Time-Series Analysis: Stochastic Models

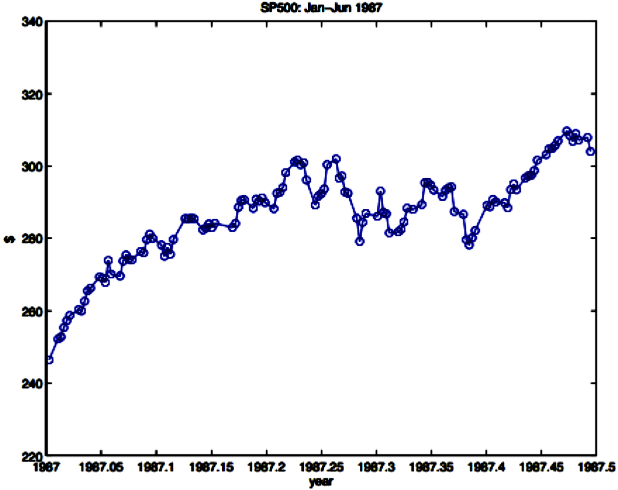
Evgeny Burnaev

Skoltech, Moscow, Russia

Outline

Example

- SP500: Jan-Jun 1987



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

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

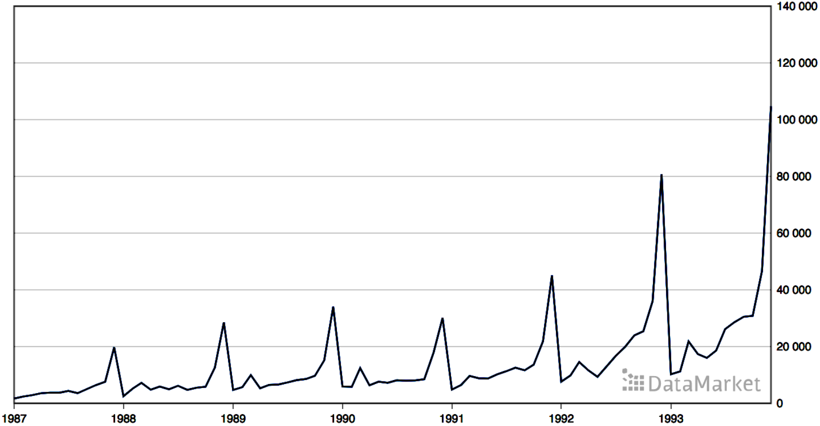
- Compact description of data
- Interpretation
- Forecasting
- Control
- Hypothesis testing
- Simulation

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

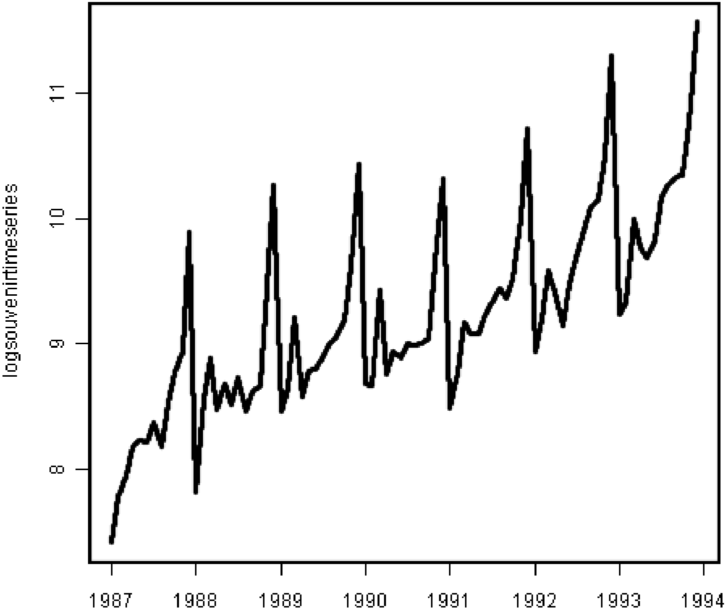
- Compact description of data
- Interpretation
- Forecasting
- Control
- Hypothesis testing
- Simulation

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

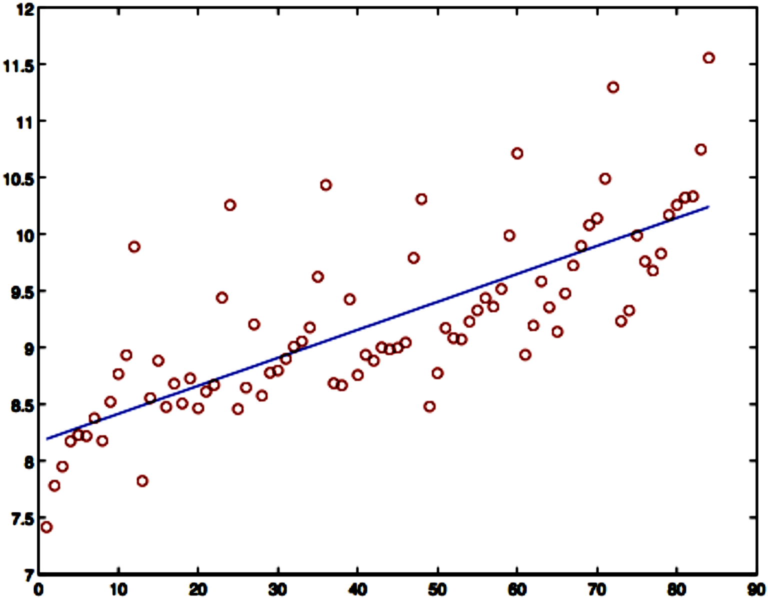
- Monthly sales for a souvenir shop at a beach resort town in Queensland (Makridakis, Wheelwright and Hyndman, 1998)



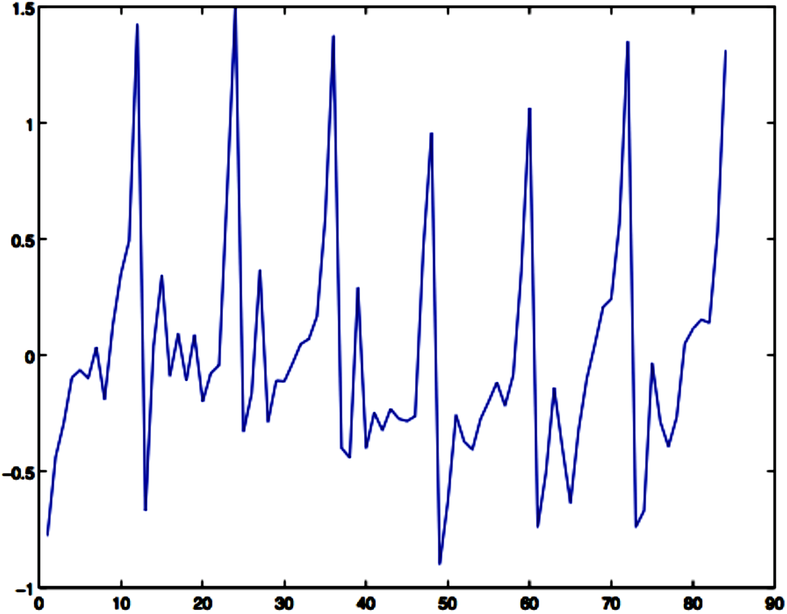
Transformed data

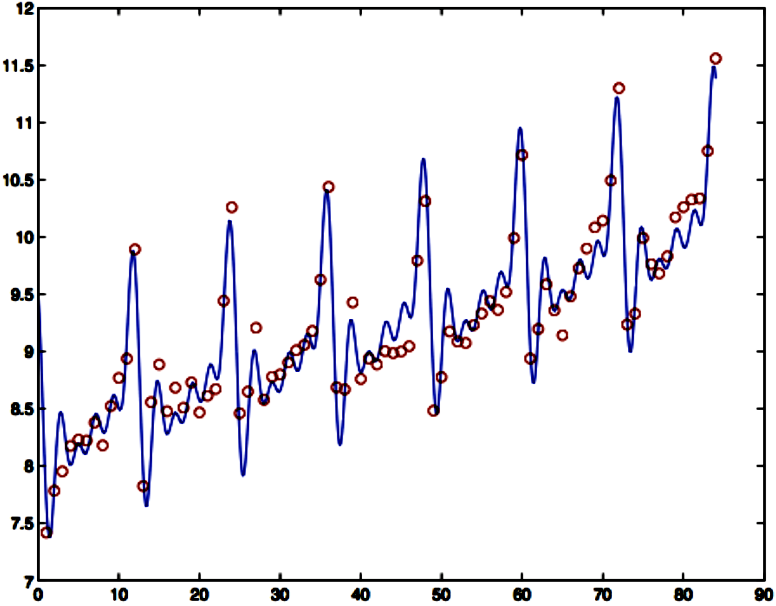


Trend



Residuals




12
11.5
11
10.5
10
9.5
9
8.5
8
7.5
7
0
10
20
30
40
50
60
70
80
90

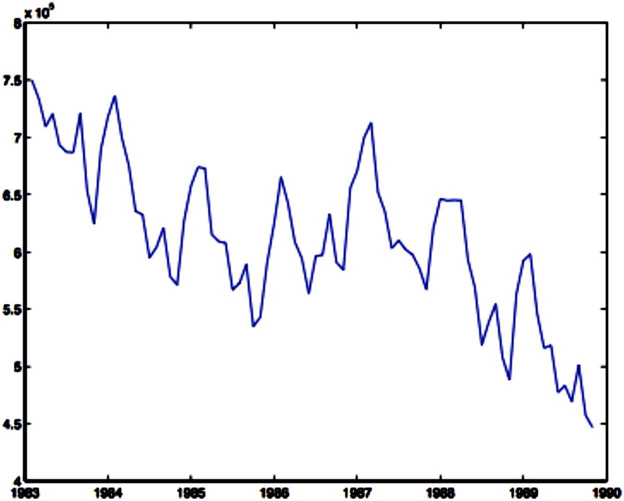
- Compact description of data ⇒
  Example: Classical decomposition $y_t = x_t + s_t + \varepsilon_t$
- Interpretation ⇒ Example: Seasonal adjustment
- Forecasting ⇒ Example: Predict sales
- Control
- Hypothesis testing
- Simulation

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

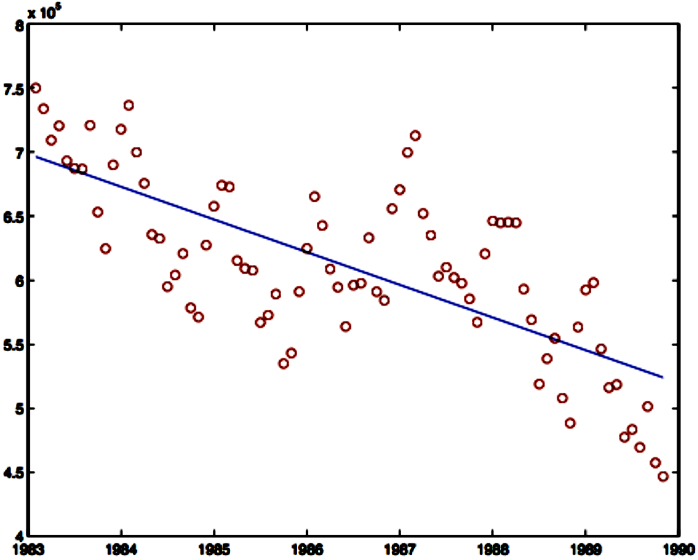
- Compact description of data ⇒
  Example: Classical decomposition $y_t = x_t + s_t + \varepsilon_t$
- Interpretation ⇒ Example: Seasonal adjustment
- Forecasting ⇒ Example: Predict sales
- Control
- Hypothesis testing
- Simulation

- 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

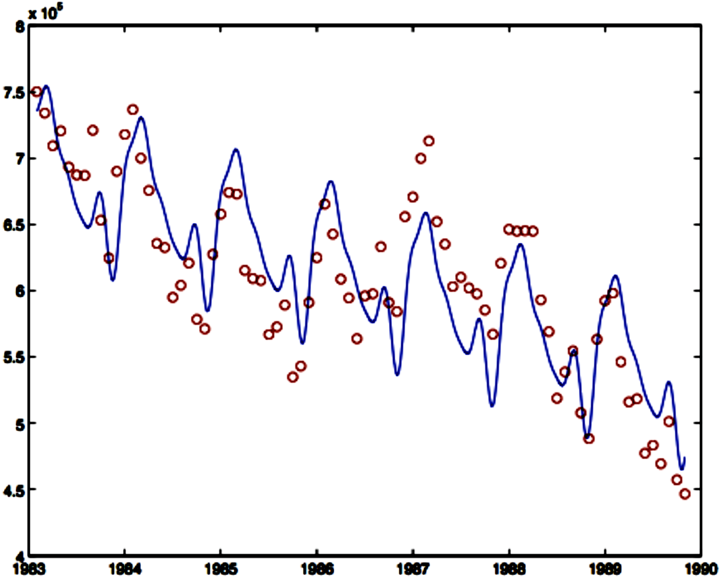
- Monthly number of unemployed people in Australia (Hipel and McLeod, 1994)



Trend



Trend + seasonal variation



- Compact description of data ⇒

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

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

- Compact description of data ⇒

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

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

- 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

- Compact description of data ⇒

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

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

- Compact description of data ⇒

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

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

- 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

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

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

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

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

- 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 retrict our attention to **second-order** properties only:

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

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

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

- Not interesting for forecasting

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

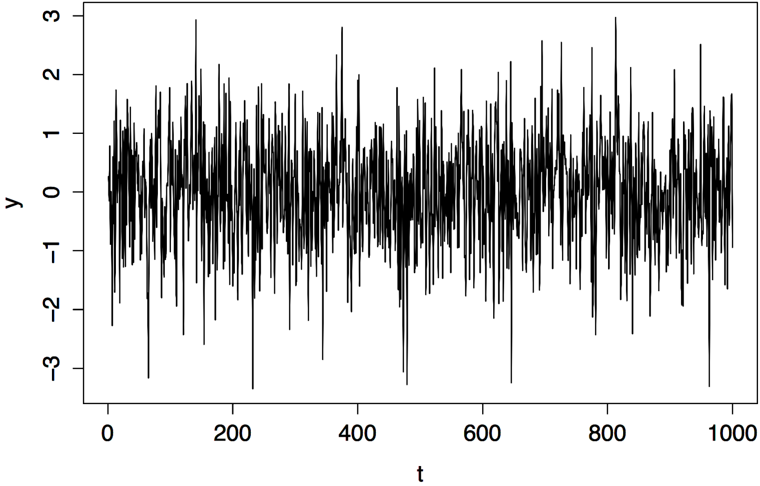
White noise model

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

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

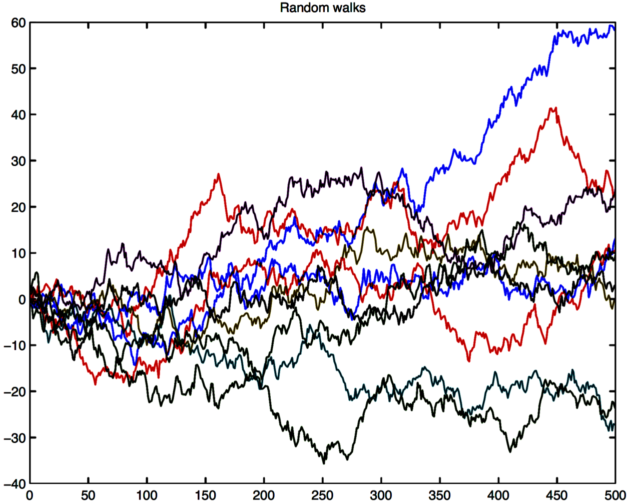
- Not interesting for forecasting

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

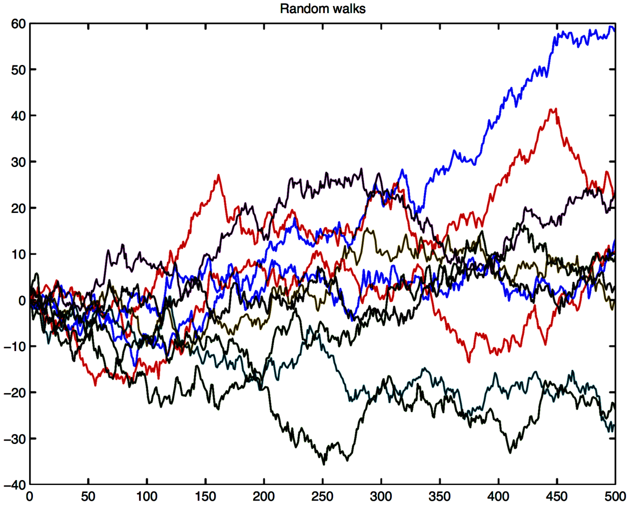

3
2
1
0
−1
−2
−3
y
0
200
400
600
800
1000
t

Random walk

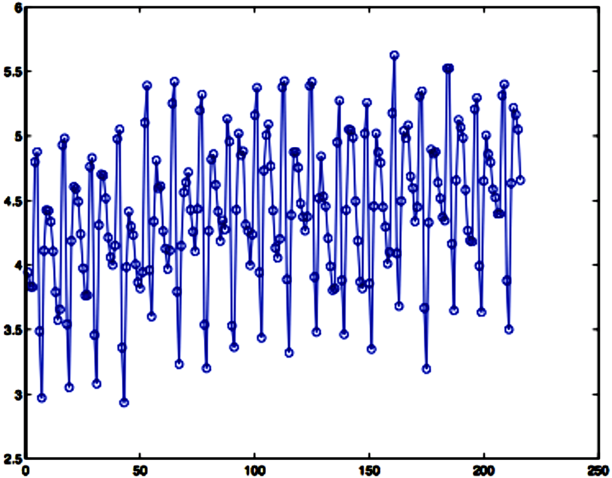
- $y_t = \sum_{i=1}^{t} \varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0, 1)$
- Stock prices on successive days; the path traced by a molecule as it travels in a liquid or a gas; the search path of a foraging animal



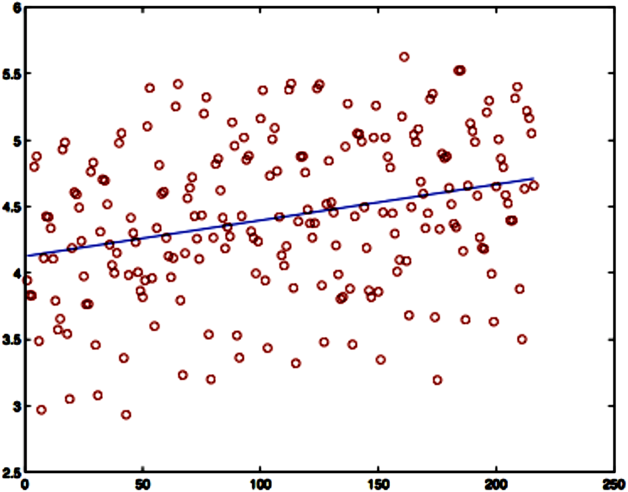
- $y_t = \sum_{i=1}^{t} \varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0, 1)$
- Stock prices on successive days; the path traced by a molecule as it travels in a liquid or a gas; the search path of a foraging animal



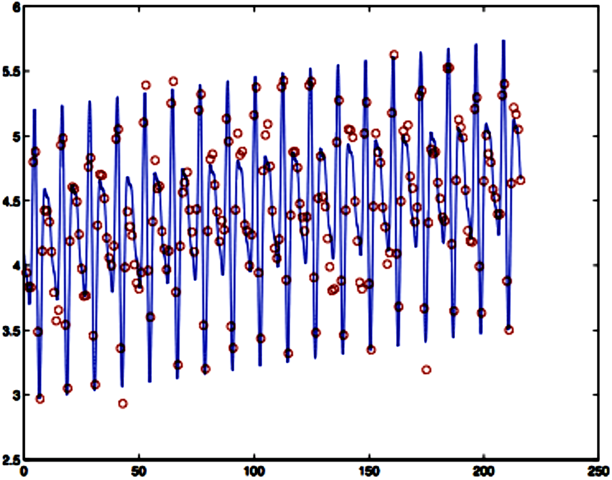
$$y_t = x_t + s_t + \varepsilon_t = \beta_0 + \beta_1 t + \sum_i (\beta_i \cos(\lambda_i t) + \gamma_i \sin(\lambda_i t)) + \varepsilon_t$$



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



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

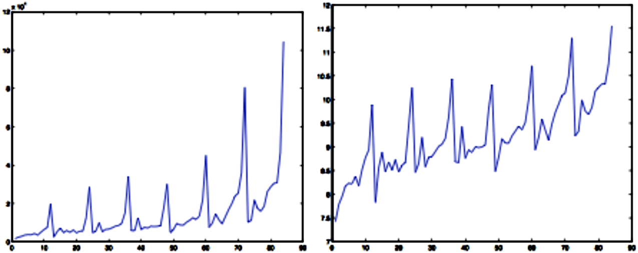


- Plot time series. Look for trends, seasonal components, step changes, outliers
- Transform data so that residuals are **stationary**
  - Estimate and subtract x_t, s_t
  - Differencing
  - Nonlinear transformations (e.g. $\log(\cdot)$, $\sqrt{\cdot}$)
- Fit model to residuals

- 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

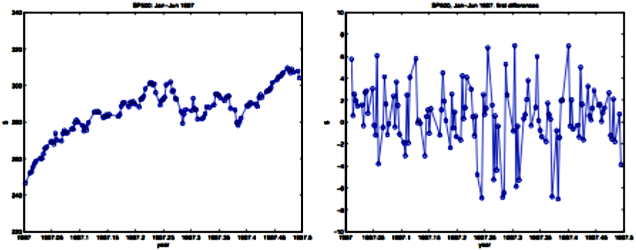
- Plot time series. Look for trends, seasonal components, step changes, outliers
- Transform data so that residuals are **stationary**
  - Estimate and subtract x_t, s_t
  - Differencing
  - Nonlinear transformations (e.g. $\log(\cdot)$, $\sqrt{\cdot}$)
- Fit model to residuals

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



Nonlinear transformation

- 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

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

- $\{y_t\}_{t\geq 1}$ is **strictly stationary** if for all k, $t_1, \ldots, t_k$, $z_1, \ldots, z_k$ and h

$$\mathbb{P}[y_{t_1} \leq z_1, \ldots, y_{t_k} \leq z_k] = \mathbb{P}[y_{t_1+h} \leq z_1, \ldots, y_{t_k+h} \leq z_k],$$

  i.e. shifting the time axis does not affect the distribution

- We shall consider **second-order properties** only

- $\{y_t\}_{t\geq 1}$ is **strictly stationary** if for all k, $t_1, \ldots, t_k$, $z_1, \ldots, z_k$ and h

$$\mathbb{P}[y_{t_1} \leq z_1, \ldots, y_{t_k} \leq z_k] = \mathbb{P}[y_{t_1+h} \leq z_1, \ldots, y_{t_k+h} \leq z_k],$$

  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

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

- 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

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

- 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

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

- 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

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

- 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

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

- **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**: 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

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

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

- 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

- **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

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

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

- 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

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

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

- 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

- **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

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

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

- 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

- **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

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

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

- 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

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

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

- 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

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

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

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

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

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

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

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

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

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

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

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

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

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**)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

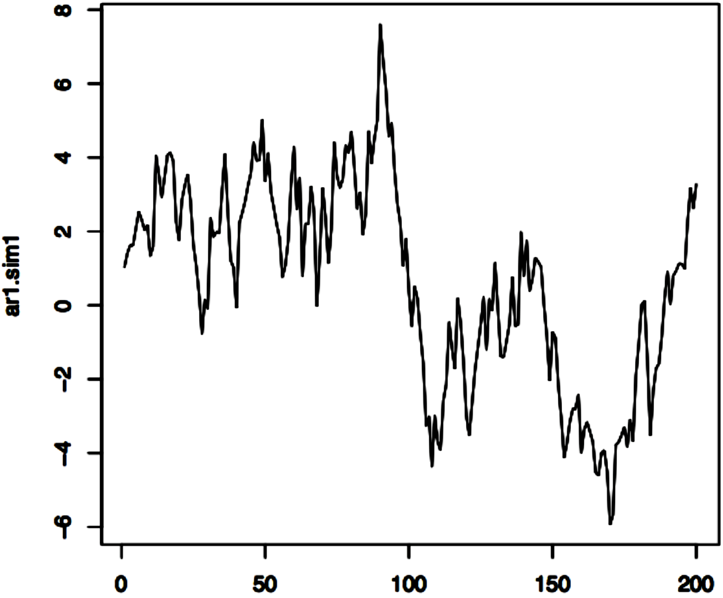
$$\begin{aligned}
\mathbb{E}[y_t] &= 0, \ \mathbb{E}[y_t^2] = \frac{\sigma^2}{1 - \varphi^2} \\
\gamma_y(h) &= \mathrm{Cov}\left(\varphi y_{t+h-1} + \varepsilon_{t+h}, y_t\right) = \\
&= \varphi \mathrm{Cov}(y_{t+h-1}, y_t) \\
&= \varphi \gamma_y(h-1) \\
&= \varphi^{|h|} \gamma_y(0) \ \text{(check for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

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

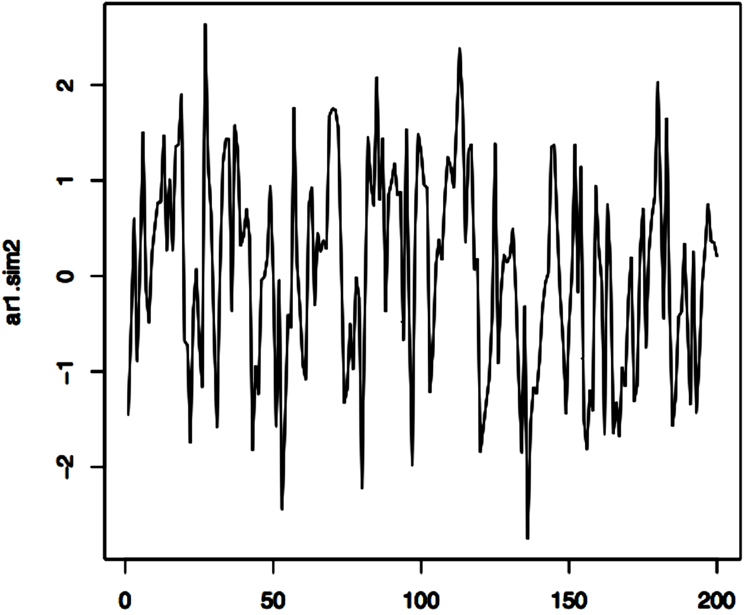
$$y_t = \varphi y_{t-1} + \varepsilon_t, \ \{\varepsilon_t\}_{t \geq 1} \sim WN(0, \sigma^2)$$

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

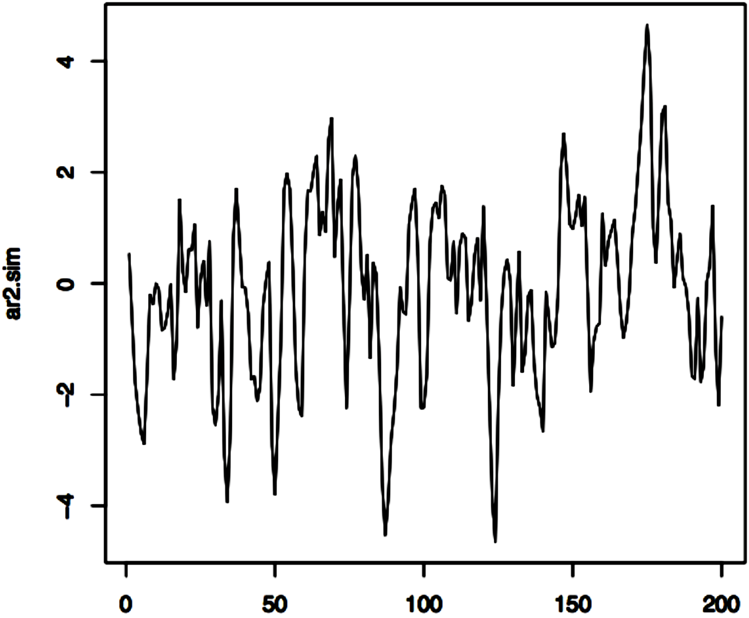
$$\begin{aligned}
\mathbb{E}[y_t] = 0, \ \mathbb{E}[y_t^2] &= \frac{\sigma^2}{1 - \varphi^2} \\
\gamma_y(h) &= \mathrm{Cov}\left(\varphi y_{t+h-1} + \varepsilon_{t+h}, y_t\right) = \\
&= \varphi \mathrm{Cov}(y_{t+h-1}, y_t) \\
&= \varphi \gamma_y(h-1) \\
&= \varphi^{|h|} \gamma_y(0) \ \text{(check for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


ar1.sim1
8
6
4
2
0
-2
-4
-6
0
50
100
150
200

AR(1): $\varphi = 0.5$



AR(2): $\varphi_1 = 0.9$, $\varphi_2 = 0.2$



- Sample autocovariance function

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

- $\approx$ sample covariance of $(y_1, y_{h+1}), \ldots, (y_{n-h}, y_n)$, 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 autocovariance function

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

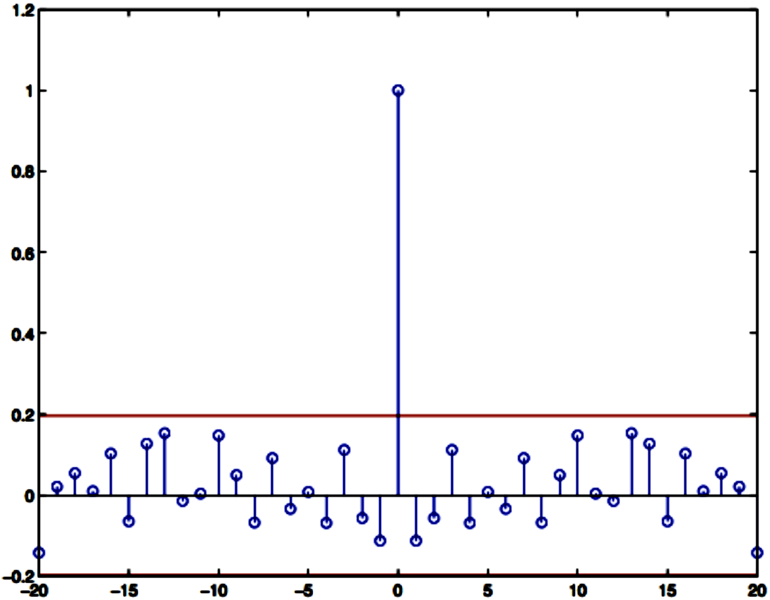
- $\approx$ sample covariance of $(y_1, y_{h+1}), \ldots, (y_{n-h}, y_n)$, except that
  - we normalize by n instead of $n-h$, and
  - we subtract the full sample mean
- We estimate sample variance and obtain sample autocorrelation function

- Sample autocovariance function

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

- $\approx$ sample covariance of $(y_1, y_{h+1}), \ldots, (y_{n-h}, y_n)$, 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

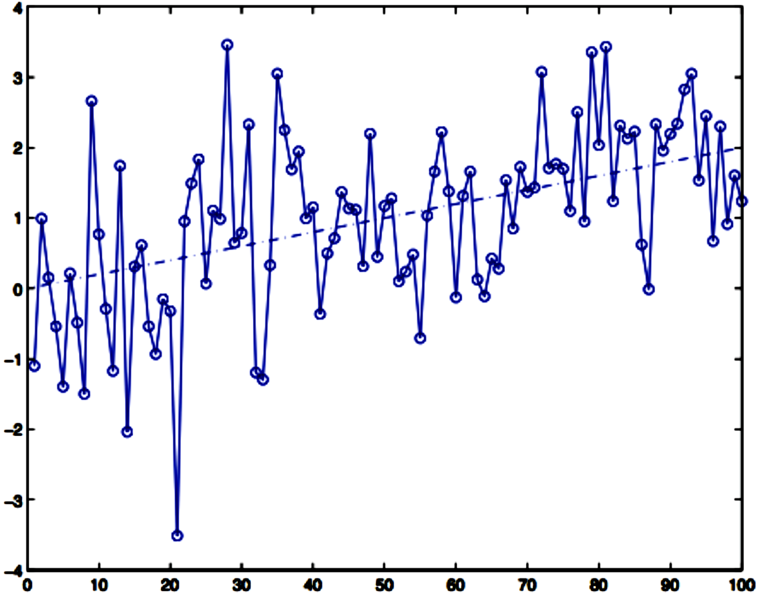
Sample ACF for Gaussian noise



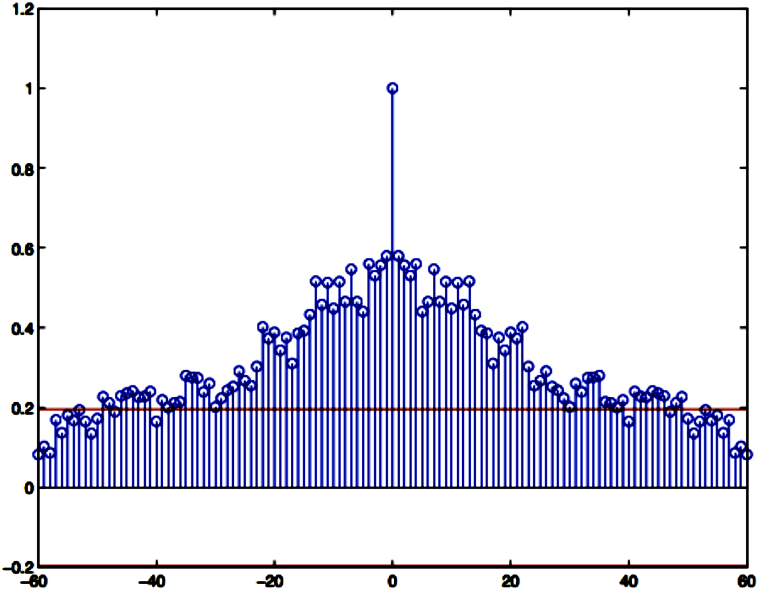
We can recognize the sample autocorrelation functions of many non-white (even non-stationary) time-series

Time series:		**Sample ACF:**
White	$Rightarrow$	zero
Trend	$Rightarrow$	Slow decay
Periodic	$Rightarrow$	Periodic
MA(q)	$Rightarrow$	Zero for $\|h\| > q$
AR(p)	$Rightarrow$	Decays to zero exponentially

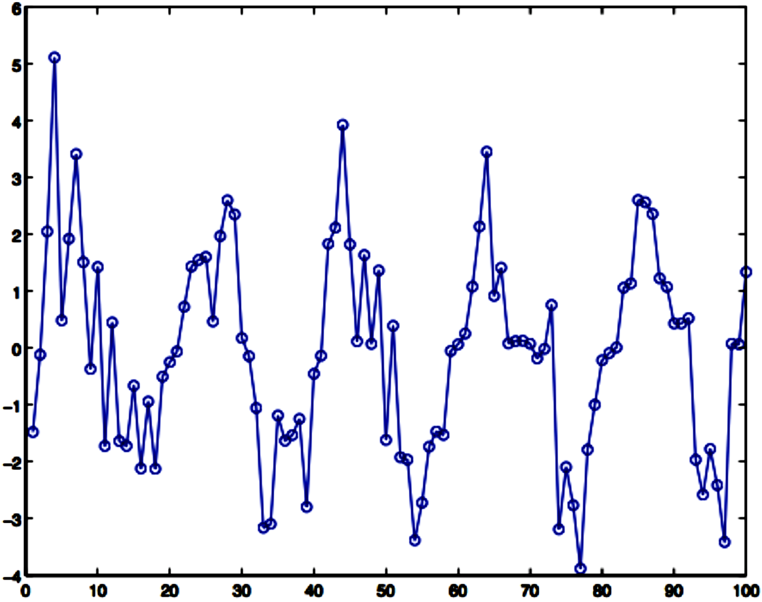
Trend



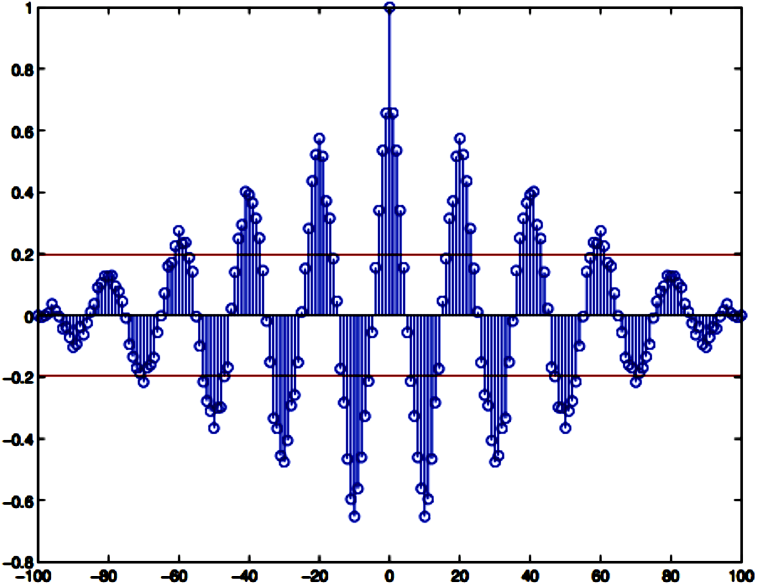
Sample ACF: Trend



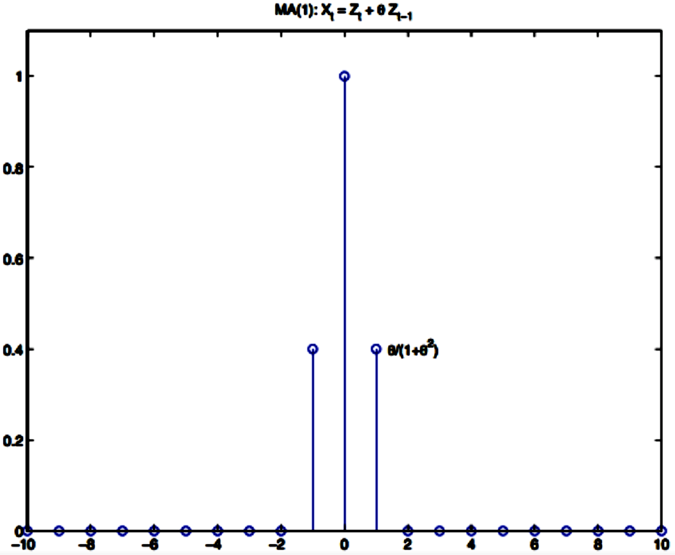
Periodic

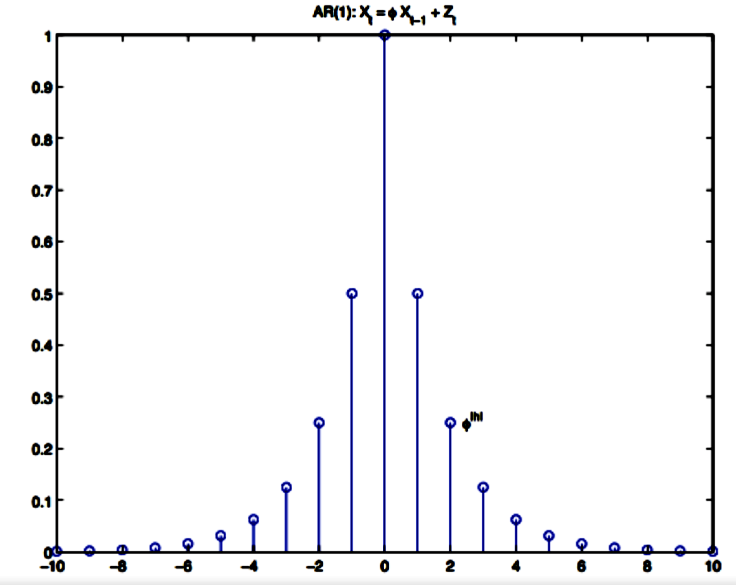


Sample ACF: Trend



ACF: MA(1)




AR(1): $X_t = \phi X_{t-1} + Z_t$
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−10
−8
−6
−4
−2
0
2
4
6
8
10
$\phi^{|h|}$

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

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

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

$$\widehat{\varphi} = \arg\min_{\varphi} \sum_{t=p+1}^{n} \left[y_t - \varphi_1 y_{t-1} - \ldots - \varphi_p y_{t-p}\right]^2$$

- In matrix form for

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

then

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

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

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

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

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

- In matrix form for

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

then

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

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

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

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

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

- In matrix form for

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

then

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

From linear to nonlinear setting

- 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

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

  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

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

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

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

where the missing information is lumped into a noise term ε_t

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

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

where

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

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

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

and

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

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

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

where the missing information is lumped into a noise term ε_t

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

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

where

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

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

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

and

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

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

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

where the missing information is lumped into a noise term ε_t

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

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

where

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

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

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

and

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