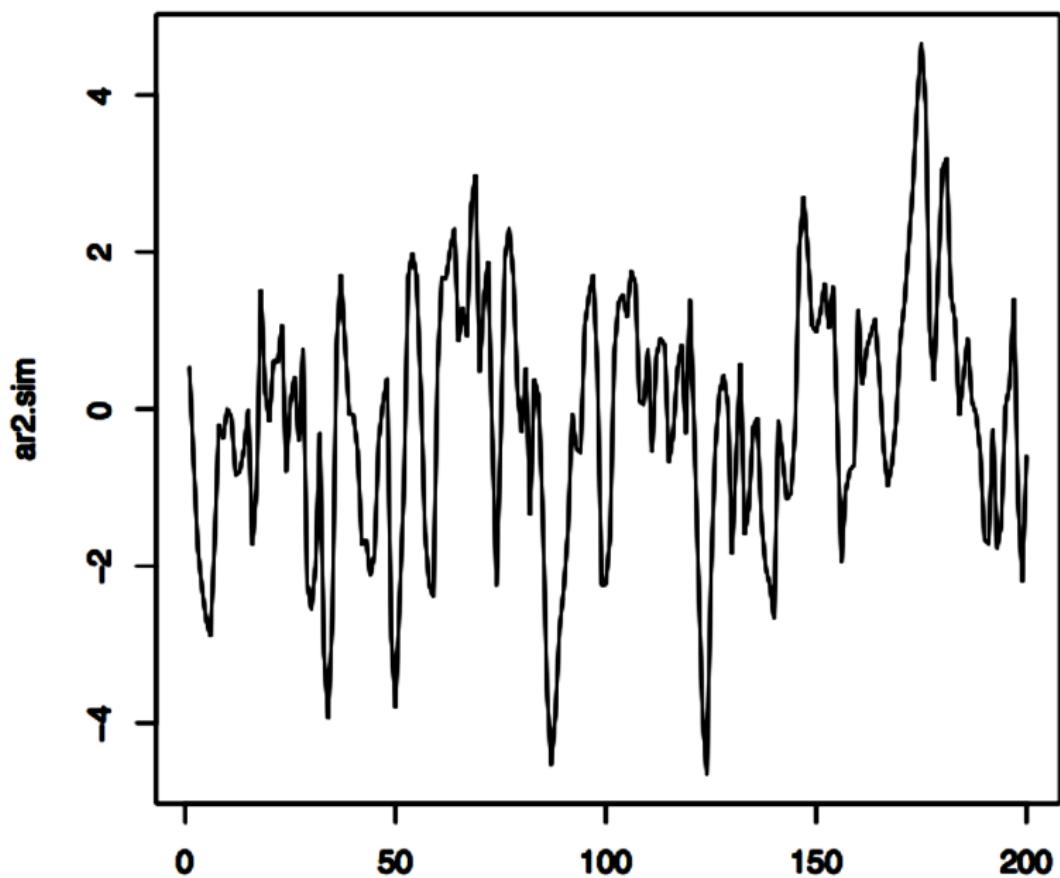
Assume that y_t is stationary (i.e. $|\varphi| < 1$), then

$$\mathbb{E}[y_t] = 0, \quad \mathbb{E}[y_t^2] = \frac{\sigma^2}{1 - \varphi^2}$$

$$\begin{aligned}\gamma_y(h) &= \text{Cov}(\varphi y_{t+h-1} + \varepsilon_{t+h}, y_t) = \\ &= \varphi \text{Cov}(y_{t+h-1}, y_t) \\ &= \varphi \gamma_y(h-1) \\ &= \varphi^{|h|} \gamma_y(0) \quad (\text{check for } h > 0 \text{ and } h < 0) \\ &= \frac{\varphi^{|h|} \sigma^2}{1 - \varphi^2}\end{aligned}$$





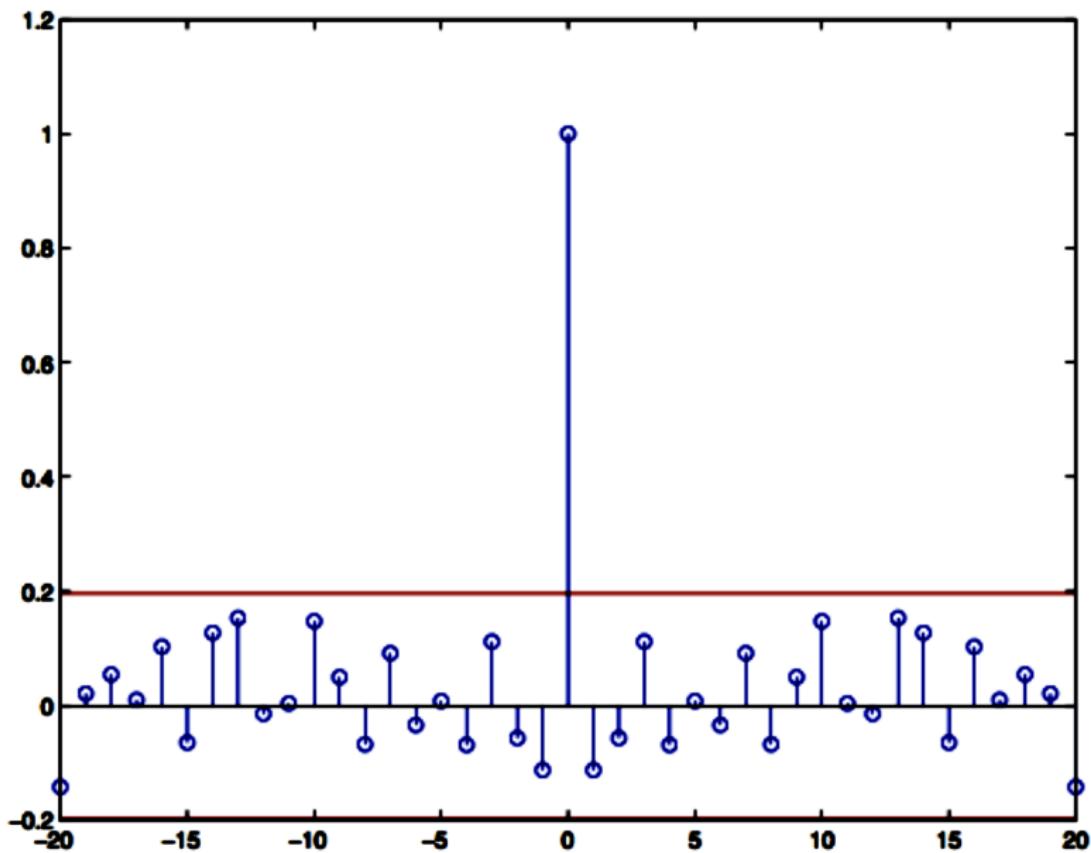


- Sample autocovariance function

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (y_{t+|h|} - \bar{y})(y_t - \bar{y})$$

- \approx sample covariance of $(y_1, y_{h+1}), \dots, (y_{n-h}, y_n)$, except that
 - we normalize by n instead of $n - h$, and
 - we subtract the full sample mean
- We estimate sample variance and obtain sample autocorrelation function

Sample ACF for Gaussian noise



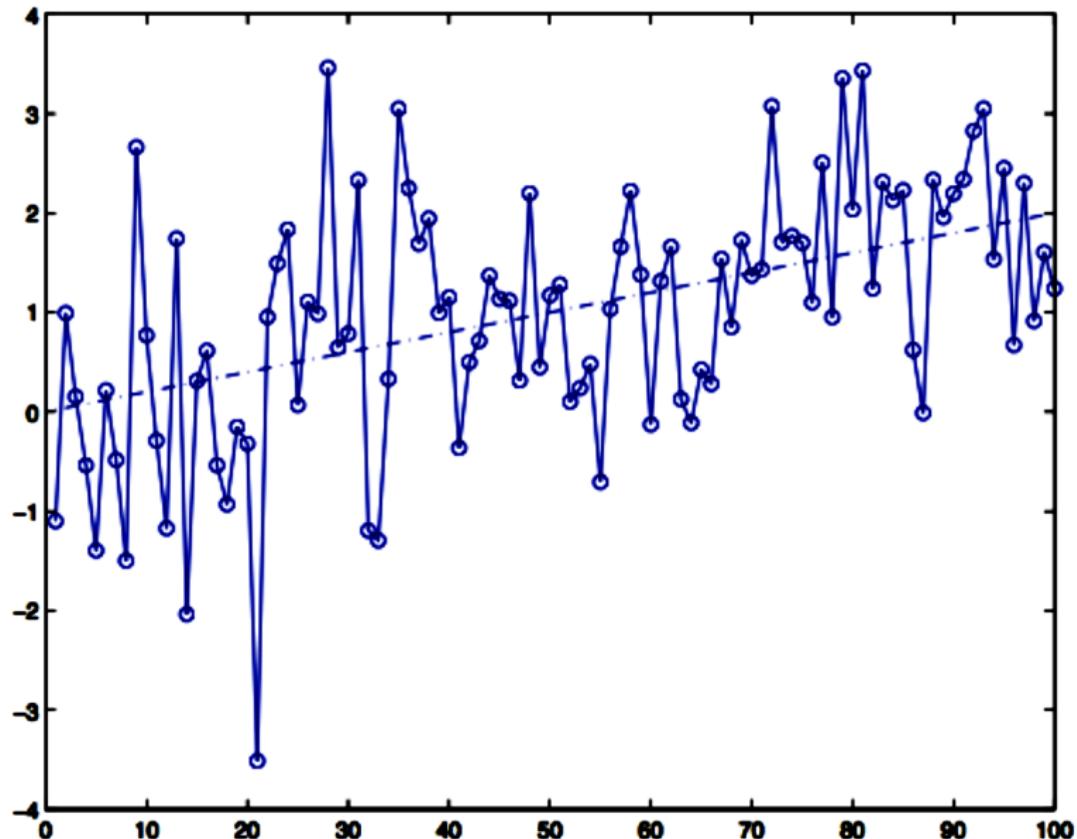
We can recognize the sample autocorrelation functions of many non-white (even non-stationary) time-series

- **Time series:**

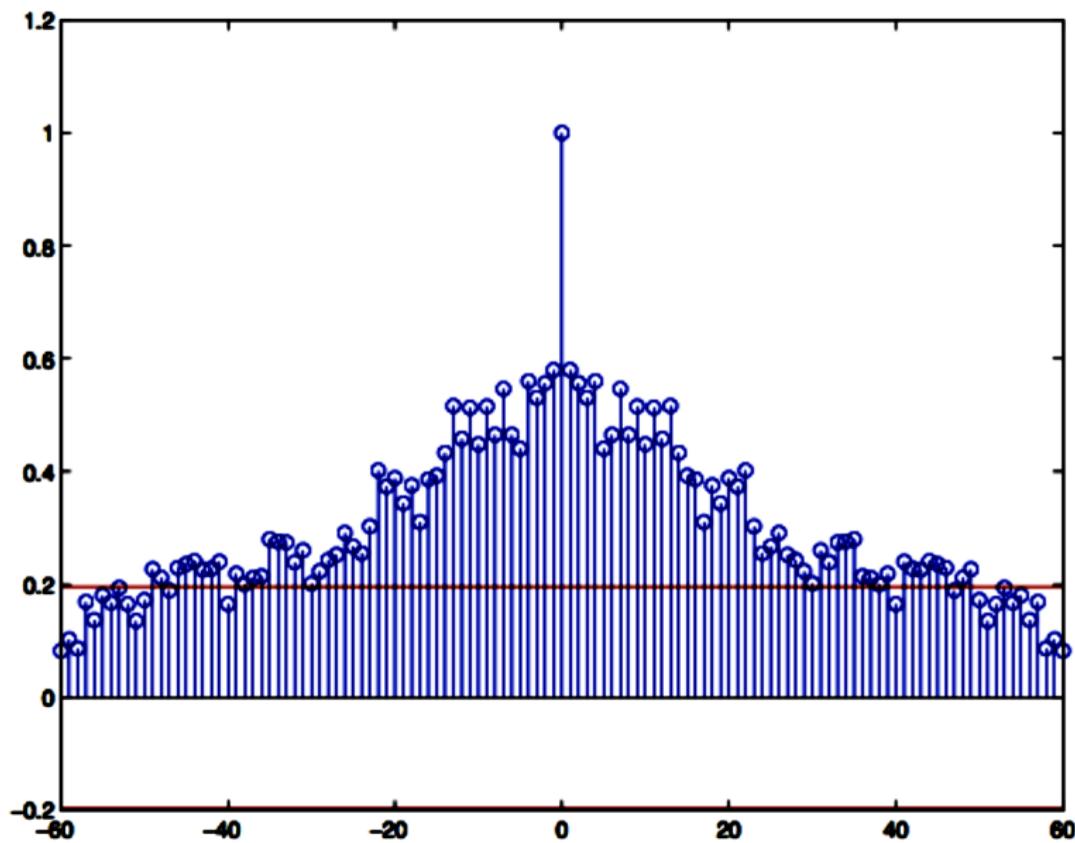
- White
- Trend
- Periodic
- MA(q)
- AR(p)

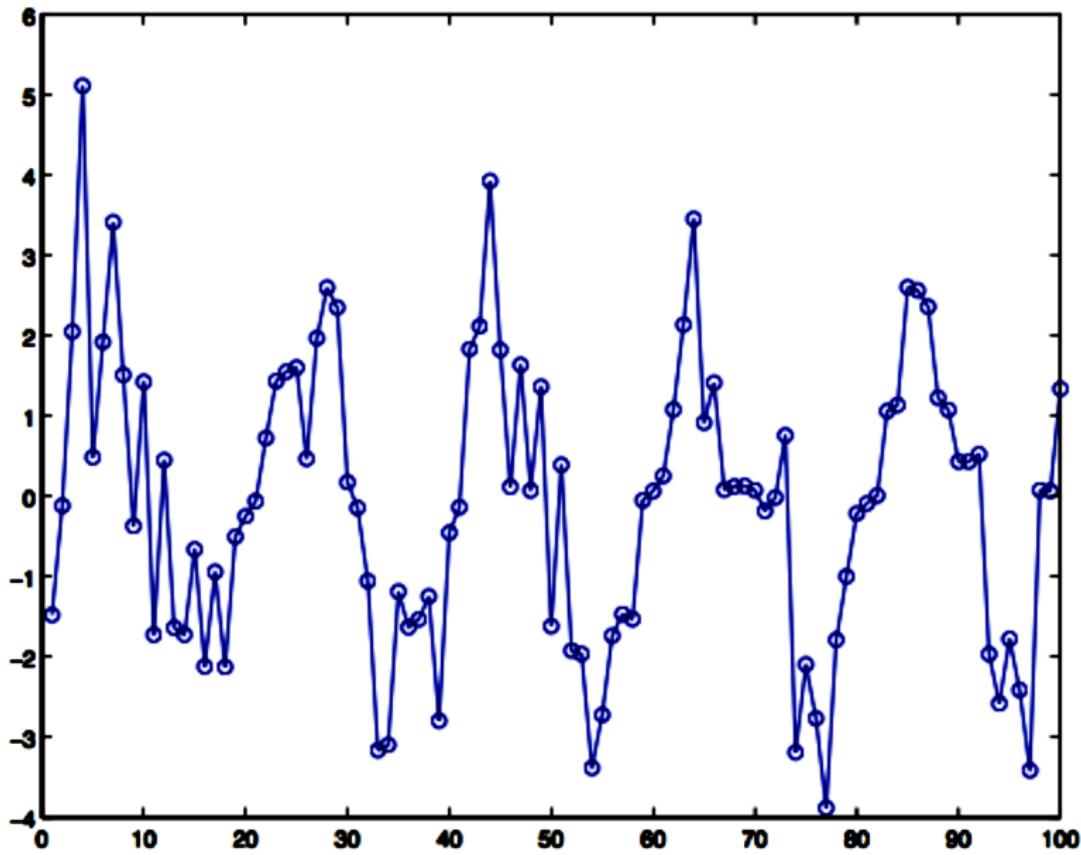
- **Sample ACF:**

- Rightarrow zero*
- Rightarrow Slow decay*
- Rightarrow Periodic*
- Rightarrow Zero for $|h| > q$*
- Rightarrow Decays to zero exponentially*

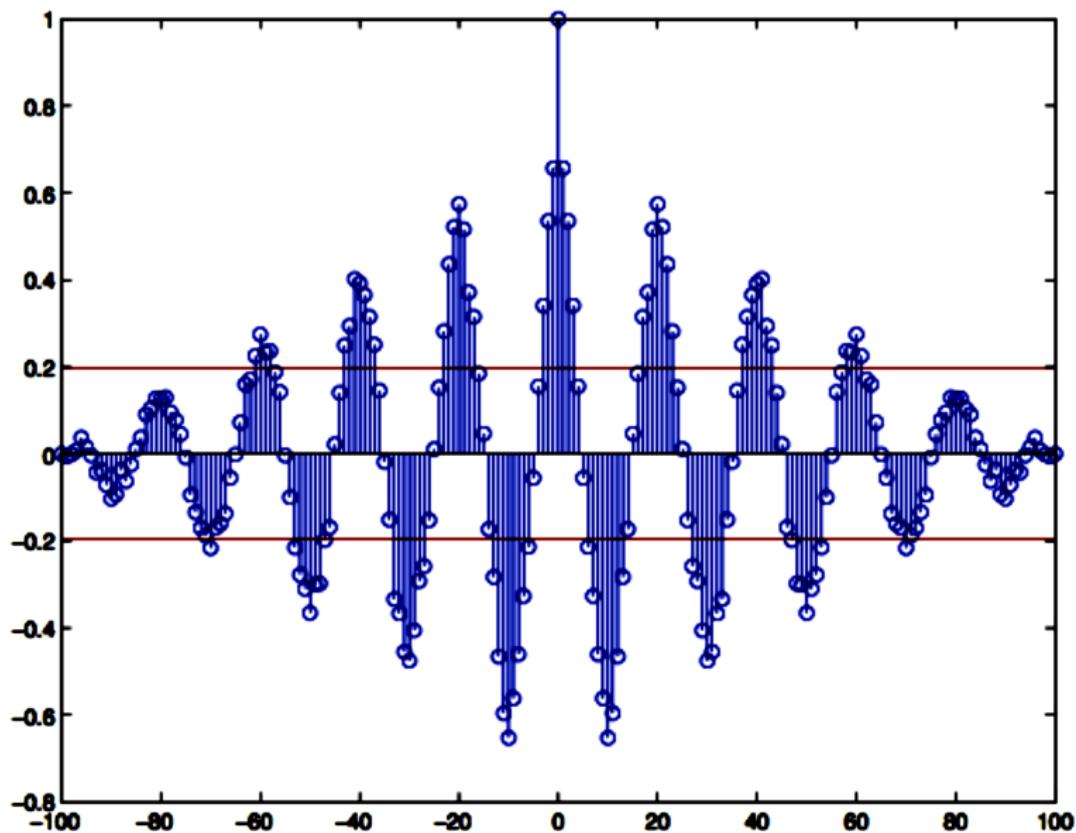


Sample ACF: Trend

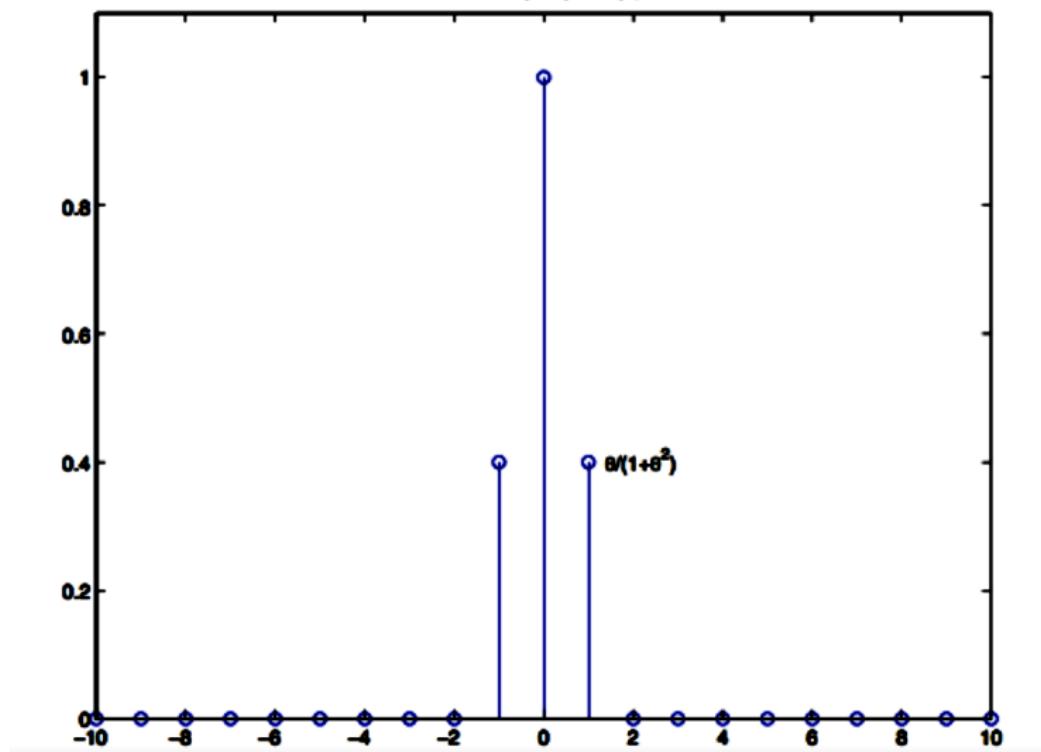




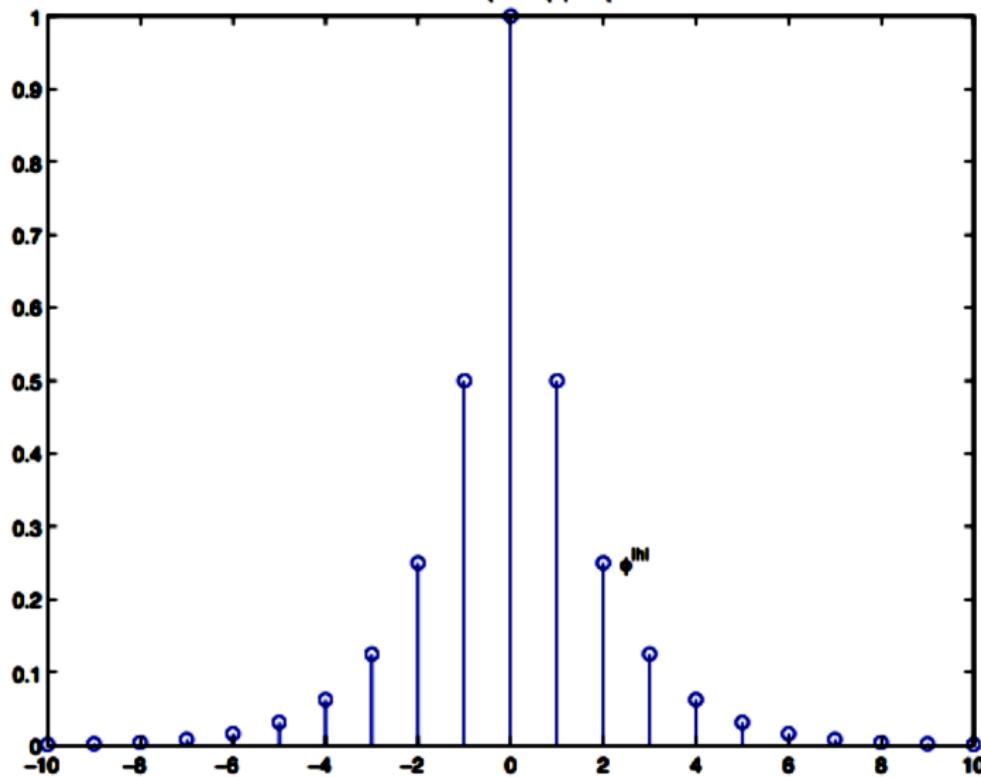
Sample ACF: Trend



$$\text{MA}(1): X_t = Z_t + \theta Z_{t-1}$$



$$\text{AR}(1): X_t = \phi X_{t-1} + Z_t$$



- An **ARMA(p,q) process** $\{y_t\}_{t \geq 1}$ is a stationary process that satisfies

$$y_t - \varphi_1 y_{t-1} - \dots - \varphi_p y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q},$$

where $\{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$

- Given n observations, in case of **AR(p) process** the parameters can be estimated by least-squares

$$\hat{\varphi} = \arg \min_{\varphi} \sum_{t=p+1}^n [y_t - \varphi_1 y_{t-1} - \dots - \varphi_p y_{t-p}]^2$$

- In matrix form for

$$\mathbf{X} = \begin{bmatrix} y_{n-1} & y_{n-2} & \dots & y_{n-p-1} \\ y_{n-2} & y_{n-3} & \dots & y_{n-p-2} \\ \vdots & \vdots & \vdots & \vdots \\ y_p & y_{p-1} & \dots & y_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_n \\ y_{n-1} \\ \vdots \\ y_{p+1} \end{bmatrix}$$

then

$$\hat{\varphi} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

- AR models assume that the relation between past and future is linear
- Nonlinear Auto Regressive (NAR) formulation

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}) + \varepsilon_t,$$

where the missing information is lumped into a noise term ε_t

- We will consider this relationship as a particular instance of a dependence

$$y_t = f(\mathbf{x}_t) + \varepsilon, \quad y_t \in \mathbb{R}^1, \quad \mathbf{x}_t \in \mathbb{R}^p,$$

where

$$\mathbf{x}_t = [y_{t-1}, y_{t-2}, \dots, y_{t-p}]$$

- We train $f(\cdot)$ using ML regression algorithm and a sample (\mathbf{X}, \mathbf{y}) , where

$$\mathbf{X} = \{\mathbf{x}_{p+1}, \mathbf{x}_{p+2}, \dots, \mathbf{x}_T\}$$

and

$$\mathbf{y} = \{y_{p+1}, y_{p+2}, \dots, y_T\}$$