Einführung in Web- und Data-Science

Time Series Analysis – ARIMA

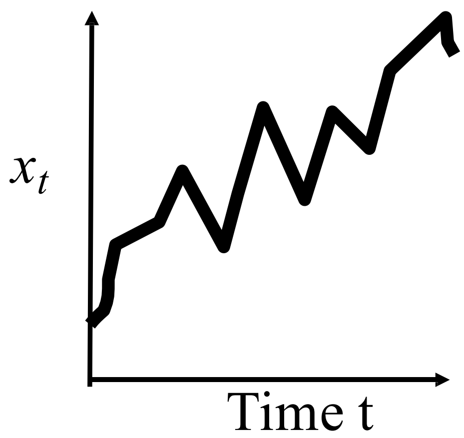
Prof. Dr. Ralf Möller
Universität zu Lübeck
Institut für Informationssysteme

Acknowledgements

- Introduction to Time Series Analysis, Raj Jain, Washington University in Saint Louis http://www.cse.wustl.edu/~jain/cse567-13/

Time Series: Definition

- Time series = stochastic process = sequence of randvars
- A sequence of observations over time



- Examples:
  - Price of a stock over successive days
  - Sizes of video frames
  - Sizes of packets over network
  - Sizes of queries to a database system
  - Number of active virtual machines in a cloud
  - …

Introduction

- Two questions of paramount importance when a data scientist examines time series data:
  - Do the data exhibit a discernible pattern?
  - Can this be exploited to make meaningful forecasts?

Autoregressive Models

- Predict the variable as a linear regression of the immediate past value: $\hat{x}_t = a_0 + a_1 x_{t-1}$
- Here, $\hat{x}_t$ is the best estimate of x_t given the history $\{x_0, x_1, \ldots, x_{t-1}\}$
- Even though we know the complete past history, we assume that x_t can be predicted based on just x_{t-1}.
- Auto-Regressive = Regression on Self
- Error: $e_t = x_t - \hat{x}_t = x_t - a_0 - a_1 x_{t-1}$
- Model: $x_t = a_0 + a_1 x_{t-1} + e_t$
- Best a_0 and a_1 ⇒ minimize the sum of squares of errors

$$\sum_{t=1}^{n} (x_t - \hat{x}_t)^2 = \sum_{t=1}^{n} (x_t - a_0 - a_1 x_{t-1})^2$$

Example 1

- The number of disk accesses for 50 database queries were measured to be: 73, 67, 83, 53, 78, 88, 57, 1, 29, 14, 80, 77, 19, 14, 41, 55, 74, 98, 84, 88, 78, 15, 66, 99, 80, 75, 124, 103, 57, 49, 70, 112, 107, 123, 79, 92, 89, 116, 71, 68, 59, 84, 39, 33, 71, 83, 77, 37, 27, 30.
- For this data:

$$\sum_{t=2}^{50} x_t = 3313 \quad \sum_{t=2}^{50} x_{t-1} = 3356$$

$$\sum_{t=2}^{50} x_t x_{t-1} = 248147 \quad \sum_{t=2}^{50} x_{t-1}^2 = 272102 \quad n = 49$$

$$a_0 = \frac{\sum x_t \sum x_{t-1}^2 - \sum x_{t-1} \sum x_t x_{t-1}}{n \sum x_{t-1}^2 - \left(\sum x_{t-1}\right)^2}$$

$$= \frac{3313 \times 272102 - 3356 \times 248147}{49 \times 272102 - 3356^2} = 33.181$$

Example 1 (ctnd.)

$$a_1 = \frac{n\sum x_t x_{t-1} - \sum x_t \sum x_{t-1}}{n\sum x_{t-1}^2 - \left(\sum x_{t-1}\right)^2}$$

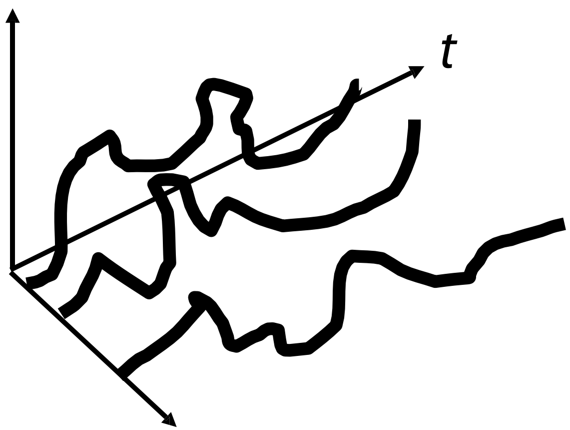
$$= \frac{49 \times 248147 - 3313 \times 3356}{49 \times 272102 - 3356^2} = 0.503$$

SSE = 32995.57

SSE = Sum of squares error

Stationary Process

Each realization of a random process will be different:



- x is function of the realization i (space) and time t: $x(i,t)$
- We can study the distribution of x_t in space
- Each x_t has a distribution, e.g., Normal $f(x_t) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x_t-\mu)^2}{2\sigma^2}}$
- If this same distribution (normal) with the same parameters μ, σ applies to $x_{t+1}, x_{t+2}, \ldots$, we say x_t is stationary

Stationary Process (ctnd.)

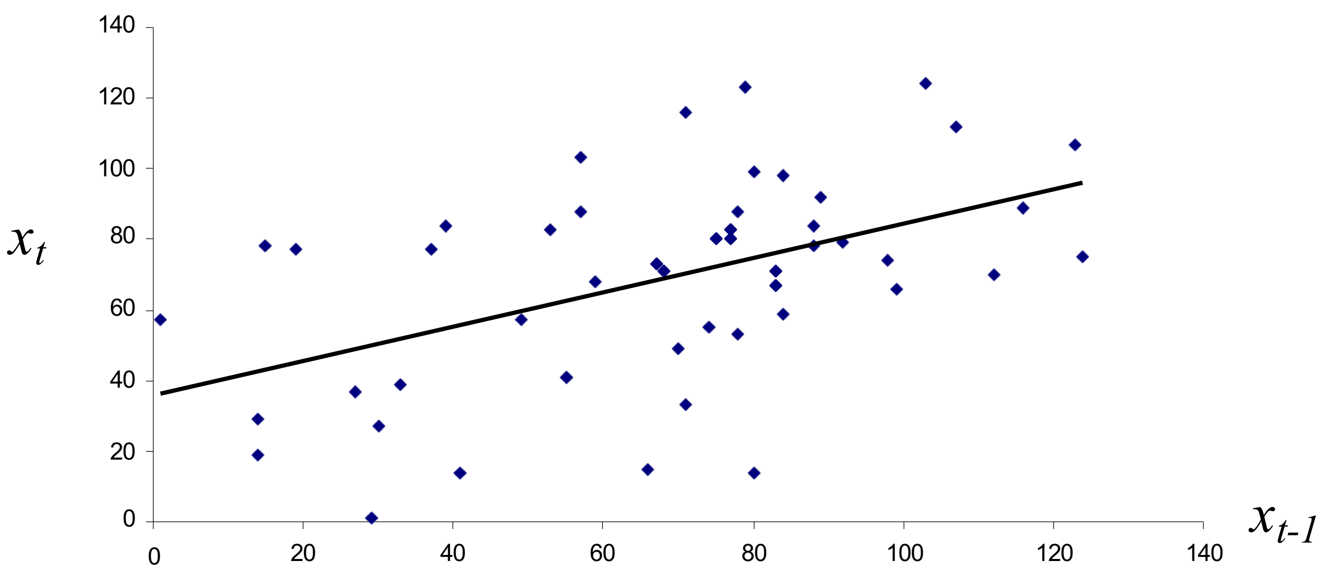
- Stationary = Standing in time
  ⇒ Distribution does not change with time
- Similarly, the joint distribution of x_t and x_{t-k} depends only on k not on t

Assumptions

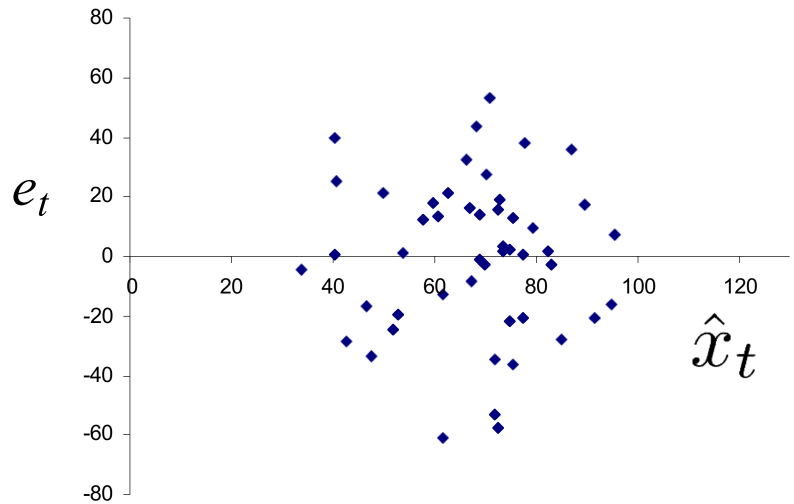
- Linear relationship between successive values
- Normal independent identically distributed (iid) errors:
  - Normal errors
  - Independent errors
- Additive errors
- x_t is a stationary process

Visual Tests

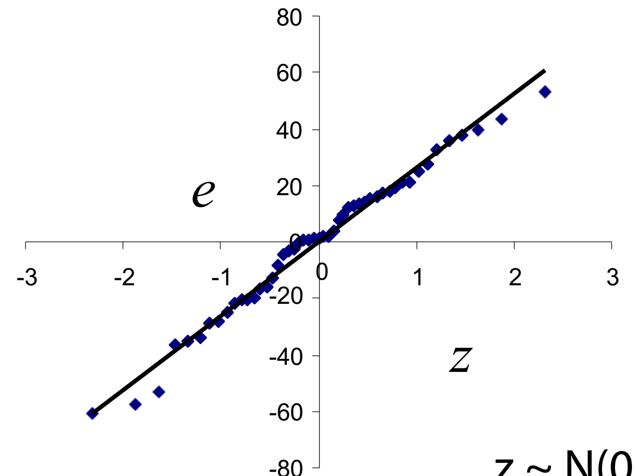
1. x_t vs. x_{t-1} for linearity
2. Errors e_t vs. predicted values $\hat{x}_t$ for additivity
3. Q-Q Plot of errors for Normality
4. Errors e_t vs. t for stationarity
5. Correlations for independence



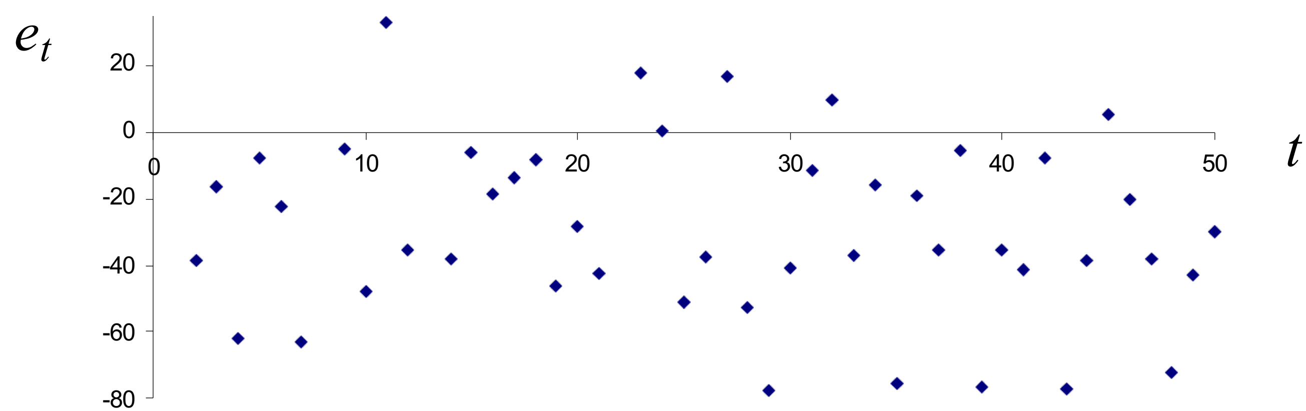
Visual Tests (cntd)



Q–Q plot



z ~ N(0,1)



A Q–Q (quantile-quantile) plot is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other.

AR(p) Model

- x_t is a function of the last p values:

$$x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_p x_{t-p} + e_t$$

- AR(2): $x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + e_t$

- AR(3): $x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + a_3 x_{t-3} + e_t$

Backward Shift Operator

$$\mathrm{B}(x_t) = x_{t-1}$$

Similarly,

$$\mathrm{B}(\mathrm{B}(x_t)) = \mathrm{B}(x_{t-1}) = x_{t-2}$$

Or

$$\mathrm{B}^2 x_t = x_{t-2}$$

$$\mathrm{B}^3 x_t = x_{t-3}$$

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

Using this notation, AR(p) model is

$$\begin{aligned}
x_t - a_1 x_{t-1} - a_2 x_{t-2} - \cdots - a_p x_{t-p} &= a_0 + e_t \\
x_t - a_1 B x_t - a_2 B^2 x_t - \cdots - a_p B^p x_t &= a_0 + e_t \\
(1 - a_1 B - a_2 B^2 - \cdots - a_p B^p) x_t &= a_0 + e_t \\
\phi_p(B) x_t &= a_0 + e_t
\end{aligned}$$

Here, $\phi_p(B)$ is a polynomial of degree p

AR(p) Parameter Estimation

$$x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + e_t$$

- The coefficients a_i can be estimated by minimizing SSE using Multiple Linear Regression

$$\text{SSE} = \sum e_t^2 = \sum_{t=3}^{n} (x_t - a_0 - a_1 x_{t-1} - a_2 x_{t-2})^2$$

- Optimal a_0, a_1, and a_2 ⇒ Minimize SSE
⇒Set the first differential to zero:

$$\begin{aligned}
\frac{d}{da_0}\text{SSE} &= \sum_{t=3}^{n} -2(x_t - a_0 - a_1 x_{t-1} - a_2 x_{t-2}) = 0 \\
\frac{d}{da_1}\text{SSE} &= \sum_{t=3}^{n} -2x_{t-1}(x_t - a_0 - a_1 x_{t-1} - a_2 x_{t-2}) = 0 \\
\frac{d}{da_2}\text{SSE} &= \sum_{t=3}^{n} -2x_{t-2}(x_t - a_0 - a_1 x_{t-1} - a_2 x_{t-2}) = 0
\end{aligned}$$

AR(p) Parameter Estimation (Cont)

The equations can be written as:

$$\begin{bmatrix} n-2 & \sum x_{t-1} & \sum x_{t-2} \\ \sum x_{t-1} & \sum x_{t-1}^2 & \sum x_{t-1}x_{t-2} \\ \sum x_{t-2} & \sum x_{t-1}x_{t-2} & \sum x_{t-2}^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum x_t \\ \sum x_t x_{t-1} \\ \sum x_t x_{t-2} \end{bmatrix}$$

Note: All sums are for t=3 to n. n-2 terms

Multiplying by the inverse of the first matrix, we get:

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} n-2 & \sum x_{t-1} & \sum x_{t-2} \\ \sum x_{t-1} & \sum x_{t-1}^2 & \sum x_{t-1}x_{t-2} \\ \sum x_{t-2} & \sum x_{t-1}x_{t-2} & \sum x_{t-2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_t \\ \sum x_t x_{t-1} \\ \sum x_t x_{t-2} \end{bmatrix}$$

Example 2

Consider the data of Example 1 and fit an AR(2) model:

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} n-2 & \sum x_{t-1} & \sum x_{t-2} \\ \sum x_{t-1} & \sum x_{t-1}^2 & \sum x_{t-1}x_{t-2} \\ \sum x_{t-2} & \sum x_{t-1}x_{t-2} & \sum x_{t-2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_t \\ \sum x_t x_{t-1} \\ \sum x_t x_{t-2} \end{bmatrix}$$

$$= \begin{bmatrix} 48 & 3283 & 3329 \\ 3283 & 266773 & 247337 \\ 3329 & 247337 & 271373 \end{bmatrix}^{-1} \begin{bmatrix} 3246 \\ 243256 \\ 229360 \end{bmatrix} = \begin{bmatrix} 39.979 \\ 0.587 \\ -0.180 \end{bmatrix}$$

SSE= 31969.99

(3% lower than 32995.57 for AR(1) model)

Summary AR(p)

- Assumptions:
  - Linear relationship between x_t and $\{x_{t-1}, \ldots, x_{t-p}\}$
  - Normal iid errors:
    - Normal errors
    - Independent errors
  - Additive errors
  - x_t is stationary

Autocorrelation

- Covariance of x_t and $x_{t\text{-}k}$ = Auto-covariance at lag k

  Autocovariance of x_t at lag $k = \text{Cov}[x_t, x_{t-k}] = E[(x_t - \mu)(x_{t-k} - \mu)]$
- For a stationary series, the statistical characteristics do not depend on time t
- Therefore, the autocovariance depends only on lag k and not on time t
- Similarly,

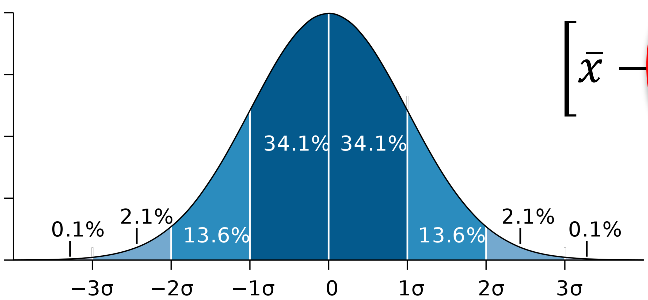
$$
\begin{aligned}
\text{Autocorrelation of } x_t \text{ at lag } k \;\; r_k &= \frac{\text{Autocovariance of } x_t \text{ at lag } k}{\text{Variance of } x_t} \\
&= \frac{\text{Cov}[x_t, x_{t-k}]}{\text{Var}[x_t]} \\
&= \frac{E[(x_t - \mu)(x_{t-k} - \mu)]}{E[(x_t - \mu)^2]}
\end{aligned}
$$

Autocorrelation (cntd.)

- Autocorrelation is dimensionless and is easier to interpret than autocovariance
- It can be shown that autocorrelations are $N(0,1/n)$ distributed, where n is the number of observations in the series
- Therefore, their 95% confidence interval is $\mp 1.96/\sqrt{n}$
  This is generally approximated as $\mp 2/\sqrt{n}$

Standard error = $1/\sqrt{n}$

$1-\alpha$	α	$z_{1-\frac{\alpha}{2}}$	$z_{1-\alpha}$
0,95	0,05	1,96	1,64
0,99	0,01	2,58	2,33
0,999	0,001	3,29	3,09

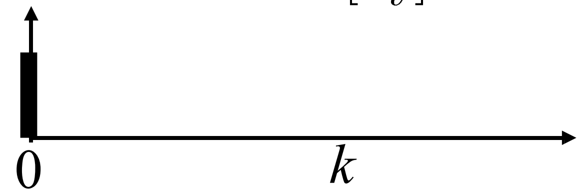


$$\left[\bar{x} - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}, \bar{x} + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\right]$$

White Noise

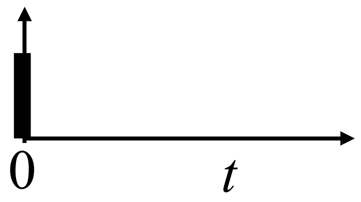
- Errors e_t are normal independent and identically distributed (IID) with zero mean and variance σ^2
- Such IID sequences are called "**white noise**" sequences.
- Properties:

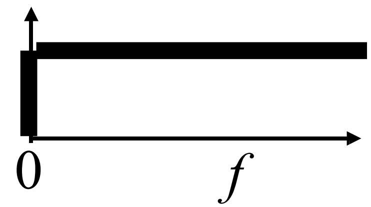
$$
\begin{aligned}
E[e_t] &= 0 \qquad \forall t \\
\mathrm{Var}[e_t] &= E[e_t^2] = \sigma^2 \qquad \forall t \\
\mathrm{Cov}[e_t, e_{t-k}] &= E[e_t e_{t-k}] = \begin{cases} \sigma^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \\
\mathrm{Cor}[e_t, e_{t-k}] &= \frac{E[e_t e_{t-k}]}{E[e_t^2]} = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}
\end{aligned}
$$



White Noise (cntd.)

- The autocorrelation function of a white noise sequence is a spike (δ- function) at k=0
- The Laplace transform of a δ- function is a constant. So in frequency domain white noise has a flat frequency spectrum





- It was incorrectly assumed that white light has no color and, therefore, has a flat frequency spectrum and so random noise with flat frequency spectrum was called white noise
- Ref: http://en.wikipedia.org/wiki/Colors_of_noise

Example 3

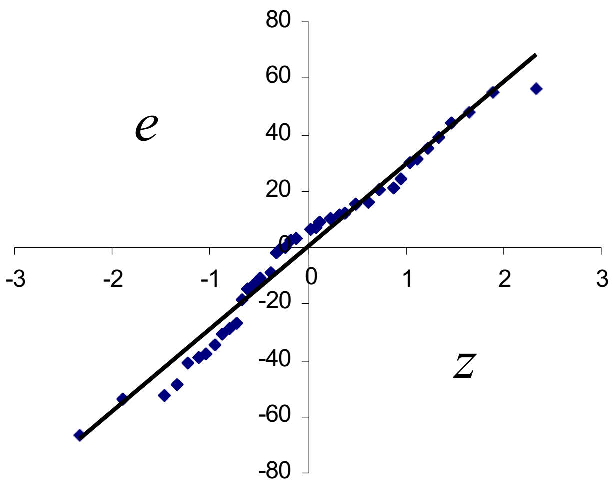
- Consider the data of Example 1. The AR(0) model is

$$x_t = a_0 + e_t$$

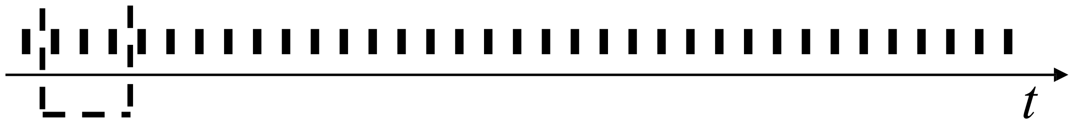
$$\sum x_t = n a_0 + \sum e_t$$

$$a_0 = \frac{1}{n} \sum x_t = 67.72$$

- SSE = 43702.08



Moving Average (MA) Models



- Moving Average of order 1: MA(1)
  $$x_t - a_0 = e_t + b_1 e_{t-1}$$
- Moving Average of order 2: MA(2)
  $$x_t - a_0 = e_t + b_1 e_{t-1} + b_2 e_{t-2}$$
- Moving Average of order q: MA(q)
  $$x_t - a_0 = e_t + b_1 e_{t-1} + b_2 e_{t-2} + \cdots + b_q e_{t-q}$$
- Moving Average of order 0: MA(0) (Note: This is also AR(0))
  $$x_t - a_0 = e_t$$
  x_t-a_0 is a white noise. a_0 is the mean of the time series

MA Models (cntd.)

- Using the backward shift operator B, MA(q):

$$\begin{aligned} x_t - a_0 &= e_t + b_1 B e_t + b_2 B^2 e_t + \cdots + b_q B^q e_t \\ &= (1 + b_1 B + b_2 B^2 + \cdots + b_q B^q) e_t \\ &= \psi_q(B) e_t \end{aligned}$$

- Here, ψ_q is a polynomial of order q

Determining MA Parameters

- Consider MA(1):

$$x_t - a_0 = e_t + b_1 e_{t-1}$$

- The parameters a_0 and b_1 cannot be estimated using standard regression formulas since we do not know errors. The errors depend on the parameters
- So the only way to find optimal a_0 and b_1 is by iteration

  ⇒ Start with some suitable values and change a_0 and b_1 until SSE is minimized and average of errors is zero

Example 4

- Consider the data of Example 1
- For these data: $\bar{x} = \frac{1}{50}\sum_{t=1}^{50} x_t = 67.72$
- We start with $a_0 = 67.72$, $b_1=0.4$
  Assuming $e_0=0$, compute all the errors and SSE
  $$\bar{e} = \frac{1}{50}\sum_{t=1}^{50} e_t = -0.152$$ and SSE = 33542.65
- We then adjust a_0 and b_1 until SSE is minimized and mean error is close to zero

Example 4 (ctnd.)

- The steps are: Starting with $a_0 = \bar{x}$ and b_1=0.4, 0.5, 0.6

a_0	b_1	$\bar{e}$	SSE	Decision
67.72	0.4	-0.15	33542.65	
67.72	0.5	-0.17	33274.55	
67.72	0.6	-0.18	34616.85	0.5 is the lowest. Try 0.45 and 0.55
67.72	0.55	-0.18	33686.88	
67.72	0.45	-0.16	33253.62	Lowest. Try 0.475 and 0.425
67.72	0.475	-0.17	33221.06	Lowest. Try 0.4875 and 0.4625
67.72	0.4875	-0.17	33236.41	
67.72	0.4625	-0.16	33227.19	b_1=0.475 is lowest. Adjust a_0
67.35	0.475	0.08	33223.45	Close to minimum SSE and zero mean.

Autocorrelations for MA(1)

- For this series, the mean is:

$$\mu = E[x_t] = a_0 + E[e_t] + b_1 E[e_{t-1}] = a_0$$

- The variance is:

$$\begin{aligned}\text{Var}[x_t] &= E[(x_t - \mu)^2] = E[(e_t + b_1 e_{t-1})^2] \\ &= E[e_t^2 + 2b_1 e_t e_{t-1} + b_1^2 e_{t-1}^2] \\ &= E[e_t^2] + 2b_1 E[e_t e_{t-1}] + b_1^2 E[e_{t-1}^2] \\ &= \sigma^2 + 2b_1 \times 0 + b_1^2 \sigma^2 = (1 + b_1^2)\sigma^2\end{aligned}$$

- The autocovariance at lag 1 is:

$$\begin{aligned}\text{Covar at lag 1} &= E[(x_t - \mu)(x_{t-1} - \mu)] \\ &= E[e_t + b_1 e_{t-1})(e_{t-1} + b_1 e_{t-2})] \\ &= E[e_t e_{t-1} + b_1 e_{t-1} e_{t-1} + b_1 e_t e_{t-2} + b_1^2 e_{t-1} e_{t-2}] \\ &= \quad 0 + b_1 E[e_{t-1}^2] + 0 + 0 \\ &= b_1 \sigma^2\end{aligned}$$

Autocorrelations for MA(1) (Cont)

- The autocovariance at lag 2 is:

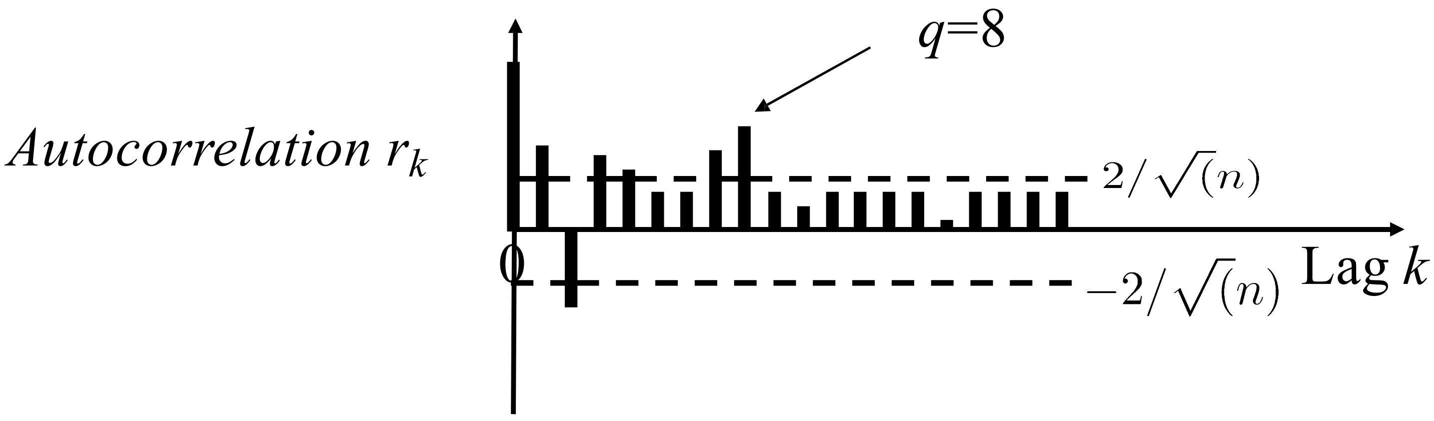
$$\begin{aligned} \text{Covar at lag 2} &= E[(x_t-\mu)(x_{t-2}-\mu] \\ &= E[(e_t+b_1e_{t-1})(e_{t-2}+b_1e_{t-3})] \\ &= E[e_te_{t-2}+b_1e_{t-1}e_{t-2}+b_1e_te_{t-3}+b_1^2e_{t-1}e_{t-3}] \\ &= 0+0+0+0=0 \end{aligned}$$

- For MA(1), the autocovariance at all higher lags (k>1) is 0
- The autocorrelation is:

$$r_k = \begin{cases} 1 & k=0 \\ \frac{b_1}{1+b_1^2} & k=1 \\ 0 & k>1 \end{cases}$$

- The autocorrelation of MA(q) series is non-zero only for lags $k \leq q$ and is zero for all higher lags.

Determining the Order MA(q)

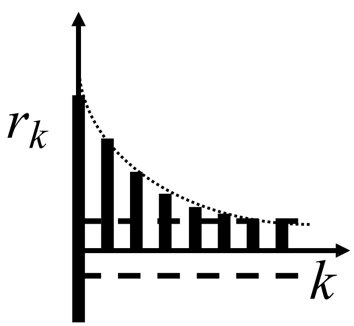


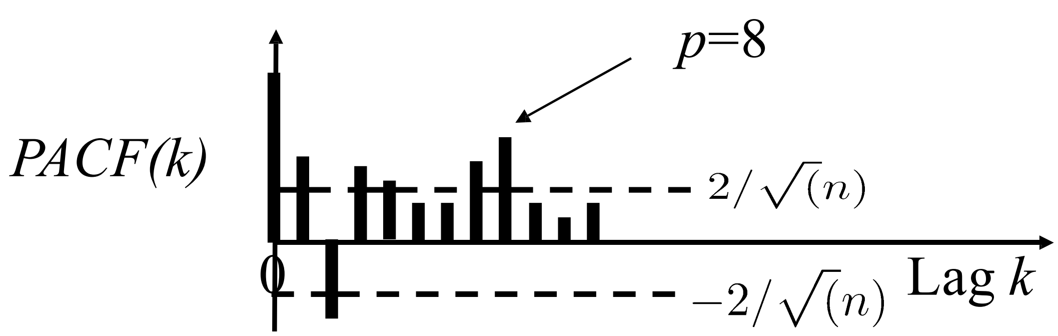
- The order of the last significant r_k determines the order of the MA(q) model

See also: Box-Jenkins Method

Determining the Order AR(p)

- ACF of AR(1) is an exponentially decreasing fn of k
- Fit AR(p) models of order p=0, 1, 2, …
- Compute the confidence intervals of a_p: $a_p \mp 2/\sqrt(n)$
- After some p, the last coefficients a_p will not be significant for all higher order models.
- This highest p is the order of the AR(p) model for the series.
- This sequence of last coefficients is also called **Partial Autocorrelation Function (PACF)**



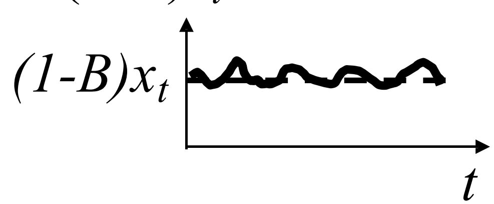


Non-Stationarity: Integrated Models

- In the white noise model AR(0): $x_t = a_0 + e_t$
- The mean a_0 is independent of time
- If it appears that the time series is increasing approximately linearly with time, the first difference of the series can be modeled as white noise: $(x_t - x_{t-1}) = a_0 + e_t$
- Or using the B operator: $(1\text{-}B)x_t = x_t\text{-}x_{t-1}$

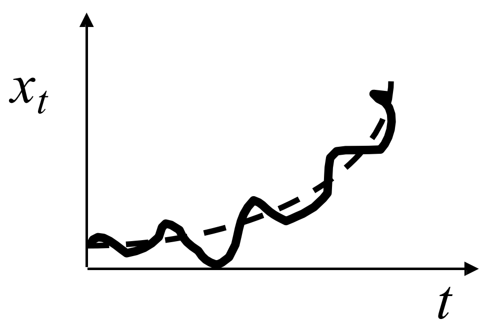
$$(1 - B)x_t = a_0 + e_t$$

- This is called an "integrated" model of order 1 or I(1). Since the errors are integrated to obtain x.
- Note that x_t is not stationary but $(1\text{-B})x_t$ is stationary.



Integrated Models (cntd.)

- If the time series is parabolic, the second difference can be modeled as white noise:
  $$(x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = a_0 + e_t$$
- Or $(1-B)^2 x_t = a_0 + e_t$
  This is an I(2) model



ARMA and ARIMA Models

- It is possible to combine AR, MA, and I models
- ARMA(p, q) Model:

$$
\begin{aligned}
x_t - a_1 x_{t-1} - \dots - a_p x_{t-p} &= a_0 + e_t + b_1 e_{t-1} + \dots + b_q e_{t-q} \\
\phi_p(B) x_t &= a_0 + \psi_q(B) e_t
\end{aligned}
$$

- ARIMA(p,d,q) Model:

$$\phi_p(B)(1-B)^d x_t = a_0 + \psi_q(B) e_t$$

Non-Stationarity due to Seasonality

- The mean temperature in December is always lower than that in November and in May it is always higher than that in March ⇒Temperature has a yearly season.
- One possible model could be I(12):

$$x_t - x_{t-12} = a_0 + e_t$$

- or

$$(1 - B)^{12} x_t = a_0 + e_t$$

Summary

- AR(1) Model:

$$x_t = a_0 + a_1 x_{t-1} + e_t$$

- MA(1) Model:

$$x_t - a_0 = e_t + b_1 e_{t-1}$$

- ARIMA(1,1,1) Model:

$$x_t - x_{t-1} = a_0 + a_1(x_{t-1} - x_{t-2}) + e_t + b_1 e_{t-1}$$