BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction to Statistical Methods

ISM Team

innovate achieve lead

BITS Pilani

Sessions 15
Time Series Analysis
(9th Sep 2023)

Time Series

- Consider the sequence of random variables Y_0, Y_1, Y_2, … which forms a family of random variables **$\{Y_t, t \in T\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$**

where Ω - Sample space

$\mathcal{F}$ - all collection of subsets A of Ω and

P – probability measure

- T – index set

Index set:

- The index set **T** is a collection of **all time functions** that can result from random experiment, usually the index **T** denote time. $\mathbf{Y_t}$ are **independent and identically distributed (iid) random variables** with mean μ_t and variance σ_t^2. Each realization of this process gives an ensemble or a data set.
- iid rvs: if each random variable has the same probability distribution as the others and all are mutually independent

- Examples of index sets:

  (1) T = $(-\infty, \infty)$ or T = $[0, \infty)$. In this case Y_t is a continuous time stochastic process.

  (2) T = $\{0, \pm 1, \pm 2, \ldots.\}$ or T = $\{0, 1, 2, \ldots\}$. In this case Y_t is a discrete time stochastic process.

- We use uppercase letter $\{Y_t\}$ to describe the process. A time series, $\{Y_t\}$ is a realization or sample function from a certain process.

- We use information from a time series to estimate parameters and properties of process $\{Y_t\}$.

Time Series ➡ Definition of Stochastic Process

- Y_t represents random quantity at time t

- In general, the value Y_t might depend on the quantity Y_{t-1} at time t-1, or even the value Y_s for other times $s < t$.

Time Series ⇨ Probability distribution of the process

- For any stochastic process with index set *T*, its probability distribution function is uniquely determined by its finite dimensional distributions.
- The *k* dimensional distribution function of a process is defined by

$$F_{Y_{t_1},Y_{t_2},\dots,Y_{t_k}} = P(Y_{t_1} \le y_{t_1}, Y_{t_2} \le y_{t_2}, \dots, Y_{t_k})$$

for any t_1, t_2, …,$t_k \in T$ and any real numbers $y_1, \dots, y_k$.

- The distribution function tells us everything we need to know about the process $\{Y_t\}$.
-

In stochastic time series,

$$Y_t,\ t \in Z = \{0, \pm 1, \pm 2, \ldots\}$$

is a family of random variables, Y_t denoting the value of the characteristic of interest at time t.

Thus, $\mathbf{Y} = (y_1, y_2, \ldots, y_n)'$ is seen as a realized value of the random vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$ with joint probability density function $f_{\mathbf{Y}}(y)$.

The joint distribution function of a finite random variables

$$\{Y_{t_1}, \ldots Y_{t_n}\},\ t_1 < t_2 < \ldots < t_n$$

from the collection $\{Y_t,\ t \in T\}$ is

$$F_{Y_{t_1}, \ldots Y_{t_n}}(y_1, \ldots, y_n) = P[Y_{t_1} \le y_{t_1}, \ldots, P[Y_{t_n} \le y_{t_n}],$$

$$(y_{t_1}, \ldots, y_{t_n}) \in R^n$$

A special kind of stochastic process is based on the assumption that the process is in a particular state of equilibrium. This type of assumption is called stationarity.

A stochastic process is strictly stationary if its properties are unaffected by a change of origin.

Time Series ⇨ Moments of Stochastic Processes

- We can describe a stochastic process via its moments, ie., $E(Y_t)$, $E(Y^2{}_t)$, $E(Y_t, Y_s)$, etc. We often use the first two moments.
- The mean function of the process is $E(Y_t) = \mu_t$.
- The variance function of the process is $V(Y_t) = \sigma^2{}_t$.

Time Series ⇨ Moments of Stochastic Processes

- The Covariance function between Y_t, and Y_s is

$$Cov\,(Y_t, Y_s) = E((Y_t - \mu_t)\,(Y_s - \mu_s))$$

- The Correlation function between Y_t, and Y_s is

$$\rho\,(Y_t, Y_s) = \frac{Cov\,(Y_t, Y_s)}{\sqrt{\sigma_t^2}\sqrt{\sigma_s^2}}$$

- These moments are functions of time

Thus, a time series stochastic process $\{Y_{t_1}, Y_{t_2}, \ldots Y_{t_k}\}$ is said to be strictly stationary, if the joint distribution of k observations $Y_{t_1}, \ldots Y_{t_k}$ made at time $t_1, \ldots, t_k$ is same as that of k+h observations $Y_{t_{1+h}}, Y_{t_{2+h}}, \ldots Y_{t_{k+h}}$ made at time points $t_{1+h}, \ldots, t_{k+h}$ for any h. That is,

$$F_{Y_{t_1}, \ldots Y_{t_k}}(y_{t_1}, \ldots, y_{t_k}) = F_{Y_{t_{1+h}}, Y_{t_{2+h}}, \ldots Y_{t_{k+h}}}(y_{t_{1+h}}, y_{t_{2+h}}, \ldots, y_{t_{k+h}})$$

Time Series ⇨ Stationary Stochastic Process in Time series

- If $\{Y_t\}$ is a strictly stationary process and $E(Y^2_t) < \infty$ then, the mean function is a constant and the variance function is also a constant.
- Moreover, for a strictly stationary process with first two moments finite, the covariance function, and the correlation function depends only on the time difference s.
- A trivial example of a strictly stationary process is a sequence of independent and identically distributed (iid) random variables.

That is, $F_{Y_{t_1}, \dots Y_{t_k}}(y_{t_1}, \dots, y_{t_k}) = F_{Y_{t_{1+h}}, \dots Y_{t_{k+h}}}(y_{t_{1+h}}, \dots, y_{t_{k+h}})$

for all possible non-empty finite distinct sets

$(t_1, \dots, t_k)$ and $(t_{1+h}, \dots, t_{k+h})$ in the index set T and all $y_{t_1}, \dots, y_{t_k}$ in the range of random variables Y_t.

Note: For reliable prediction to be made, the time series should be stationary (No systematic change such as seasonality, trend i.e. only random fluctuations)

Time Series ⇨ Stochastic Process

- Strict stationarity is too strong of a condition in practice. It is often difficult assumption to assess based on an observed time series $Y_1, \ldots, Y_k$.
- In time series analysis we often use a weaker sense of stationarity in terms of the moments of the process.
- A process is said to be *n*th-order weakly stationary if all its joint moments up to order *n* exists and are time invariant, i.e., independent of time origin.

Time Series ➡ Stochastic Process

- For example, a second-order weakly stationary process will have constant mean and variance, with the covariance and the correlation being functions of the time difference along.
- A strictly stationary process with the first two moments finite is also a second-ordered weakly stationary. But a strictly stationary process may not have finite moments and therefore may not be weakly stationary.

When k = 1, strict stationary implies that the pdf $f(y_t)$ is the same for all t, is say f(y). The stochastic process f(y) has a constant mean

$$\mu = E(Y_t) = \int_{-\infty}^{\infty} y f(y) dy$$

and a constant variance

$$\sigma^2 = E(Y_t - \mu)^2 = \int_{-\infty}^{\infty} (y - \mu)^2 \mathrm{f(y)dy}$$

provided the mean and variance exists.

The mean (μ) and variance (σ²) are estimated as

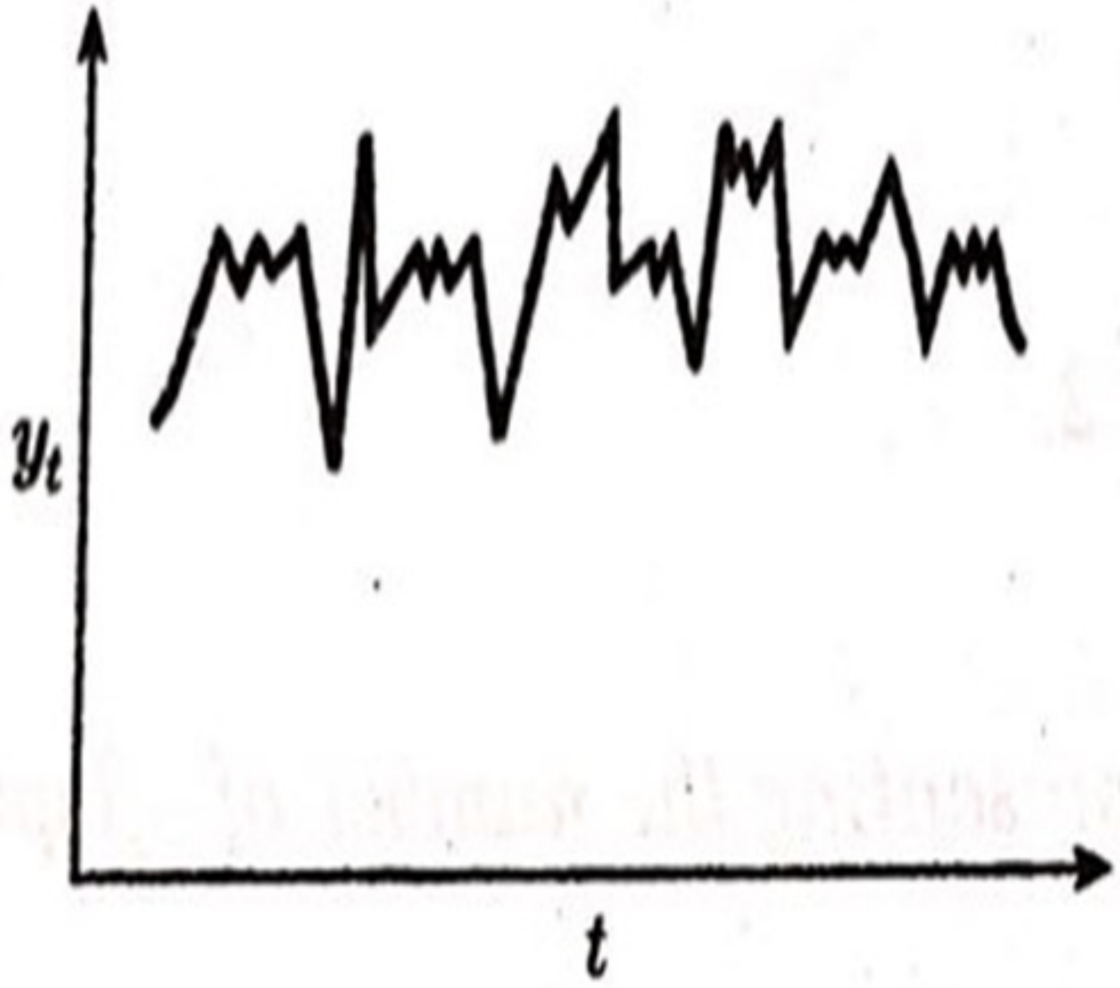
$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

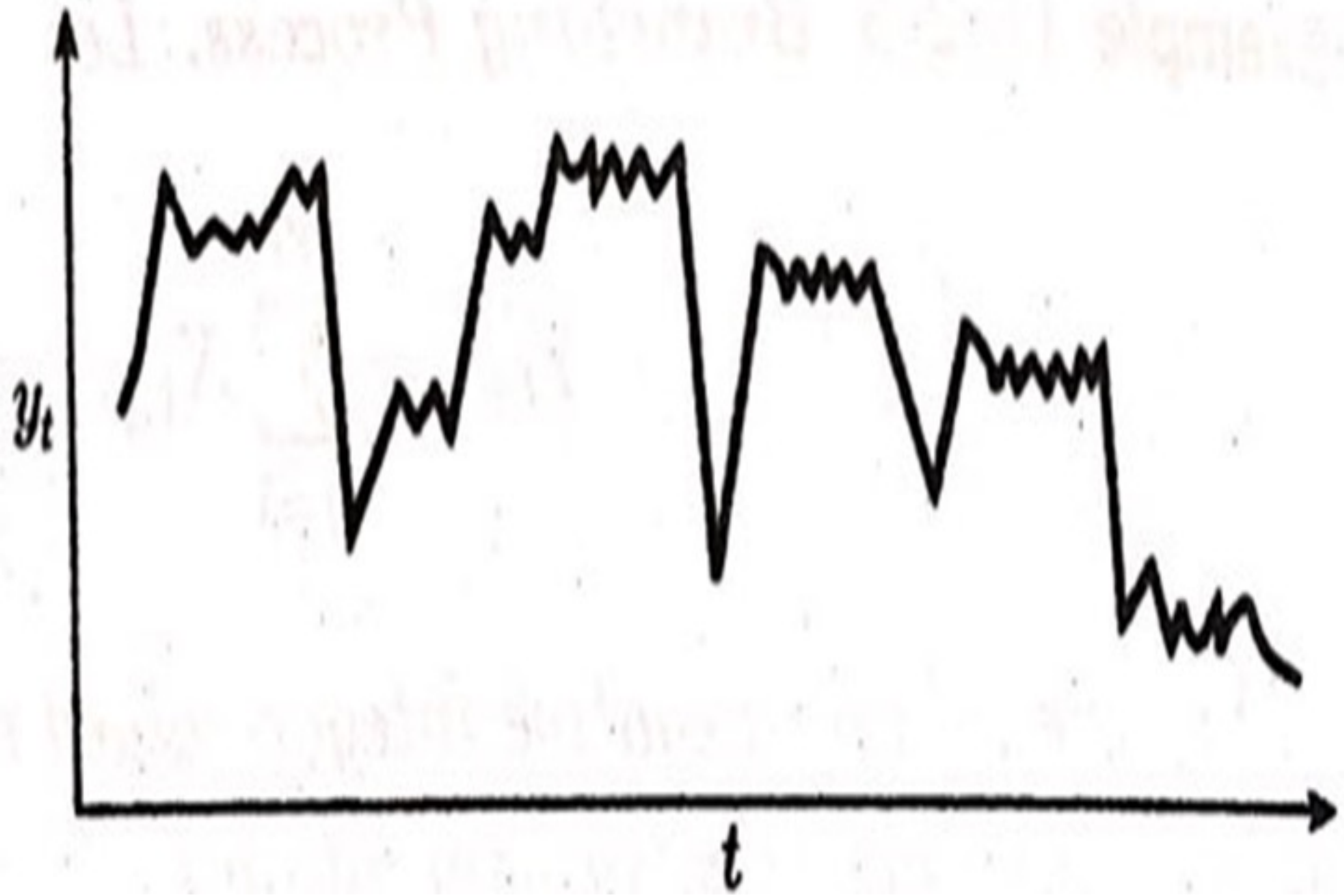
$$\hat{\sigma}_y^2 = S_y^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2$$

Thus, a stationary process remains at equilibrium about a common mean value.

However, in industry, business (ex. Stock price) and economics many time series are better represented as non-stationary and in particular, having no natural mean.

Time Series ➡ Stationary Stochastic Process in Time series





Stationary time series

Non-stationary time series

Time Series ➡ Stationary Stochastic Process in Time series

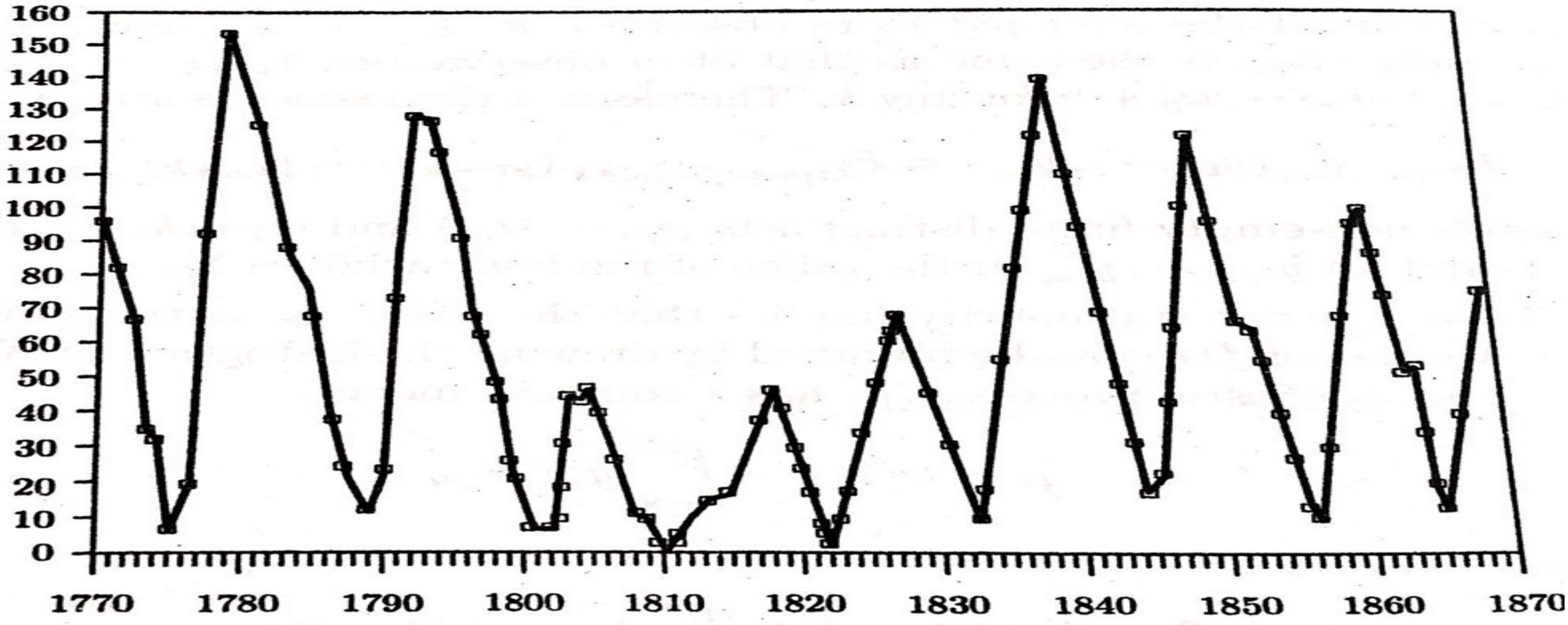


Fig 15.3 The Wolfer sunspot numbers, 1770-1869.
(Source: Box and Jenkins, 1976)

However, many non-stationary time series can be so modified that the reduced to time series obeys the original series.

The two main component which cause lack of stationarity are trend and seasonality. In fitting the stationary time series model, we therefore, assume that the trend and seasonality have been eliminated from the original series.

- Stationarity implies that the joint distribution of pair of observations y_{t_1}, y_{t_2} viz.,

$$f(y_{t_1}, y_{t_2}) = f(y_{t_1+h}, y_{t_2+h})$$

for any integer h. That is, the joint distribution of any pair of observations on time points which differ by a constant quantity is the same for such pairs.

Thus

$$f(y_t, y_{t+h}) = f(y_{t_1}, y_{t_1+h}) = f(y_{t_2}, y_{t_2+h})$$

have the same distribution.

The form of $f(y_t, y_{t+h})$ can be inferred by plotting values of (y_t, y_{t+h}); $t = 1, 2, \ldots,$ i.e., values of y_t separated by lag h.

Time Series ➡ Autocovariance

- Let $\{Y_t\}$ be a Time Series with $E\{Y_t^2\} < \infty$. The mean function of $\{Y_t\}$ is $\mu_y(t) = E(Y_t)$.
- The covariance function of $\{Y_t, Y_{t+h}\}$ is
- $\gamma_h = \mathrm{Cov}(Y_t, Y_{t+h}) = E[(Y_t-\mu)(Y_{t+h}-\mu)]$
- In general

$$\gamma_y(r, s) = \mathrm{Cov}(Y_r, Y_s) = E[\{Y_r-\mu_y(r)\}\{Y_s-\mu_y(s)\}]$$

for all integers r and s

- Clearly, with h=0, in $\gamma_h = \text{Cov}\,(Y_t, Y_{t+h}) = E[(Y_t-\mu)(Y_{t+h}-\mu)]$

$$\gamma_0 = \text{Cov}\,(Y_t, Y_{t+0}) = E[(Y_t-\mu)(Y_{t+0}-\mu)] = \sigma^2_y$$

$$|\gamma_h| \le \gamma_0, \forall\, h = 1, 2, \ldots$$

because, $|Cov(Y_{t+h}, Y_t)| \le \sqrt{V(Y_{t+h})V(Y_t)}$

Time Series ⇨ Autocovariance function

- Let $\{Y_t\}$ be a stationery Time Series.

- The Autocovariance Function (ACVF) of $\{Y_t\}$ at lag h is

$$\gamma_y(t+h, t) = \mathrm{Cov}\,(Y_{t+h}, Y_t).$$

The Autocorrelation Function (ACF) of $\{Y_t\}$ at lag h is

$$\rho_h = \frac{\gamma_h}{\gamma_0} = \frac{\mathrm{Cov}(Y_{t+h}, Y_t)}{\sigma^2}\ , \quad \boldsymbol{\rho_0 = 1}$$

Time Series ➡ Autocorrelation

- Correlation: measure of relationship between two variables are related
  - Direction of the relationship (Negative, Zero, and/ or positive
  - Degree/ Extent of relationship

Time Series ⇨ Autocorrelation

- The autocorrelation measures the degree of relationship of the same variable between the observation at the current time period and the observations at the prior time periods.
- It measures how the lagged version of the value of a variable is related to the original version of it in time series.

The value of r_k can be written as

$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2}$$

where T is the length of time series.

Time Series ⇨ Autocorrelation

- Autocorrelation between (Y_t, Y_{t-1})

$$r_1 = \frac{\sum(Y_t - \bar{Y})\,(Y_{t-1} - \bar{Y})}{\sum(Y_t - \bar{Y})^2}$$

- Autocorrelation between (Y_t, Y_{t-2})

$$r_2 = \frac{\sum(Y_t - \bar{Y})\,(Y_{t-2} - \bar{Y})}{\sum(Y_t - \bar{Y})^2}$$

– – – –

- Autocorrelation between (Y_t, Y_{t-k})

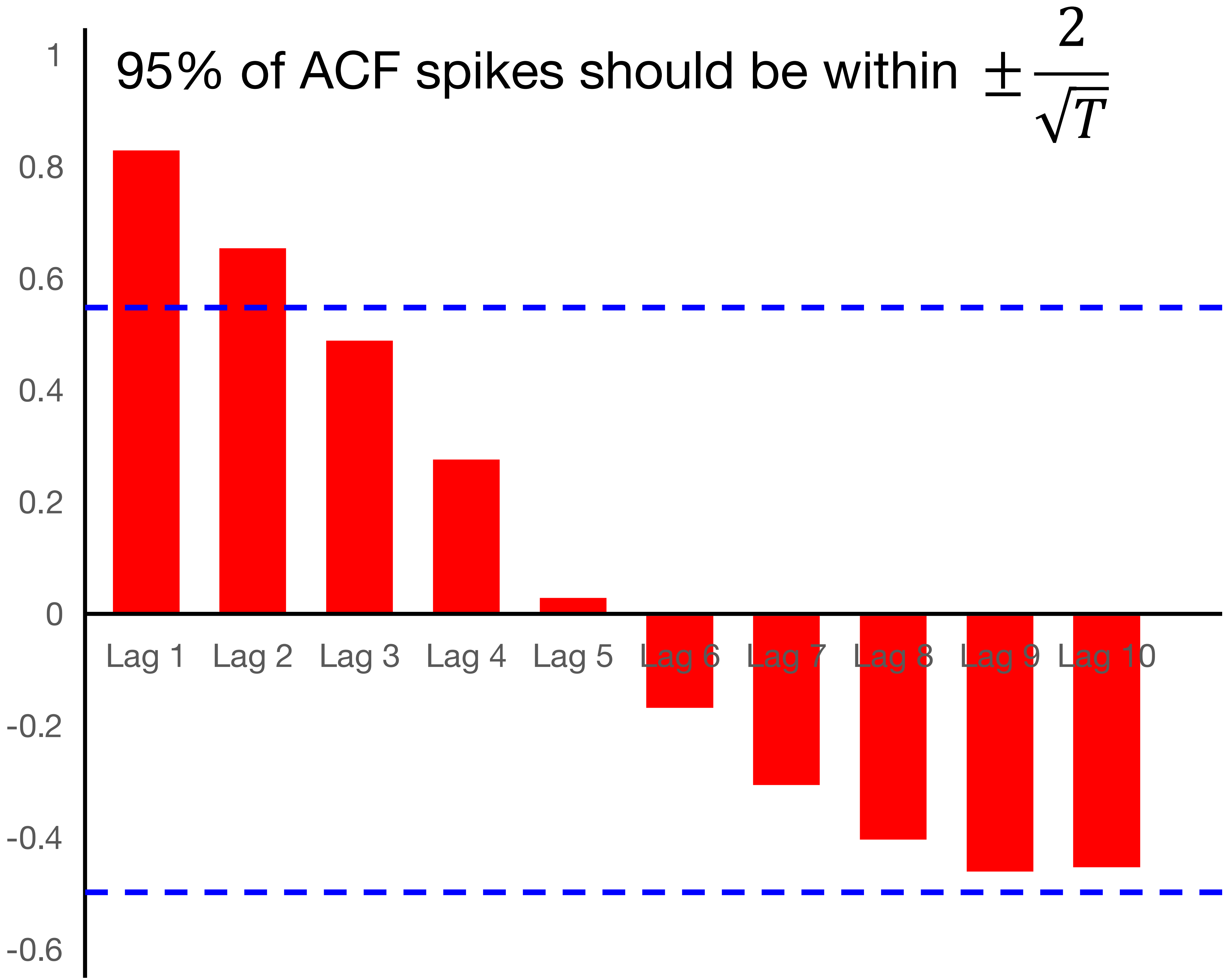
$$r_k = \frac{\sum(Y_t - \bar{Y})\,(Y_{t-k} - \bar{Y})}{\sum(Y_t - \bar{Y})^2}$$

- There are several autocorrelation coefficients, corresponding to each panel in the lag plot. For example, r_1 measures the relationship between Y_t and Y_{t-1}, r_2 measures the relationship between Y_t and Y_{t-2}, and so on rk measures the relationship between Y_t and $Y_{t\text{-}k}$.

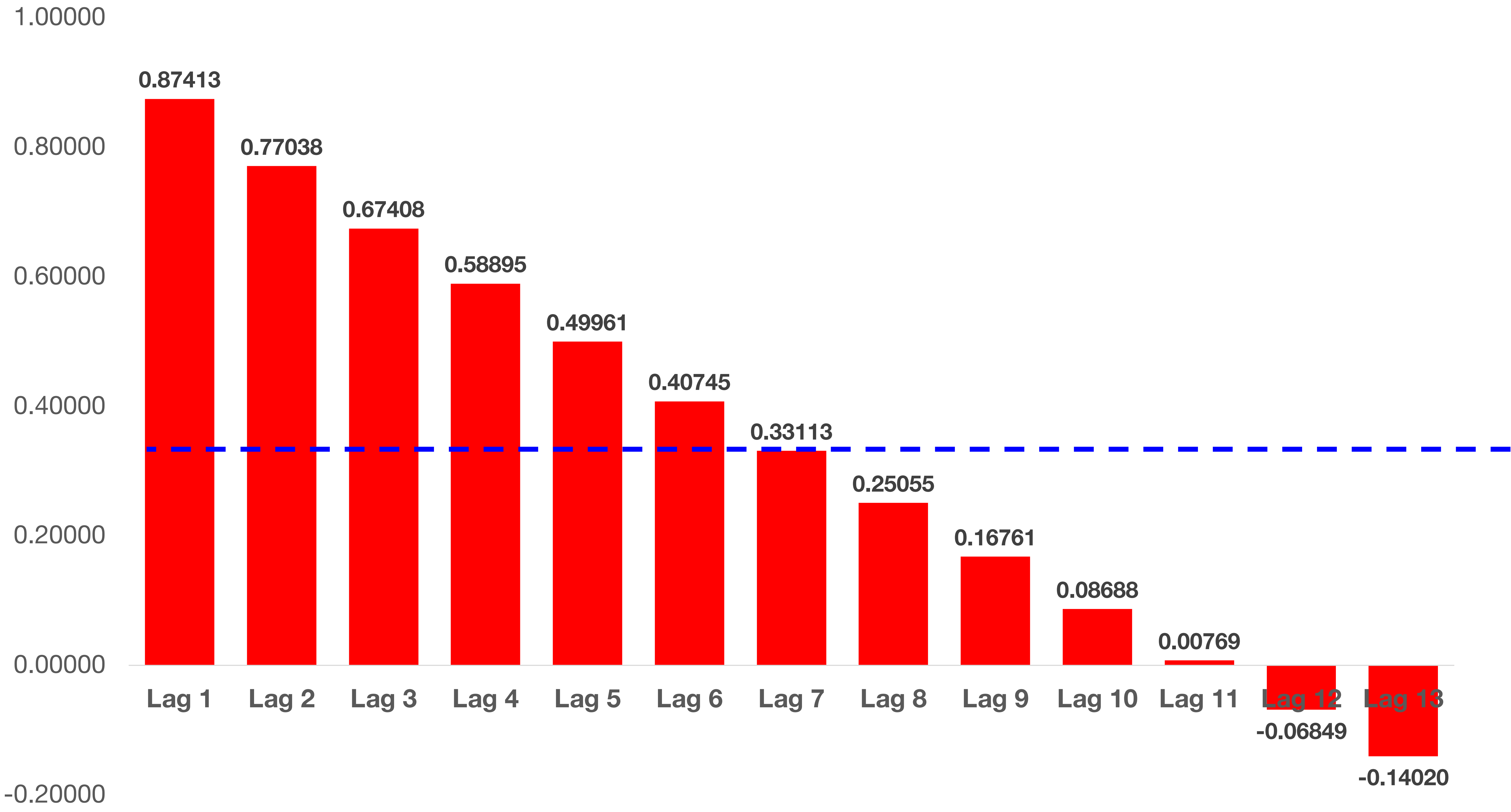
Year	2011				2012			
Quarter	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Retail quarterly sale (million) (Y_t)	147772	154400	166188	170202	173264	175371	184957	186395
Year	2013				2014			
Quarter	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Retail quarterly sale (million) (Y_t)	191130	191213	195749	198262	199980	209566	212529	213754
Year	2015				2016			
Quarter	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Retail quarterly sale (million) (Y_t)	222124	224372	229871	236260	238044	244076	244944	245799
Year	2017				2018			
Quarter	Q1	Q2	Q3	Q4	Q1	Q2		
Retail quarterly sale (million) (Y_t)	251125	254557	255223	264793	265064	277282		

Time	Value	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
1	22										
2	24	22									
3	25	24	22								
4	25	25	24	22							
5	28	25	25	24	22						
6	29	28	25	25	24	22					
7	34	29	28	25	25	24	22				
8	37	34	29	28	25	25	24	22			
9	40	37	34	29	28	25	25	24	22		
10	44	40	37	34	29	28	25	25	24	22	
11	51	44	40	37	34	29	28	25	25	24	22
12	48	51	44	40	37	34	29	28	25	25	24
13	47	48	51	44	40	37	34	29	28	25	25
14	50	47	48	51	44	40	37	34	29	28	25
15	51	50	47	48	51	44	40	37	34	29	28

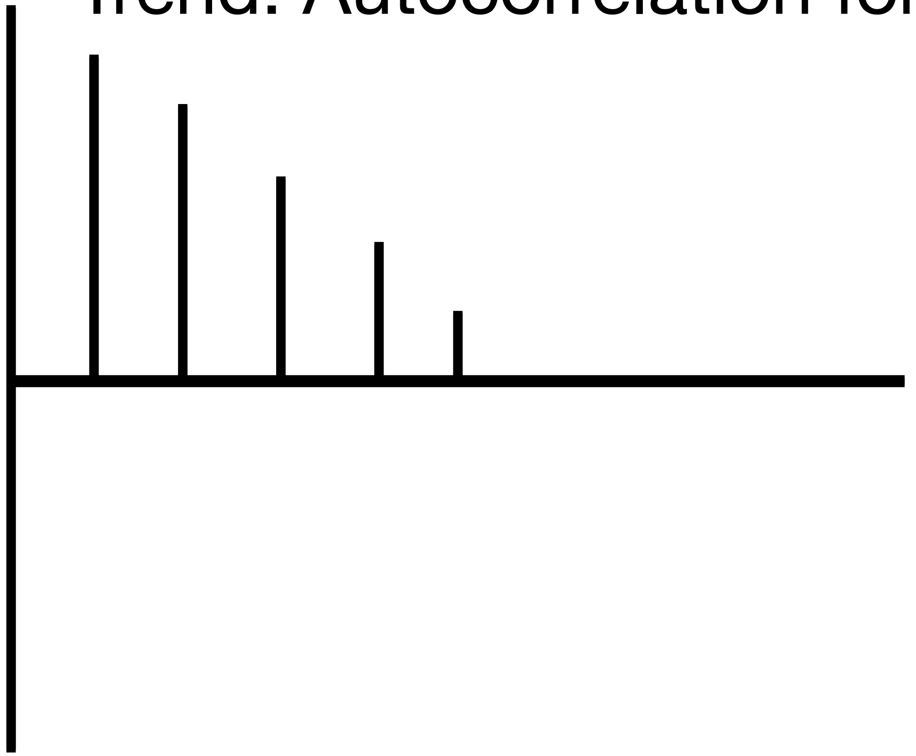
	Autocorrelation
Lag 1	0.83174
Lag 2	0.65632
Lag 3	0.49105
Lag 4	0.27864
Lag 5	0.03103
Lag 6	-0.1653
Lag 7	-0.3037
Lag 8	-0.401
Lag 9	-0.4582
Lag 10	-0.4505

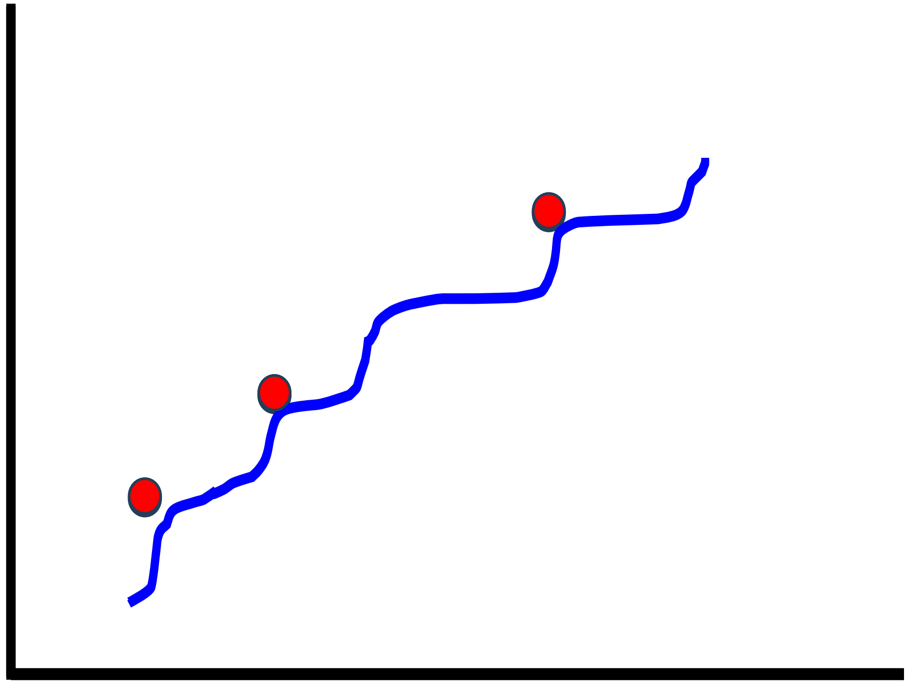


Month/ Year	Retail quarterly sale (million)	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6
Q1 - 2011	147772						
Q2 - 2011	154400	147772					
Q3 - 2011	166188	154400	147772				
Q4 - 2011	170202	166188	154400	147772			
Q1 - 2012	173264	170202	166188	154400	147772		
Q2 - 2012	175371	173264	170202	166188	154400	147772	
Q3 - 2012	184957	175371	173264	170202	166188	154400	147772
Q4 - 2012	186395	184957	175371	173264	170202	166188	154400
Q1 - 2013	191130	186395	184957	175371	173264	170202	166188
Q2 - 2013	191213	191130	186395	184957	175371	173264	170202
Q3 - 2013	195749	191213	191130	186395	184957	175371	173264
Q4 - 2013	198262	195749	191213	191130	186395	184957	175371
Q1 - 2014	199980	198262	195749	191213	191130	186395	184957
Q2 - 2014	209566	199980	198262	195749	191213	191130	186395
Q3 - 2014	212529	209566	199980	198262	195749	191213	191130
Q4 - 2014	213754	212529	209566	199980	198262	195749	191213
Q1 - 2015	222124	213754	212529	209566	199980	198262	195749
Q2 - 2015	224372	222124	213754	212529	209566	199980	198262
Q3 - 2015	229871	224372	222124	213754	212529	209566	199980
Q4 - 2015	236260	229871	224372	222124	213754	212529	209566
Q1 - 2016	238044	236260	229871	224372	222124	213754	212529
Q2 - 2016	244076	238044	236260	229871	224372	222124	213754
Q3 - 2016	244944	244076	238044	236260	229871	224372	222124
Q4 - 2016	245799	244944	244076	238044	236260	229871	224372
Q1 - 2017	251125	245799	244944	244076	238044	236260	229871
Q2 - 2017	254557	251125	245799	244944	244076	238044	236260
Q3 - 2017	255223	254557	251125	245799	244944	244076	238044
Q4 - 2017	264793	255223	254557	251125	245799	244944	244076
Q1 - 2018	265064	264793	255223	254557	251125	245799	244944
Q2 - 2018	277282	265064	264793	255223	254557	251125	245799

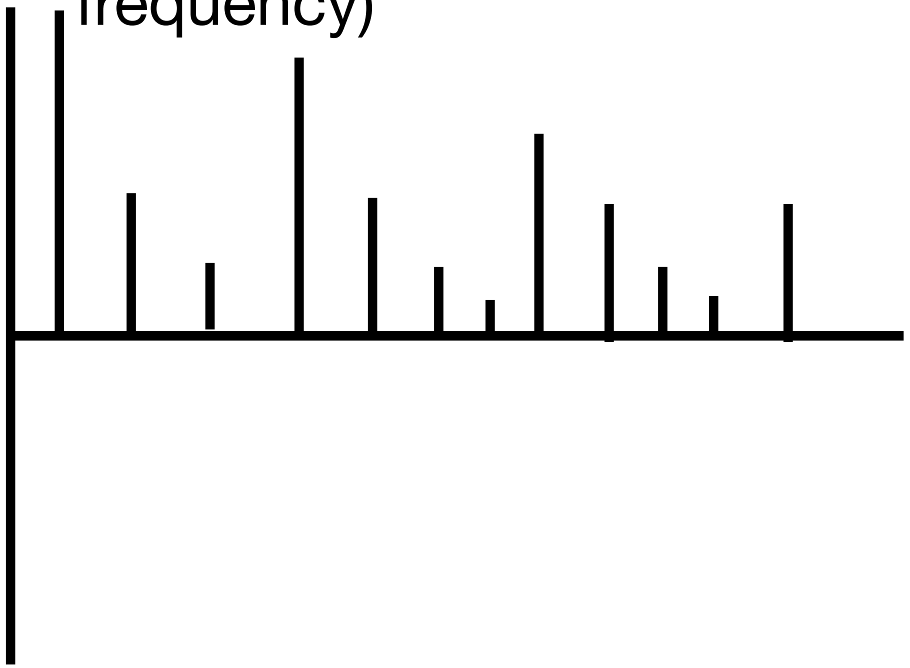

1.00000
0.80000
0.60000
0.40000
0.20000
0.00000
-0.20000
0.87413
0.77038
0.67408
0.58895
0.49961
0.40745
0.33113
0.25055
0.16761
0.08688
0.00769
-0.06849
-0.14020
Lag 1
Lag 2
Lag 3
Lag 4
Lag 5
Lag 6
Lag 7
Lag 8
Lag 9
Lag 10
Lag 11
Lag 12
Lag 13

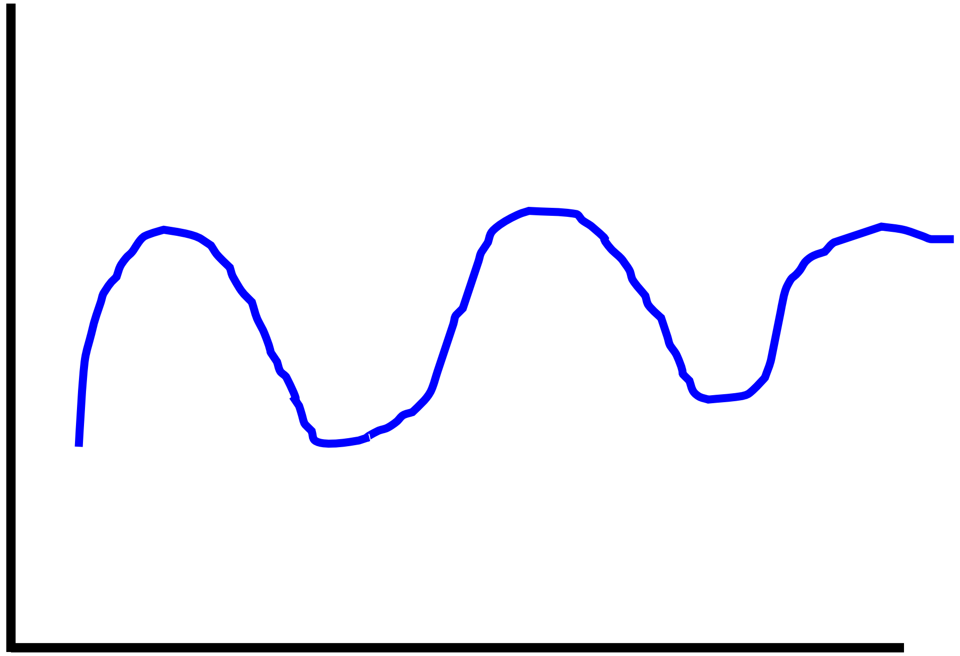
- Trend: Autocorrelation for small lags tend to be large and positive



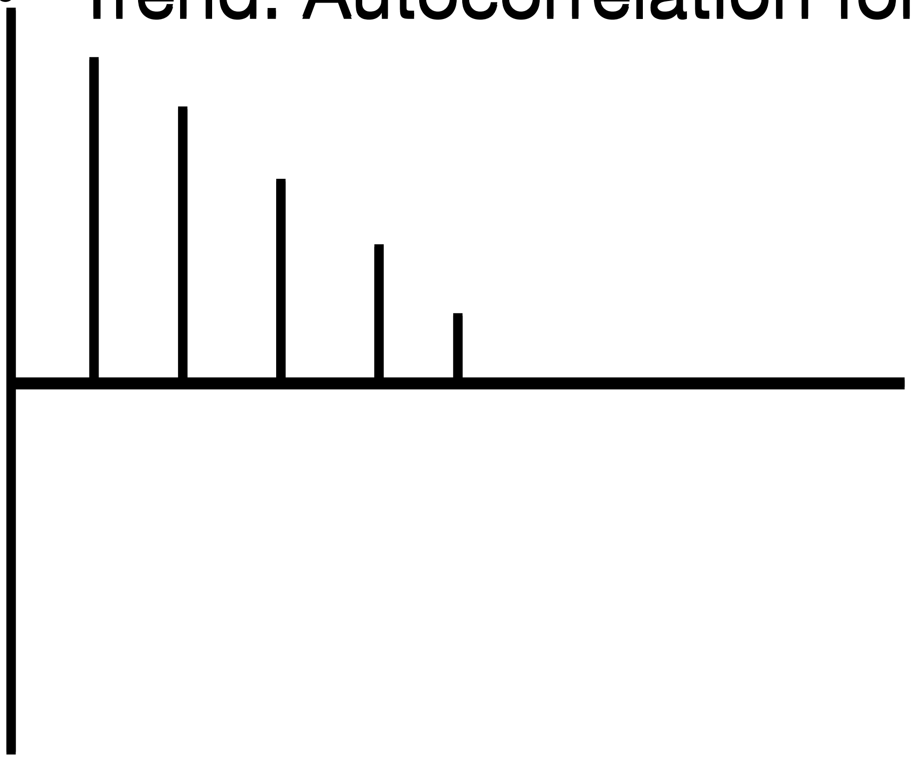


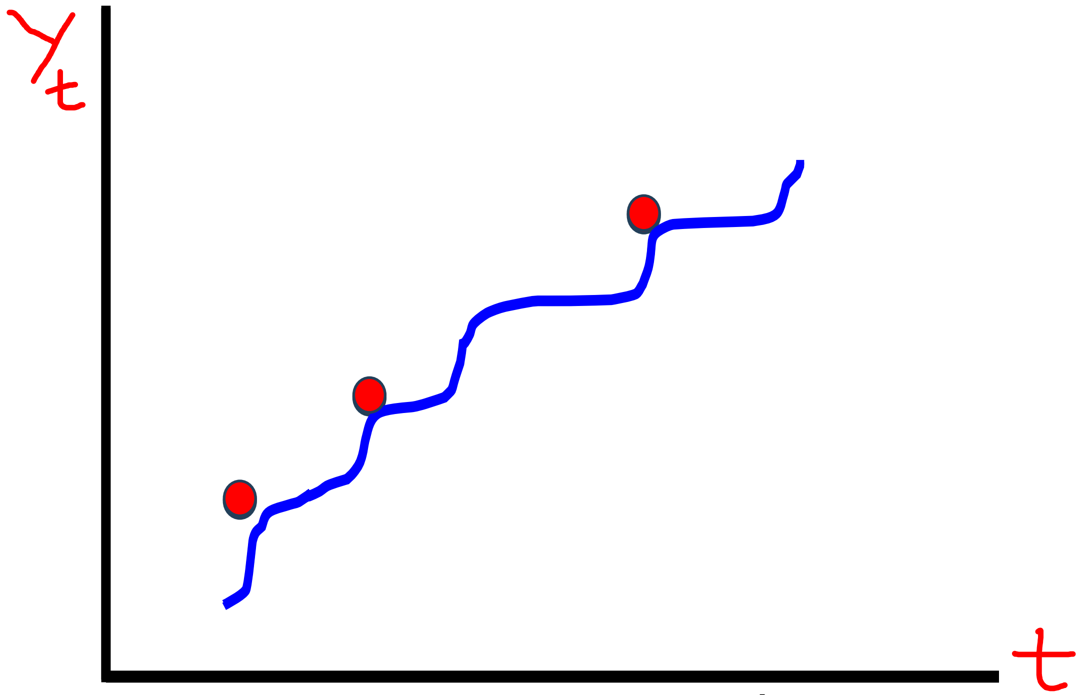
- Seasonal: Autocorrelation will be larger for smaller seasonal lags (at multiple seasonal frequency)



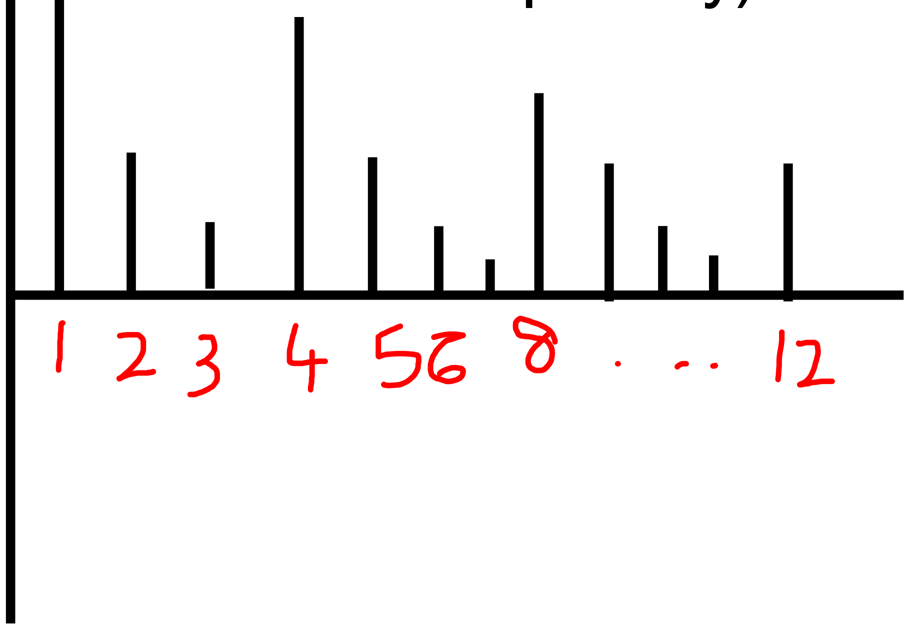


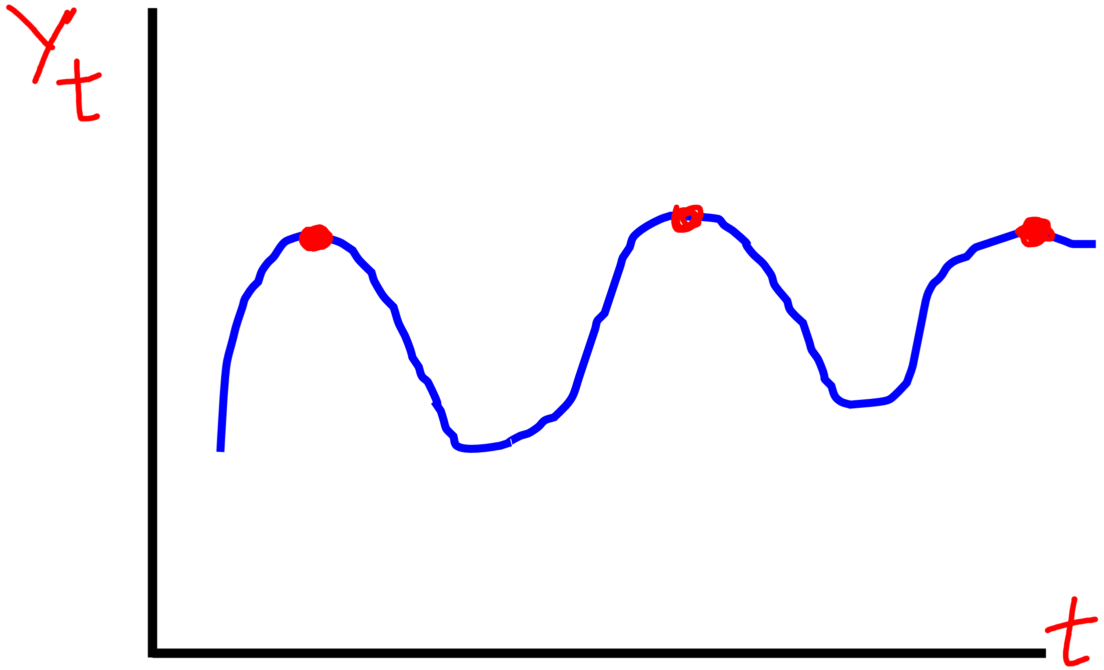
- Trend: Autocorrelation for small lags tend to be large and positive



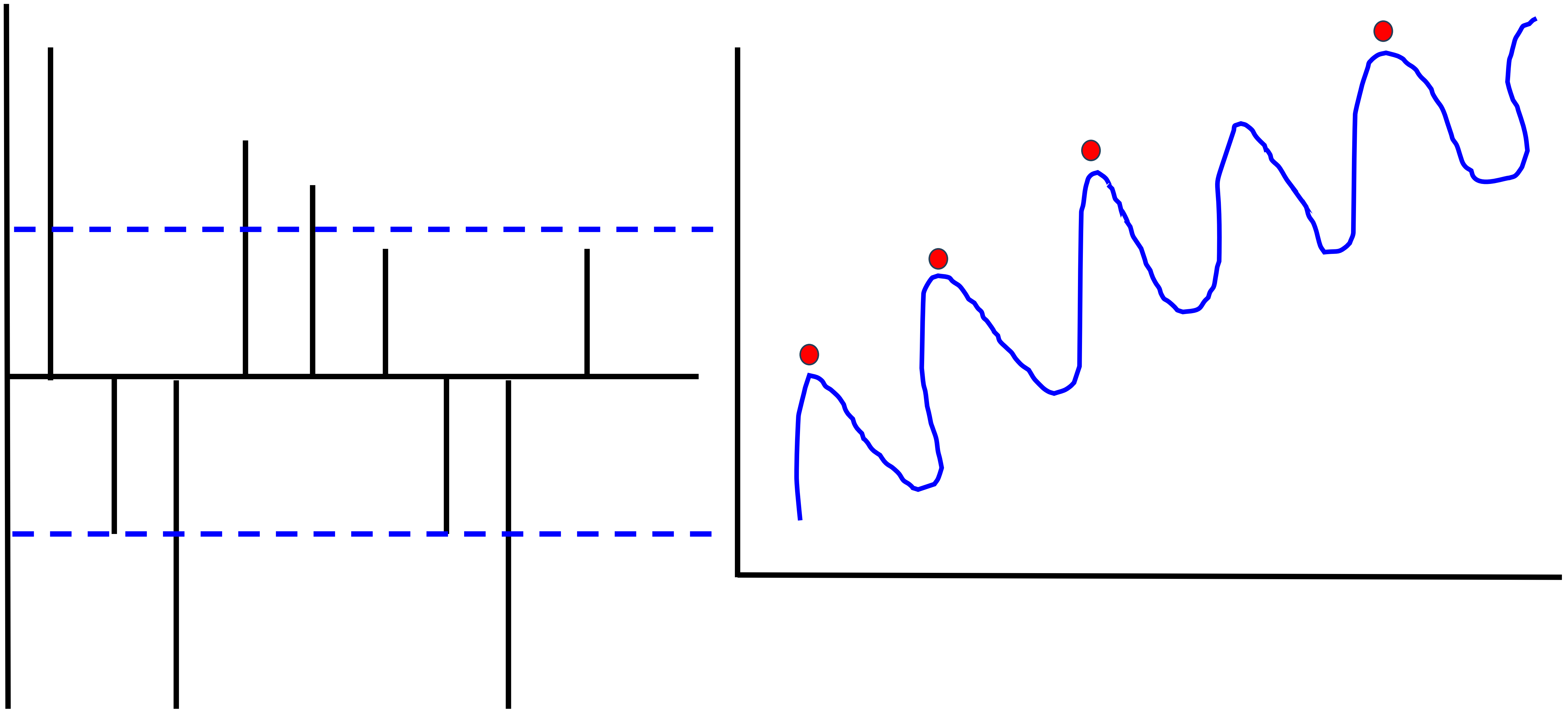


- Seasonal: Autocorrelation will be larger for smaller seasonal lags (at multiple seasonal frequency)





Seasonal series



Time Series ⇨ White Noise

Time series that show no autocorrelation are called **white noise.** That is, if the variables are independent and identically distributed with a mean of zero. This means that all variables have the same variance (σ^2) and each value has a zero correlation with all other values in the series.

Time Series ➔ White Noise

A white noise called **white** when it has the same intensity at every frequency. Its name is derived by analogy to light, which is **called** "**white**" when it contains all visible frequencies **noise.**

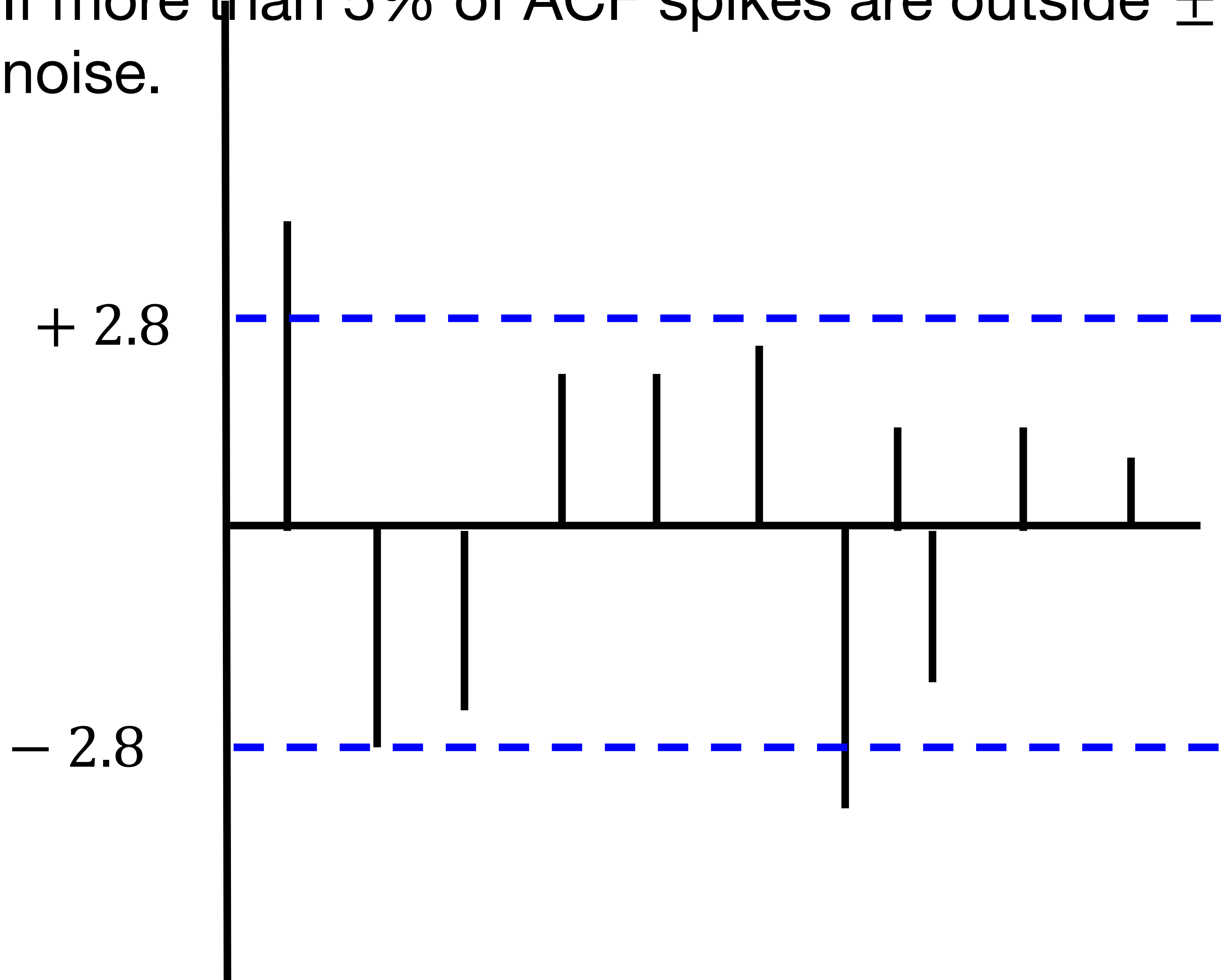
The term **noise**, in this context, came from signal processing where it was used to refer to unwanted electrical or electromagnetic energy that degrades the quality of signals and data

For white noise series, we expect each autocorrelation to be close to zero. Of course, they will not be exactly equal to zero as there is some random variation. For a white noise series, we expect 95% of the spikes in the ACF to lie within $\pm 2/\sqrt{T}$ where T is the length of the time series.

Time Series → White Noise

It is common to plot these bounds on a graph of the ACF (the blue dashed lines above). If one or more large spikes are outside these bounds, or if substantially more than 5% of spikes are outside these bounds, then the series is probably not white noise

- White Noise series
- Time series that shows no autocorrelation
- 95% of ACF spikes should be within $\pm\frac{2}{\sqrt{T}}$
- If more than 5% of ACF spikes are outside $\pm\frac{2}{\sqrt{T}}$, then the series is not white noise.

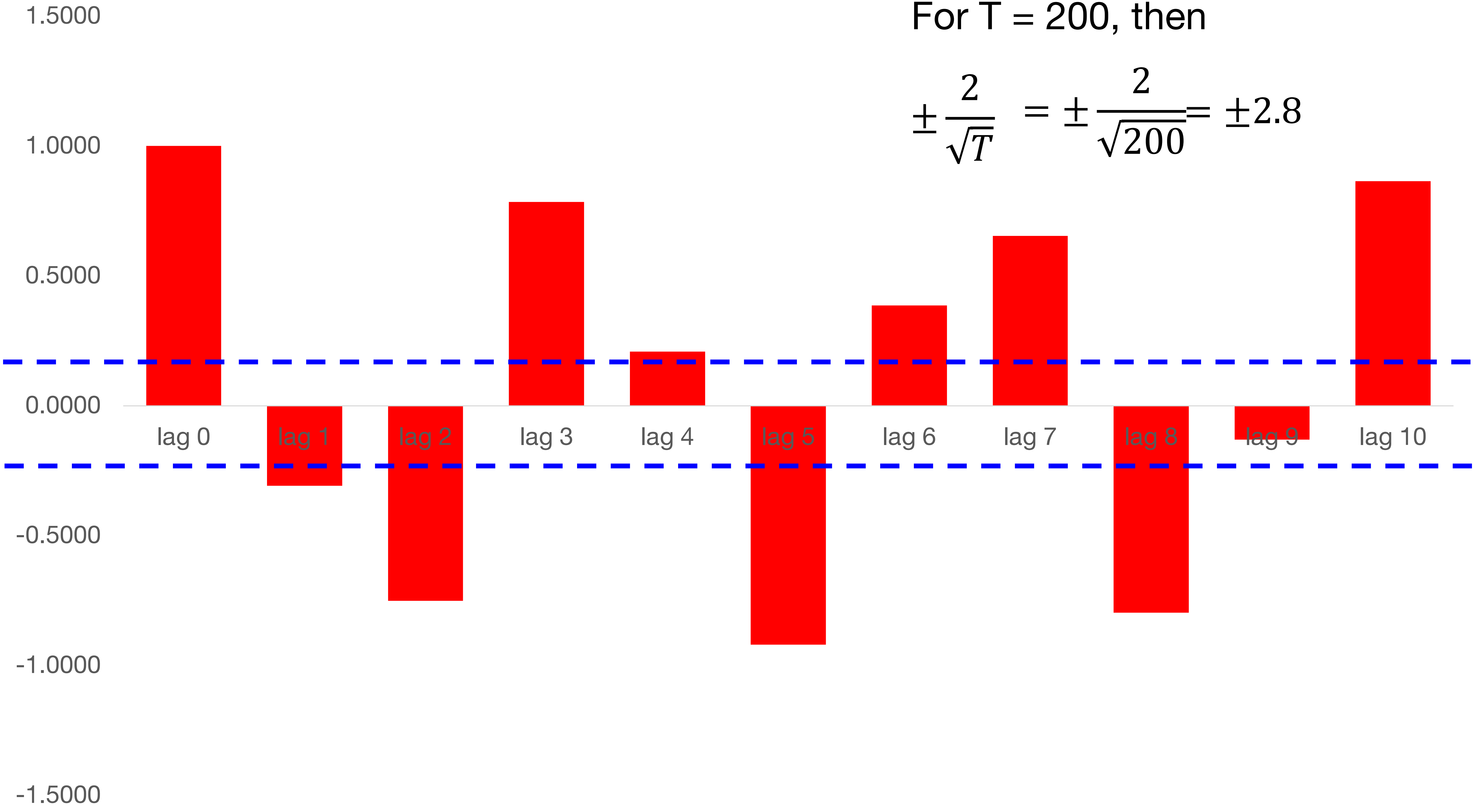


Example: If T = 50, then

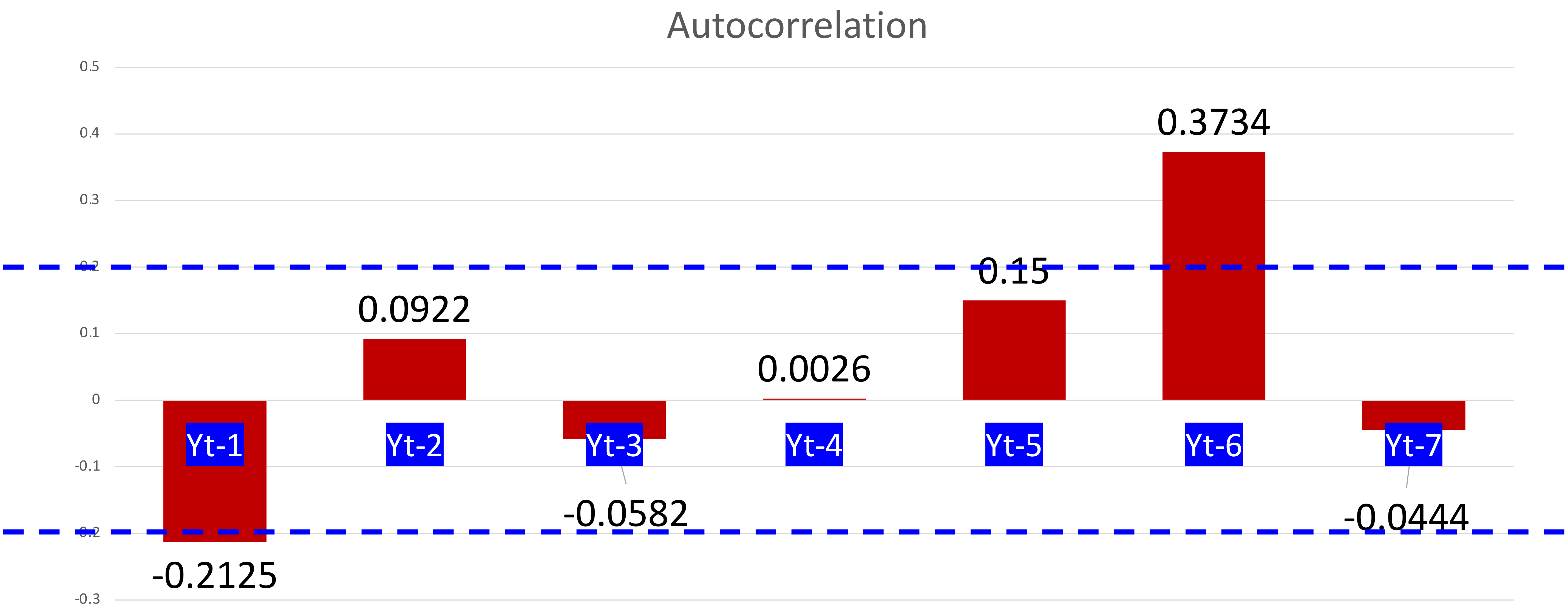
$$\pm\frac{2}{\sqrt{T}} = \pm\frac{2}{\sqrt{50}} = \pm 2.8$$

For T = 200, then

$$\pm\frac{2}{\sqrt{T}} = \pm\frac{2}{\sqrt{200}} = \pm 2.8$$



Time Series → Autocorrelation



Time Series ⇨ Test for Autocorrelation: Ljung-Box test

- A statistical test of whether any group of autocorrelation of a time series are different from zero

- A test for the overall randomness based on a number of lags

- **H_0:** The series is random or white noise or independent and identically distributed (iid)

- **H_1:** The series exhibits serial/ autocorrelation (non-random)

- A small P-value ($P < 0.05$) ⇨ Reject H_0 ⇨ the series not white noise
- A large P-value ($P > 0.05$) ⇨ Fail to Reject H_0 ⇨ the series white noise

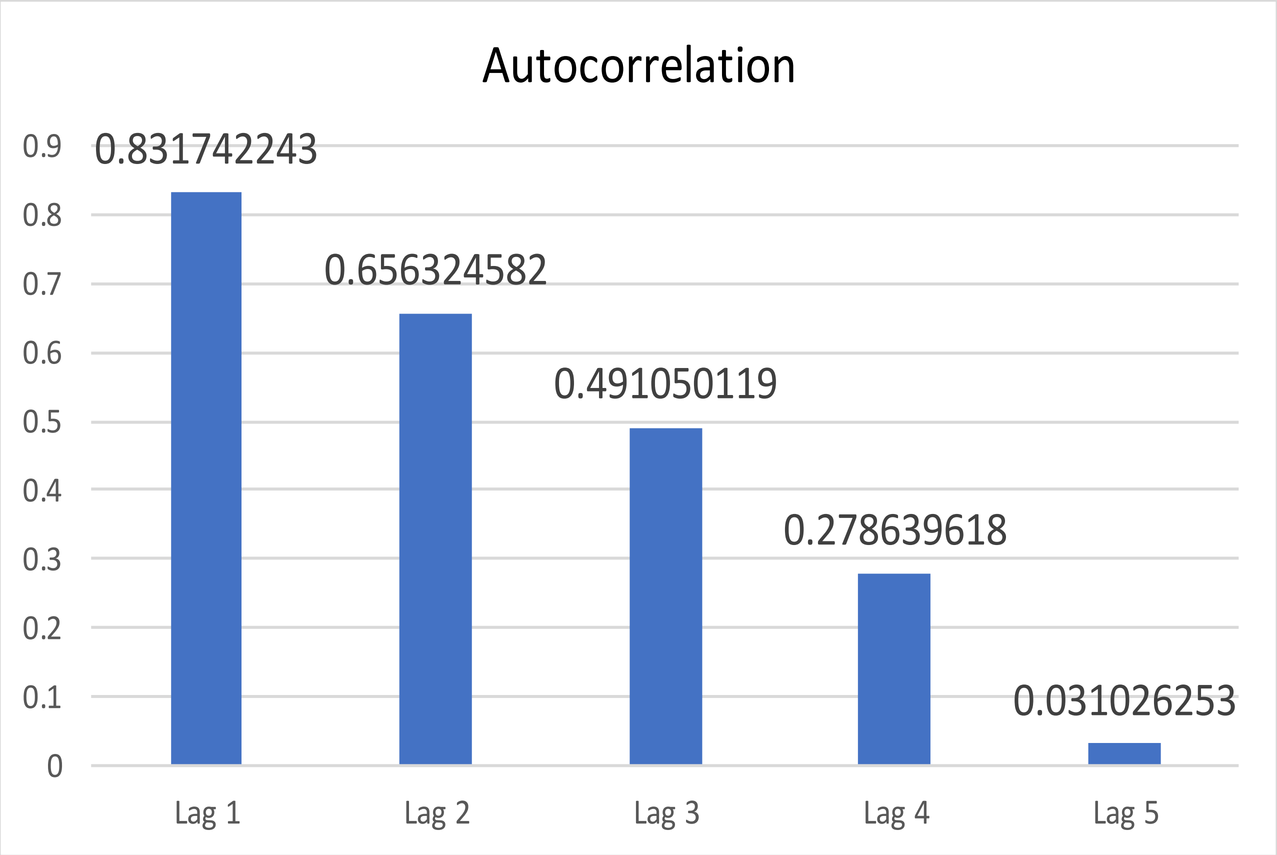
- **Note: Normally, P-value will be very large for white noise series**

innovate achieve lead

Time Series ➡ Autocorrelation

Time	Value	lag 1	lag 2	lag 3	lag 4	lag 5
1	22					
2	24	22				
3	25	24	22			
4	25	25	24	22		
5	28	25	25	24	22	
6	29	28	25	25	24	22
7	34	29	28	25	25	24
8	37	34	29	28	25	25
9	40	37	34	29	28	25
10	44	40	37	34	29	28
11	51	44	40	37	34	29
12	48	51	44	40	37	34
13	47	48	51	44	40	37
14	50	47	48	51	44	40
15	51	50	47	48	51	44

Lag 1	0.8317
Lag 2	0.6563
Lag 3	0.4911
Lag 4	0.2786
Lag 5	0.0310



Time Series ➡ Partial Autocorrelation Function (PACF)

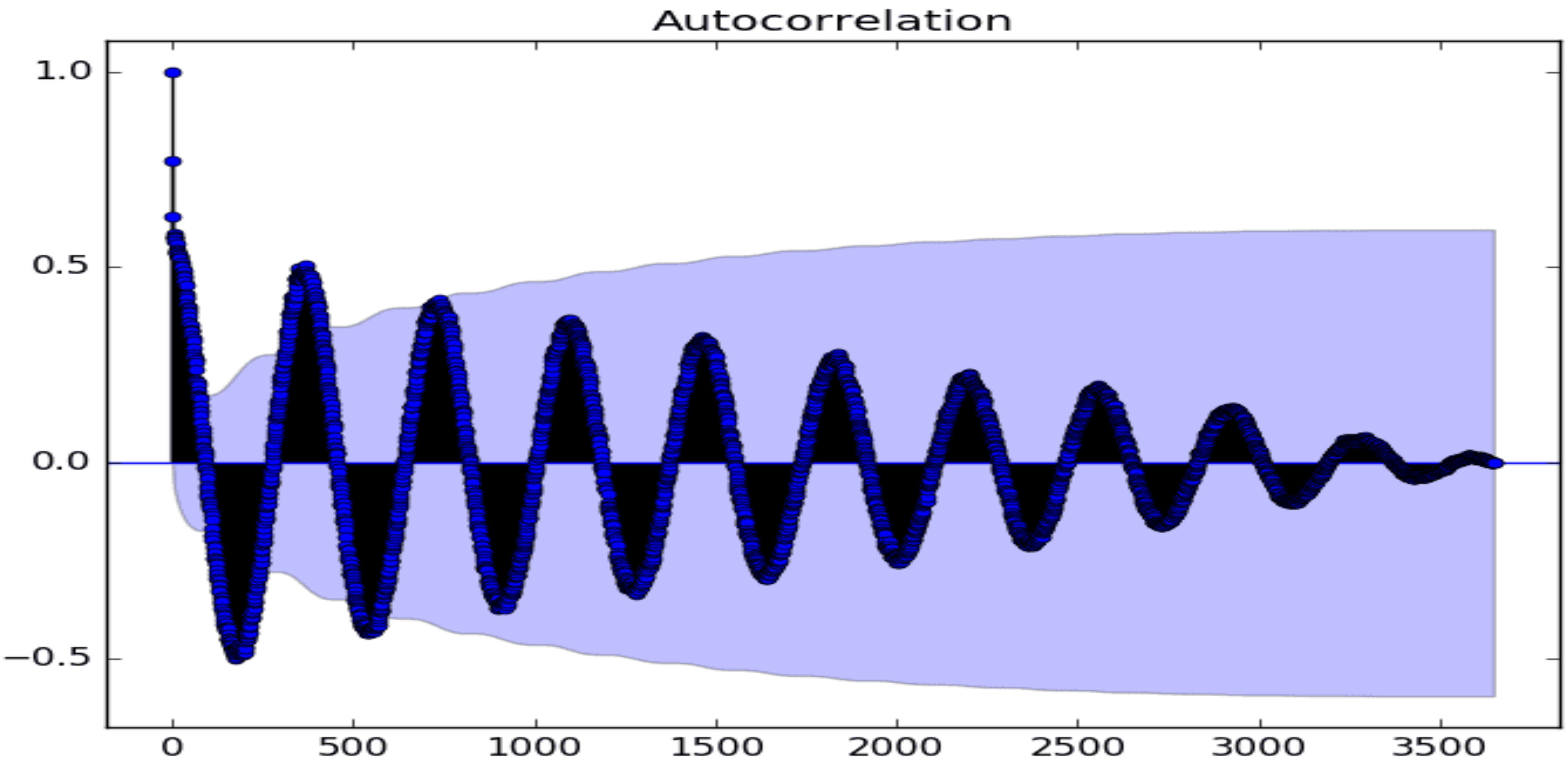
- The correlation between the observation at two time points given that we consider both observations are correlated to observations at other time periods.
  - Example: Today's stock price can be correlated to the day before yesterday and yesterday can also be correlated to the day before yesterday. Then PACF of yesterday is the ”real” correlation between today and yesterday after taking out the influence of the day before yesterday.

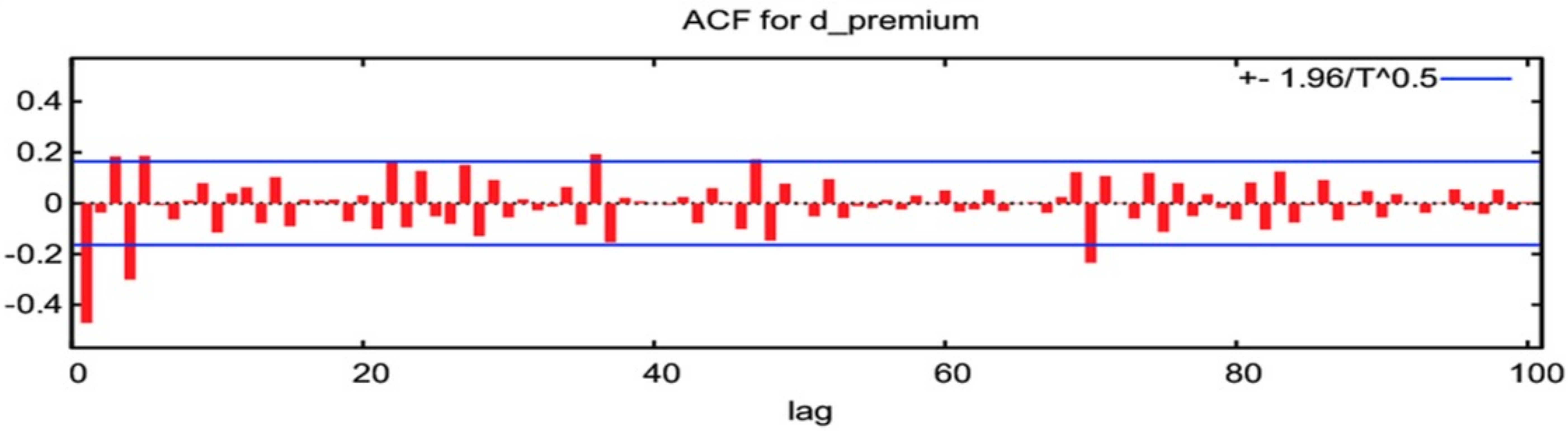
- The partial correlation between two variables is a conditional correlation taking into account their dependence on all other remaining variables
  - Eg. A third order (lag) partial autocorrelation is

$$\frac{\mathrm{Cov}(X_t, X_{t-3} | X_{t-1}, X_{t-2})}{\sqrt{\mathrm{Var}(X_t | X_{t-1}, X_{t-2})\mathrm{Var}(X_{t-3} | X_{t-1}, X_{t-2})}}$$

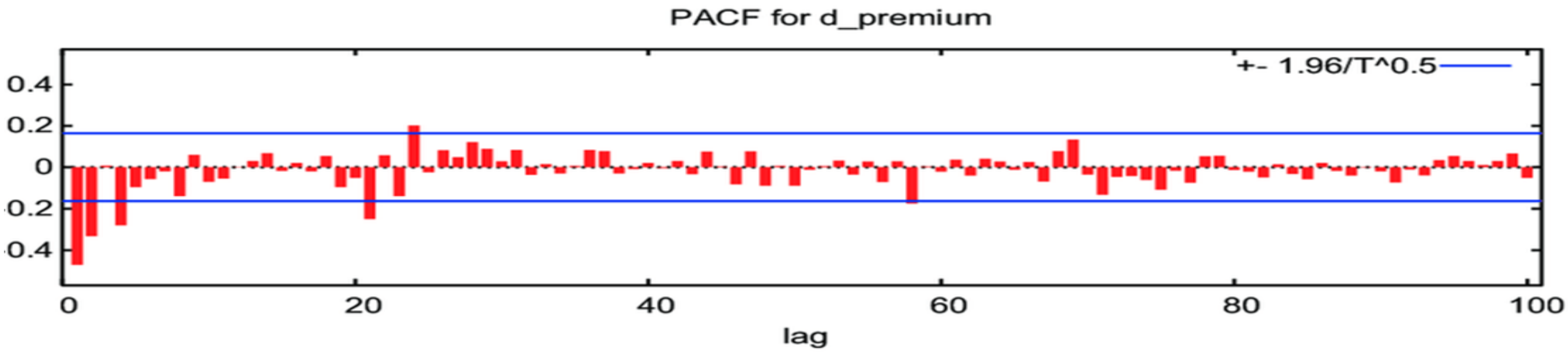
- The first order PACF & ACF are same

Time Series → Correlogram





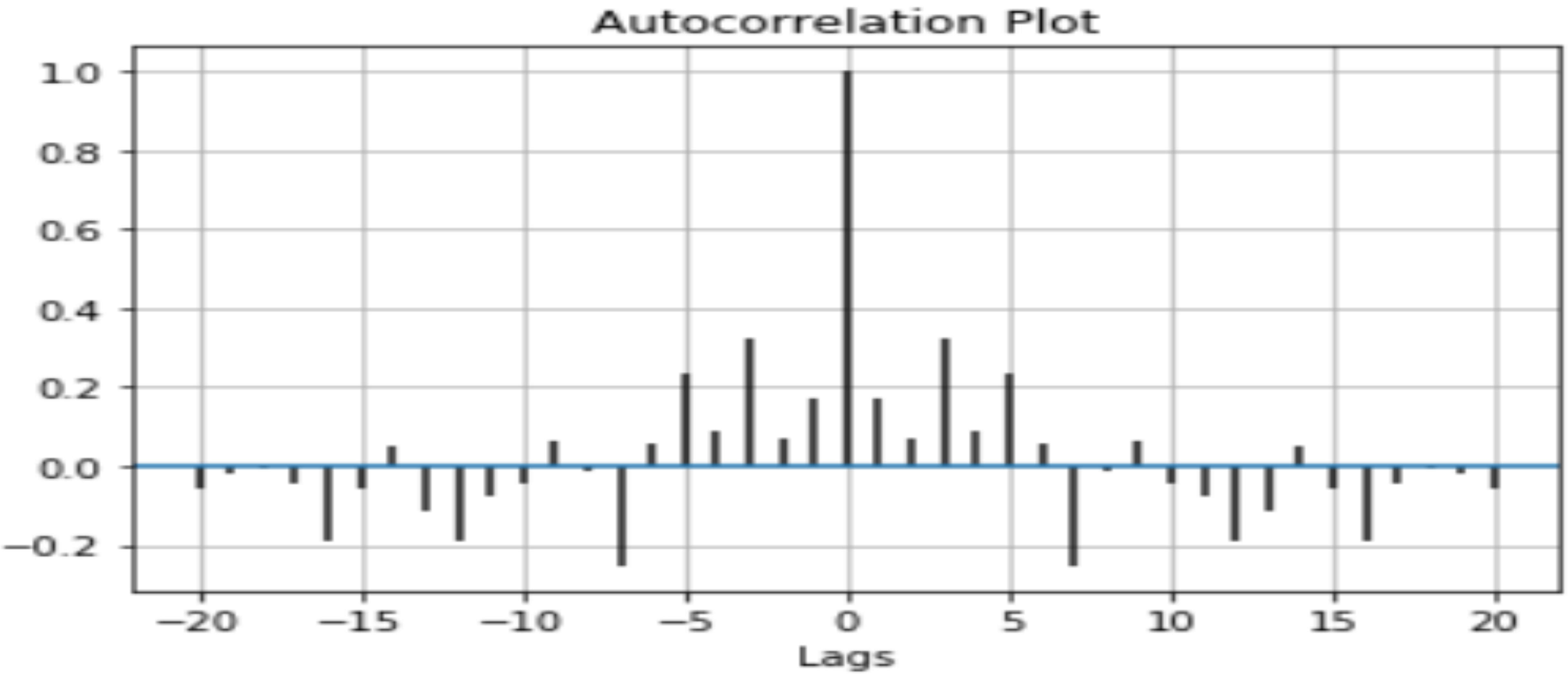
ACF for d_premium
+- 1.96/T^0.5
0.4
0.2
0
-0.2
-0.4
0
20
40
60
80
100
lag



PACF for d_premium
+- 1.96/T^0.5
0.4
0.2
0
0
20
40
60
80
100
lag

Time Series ⇨ Correlogram

The Autocorreleation plot for the data is:



Time Series ➡ Autoregressive (AR) models

- An **autoregressive (AR) model** is a representation of some type of random process; used to describe certain time-varying processes. The autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic (random) error term; thus the model is in the form of a stochastic difference equation.

- The Autoregressive model of order p denoted by AR (p) is given by

$$Y_t = c_t + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + \varepsilon_t$$

where $\phi_1, \phi_2, \phi_3, \ldots \phi_p$ are the parameter of the model; c is constant and ε_t is the **white noise**.

Time	Deflection	lag 1	lag 2	lag 3	lag 4	lag 5	lag 6	lag 7	lag 8	lag 9	lag 10
1.2001	-213										
2.2001	-564	-213									
3.2001	-35	-564	-213								
4.2001	-15	-35	-564	-213							
5.2001	141	-15	-35	-564	-213						
6.2001	115	141	-15	-35	-564	-213					
7.2001	-420	115	141	-15	-35	-564	-213				
8.2001	-360	-420	115	141	-15	-35	-564	-213			
9.2001	203	-360	-420	115	141	-15	-35	-564	-213		
10.2001	-338	203	-360	-420	115	141	-15	-35	-564	-213	
11.2001	-431	-338	203	-360	-420	115	141	-15	-35	-564	-213
12.2001	194	-431	-338	203	-360	-420	115	141	-15	-35	-564
1.2002	-220	194	-431	-338	203	-360	-420	115	141	-15	-35
2.2002	-513	-220	194	-431	-338	203	-360	-420	115	141	-15
3.2002	154	-513	-220	194	-431	-338	203	-360	-420	115	141
4.2002	-125	154	-513	-220	194	-431	-338	203	-360	-420	115
5.2002	-559	-125	154	-513	-220	194	-431	-338	203	-360	-420
6.2002	92	-559	-125	154	-513	-220	194	-431	-338	203	-360
7.2002	-21	92	-559	-125	154	-513	-220	194	-431	-338	203
8.2002	-579	-21	92	-559	-125	154	-513	-220	194	-431	-338
9.2002	-52	-579	-21	92	-559	-125	154	-513	-220	194	-431
10.2002	99	-52	-579	-21	92	-559	-125	154	-513	-220	194
11.2002	-543	99	-52	-579	-21	92	-559	-125	154	-513	-220
12.2002	-175	-543	99	-52	-579	-21	92	-559	-125	154	-513
1.2003	162	-175	-543	99	-52	-579	-21	92	-559	-125	154
2.2003	-457	162	-175	-543	99	-52	-579	-21	92	-559	-125
3.2003	-346	-457	162	-175	-543	99	-52	-579	-21	92	-559
4.2003	204	-346	-457	162	-175	-543	99	-52	-579	-21	92
5.2003	-300	204	-346	-457	162	-175	-543	99	-52	-579	-21
6.2003	-474	-300	204	-346	-457	162	-175	-543	99	-52	-579
7.2003	164	-474	-300	204	-346	-457	162	-175	-543	99	-52
8.2003	-107	164	-474	-300	204	-346	-457	162	-175	-543	99
9.2003	-572	-107	164	-474	-300	204	-346	-457	162	-175	-543
10.2003	-8	-572	-107	164	-474	-300	204	-346	-457	162	-175
11.2003	83	-8	-572	-107	164	-474	-300	204	-346	-457	162
12.2003	-541	83	-8	-572	-107	164	-474	-300	204	-346	-457
1.2004	-224	-541	83	-8	-572	-107	164	-474	-300	204	-346
2.2004	180	-224	-541	83	-8	-572	-107	164	-474	-300	204
3.2004	-420	180	-224	-541	83	-8	-572	-107	164	-474	-300
4.2004	-374	-420	180	-224	-541	83	-8	-572	-107	164	-474
5.2004	201	-374	-420	180	-224	-541	83	-8	-572	-107	164
6.2004	-236	201	-374	-420	180	-224	-541	83	-8	-572	-107
7.2004	-531	-236	201	-374	-420	180	-224	-541	83	-8	-572
8.2004	83	-531	-236	201	-374	-420	180	-224	-541	83	-8
9.2004	27	83	-531	-236	201	-374	-420	180	-224	-541	83
10.2004	-564	27	83	-531	-236	201	-374	-420	180	-224	-541
11.2004	-112	-564	27	83	-531	-236	201	-374	-420	180	-224
12.2004	131	-112	-564	27	83	-531	-236	201	-374	-420	180
1.2005	-507	131	-112	-564	27	83	-531	-236	201	-374	-420
2.2005	-254	-507	131	-112	-564	27	83	-531	-236	201	-374
3.2005	199	-254	-507	131	-112	-564	27	83	-531	-236	201
4.2005	-311	199	-254	-507	131	-112	-564	27	83	-531	-236
5.2005	-495	-311	199	-254	-507	131	-112	-564	27	83	-531
6.2005	143	-495	-311	199	-254	-507	131	-112	-564	27	83
7.2005	-46	143	-495	-311	199	-254	-507	131	-112	-564	27
8.2005	-579	-46	143	-495	-311	199	-254	-507	131	-112	-564
9.2005	-90	-579	-46	143	-495	-311	199	-254	-507	131	-112
10.2005	136	-90	-579	-46	143	-495	-311	199	-254	-507	131
11.2005	-472	136	-90	-579	-46	143	-495	-311	199	-254	-507
12.2005	-338	-472	136	-90	-579	-46	143	-495	-311	199	-254
1.2006	202	-338	-472	136	-90	-579	-46	143	-495	-311	199
2.2006	-287	202	-338	-472	136	-90	-579	-46	143	-495	-311
3.2006	-477	-287	202	-338	-472	136	-90	-579	-46	143	-495
4.2006	169	-477	-287	202	-338	-472	136	-90	-579	-46	143
5.2006	-124	169	-477	-287	202	-338	-472	136	-90	-579	-46
6.2006	-568	-124	169	-477	-287	202	-338	-472	136	-90	-579
7.2006	17	-568	-124	169	-477	-287	202	-338	-472	136	-90
8.2006	48	17	-568	-124	169	-477	-287	202	-338	-472	136
9.2006	-568	48	17	-568	-124	169	-477	-287	202	-338	-472
10.2006	-135	-568	48	17	-568	-124	169	-477	-287	202	-338
11.2006	162	-135	-568	48	17	-568	-124	169	-477	-287	202
12.2006	-430	162	-135	-568	48	17	-568	-124	169	-477	-287
1.2007	-422	-430	162	-135	-568	48	17	-568	-124	169	-477
2.2007	172	-422	-430	162	-135	-568	48	17	-568	-124	169
3.2007	-74	172	-422	-430	162	-135	-568	48	17	-568	-124
4.2007	-577	-74	172	-422	-430	162	-135	-568	48	17	-568
5.2007	-13	-577	-74	172	-422	-430	162	-135	-568	48	17
6.2007	92	-13	-577	-74	172	-422	-430	162	-135	-568	48
7.2007	-534	92	-13	-577	-74	172	-422	-430	162	-135	-568
8.2007	-243	-534	92	-13	-577	-74	172	-422	-430	162	-135
9.2007	194	-243	-534	92	-13	-577	-74	172	-422	-430	162
10.2007	-355	194	-243	-534	92	-13	-577	-74	172	-422	-430
11.2007	-465	-355	194	-243	-534	92	-13	-577	-74	172	-422
12.2007	156	-465	-355	194	-243	-534	92	-13	-577	-74	172
1.2008	-81	156	-465	-355	194	-243	-534	92	-13	-577	-74
2.2008	-578	-81	156	-465	-355	194	-243	-534	92	-13	-577
3.2008	-64	-578	-81	156	-465	-355	194	-243	-534	92	-13
4.2008	139	-64	-578	-81	156	-465	-355	194	-243	-534	92
5.2008	-449	139	-64	-578	-81	156	-465	-355	194	-243	-534
6.2008	-384	-449	139	-64	-578	-81	156	-465	-355	194	-243
7.2008	193	-384	-449	139	-64	-578	-81	156	-465	-355	194
8.2008	-198	193	-384	-449	139	-64	-578	-81	156	-465	-355
9.2008	-538	-198	193	-384	-449	139	-64	-578	-81	156	-465
10.2008	110	-538	-198	193	-384	-449	139	-64	-578	-81	156
11.2008	-44	110	-538	-198	193	-384	-449	139	-64	-578	-81
12.2008	-577	-44	110	-538	-198	193	-384	-449	139	-64	-578
1.2009	-6	-577	-44	110	-538	-198	193	-384	-449	139	-64
2.2009	66	-6	-577	-44	110	-538	-198	193	-384	-449	139
3.2009	-552	66	-6	-577	-44	110	-538	-198	193	-384	-449
4.2009	-164	-552	66	-6	-577	-44	110	-538	-198	193	-384
5.2009	[illegible]	-164	-552	66	-6	-577	-44	110	-538	-198	193
6.2009	[illegible]	[illegible]	-164	-552	66	-6	-577	-44	110	-538	-198
7.2009	[illegible]	[illegible]	[illegible]	-164	-552	66	-6	-577	-44	110	-538
8.2009	[illegible]	[illegible]	[illegible]	[illegible]	-164	-552	66	-6	-577	-44	110
9.2009	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-164	-552	66	-6	-577	-44
10.2009	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-164	-552	66	-6	-577
11.2009	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-164	-552	66	-6
12.2009	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-164	-552	66
1.2010	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-164	-552
2.2010	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-164
3.2010	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4.2010	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5.2010	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6.2010	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7.2010	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
8.2010	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9.2010	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10.2010	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
11.2010	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
12.2010	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.2011	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2.2011	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3.2011	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4.2011	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5.2011	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6.2011	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7.2011	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
8.2011	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9.2011	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10.2011	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
11.2011	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
12.2011	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.2012	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2.2012	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3.2012	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4.2012	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5.2012	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6.2012	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7.2012	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
8.2012	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
9.2012	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10.2012	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
11.2012	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
12.2012	190	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.2013	-391	190	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
2.2013	-406	-391	190	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3.2013	194	-406	-391	190	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
4.2013	-186	194	-406	-391	190	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
5.2013	-553	-186	194	-406	-391	190	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6.2013	83	-553	-186	194	-406	-391	190	[illegible]	[illegible]	[illegible]	[illegible]
7.2013	-13	83	-553	-186	194	-406	-391	190	[illegible]	[illegible]	[illegible]
8.2013	-577	-13	83	-553	-186	194	-406	-391	190	[illegible]	[illegible]
9.2013	-49	-577	-13	83	-553	-186	194	-406	-391	190	[illegible]
10.2013	103	-49	-577	-13	83	-553	-186	194	-406	-391	190
11.2013	-515	103	-49	-577	-13	83	-553	-186	194	-406	-391
12.2013	-280	-515	103	-49	-577	-13	83	-553	-186	194	-406
1.2014	201	-280	-515	103	-49	-577	-13	83	-553	-186	194
2.2014	300	201	-280	-515	103	-49	-577	-13	83	-553	-186
3.2014	-506	300	201	-280	-515	103	-49	-577	-13	83	-553
4.2014	131	-506	300	201	-280	-515	103	-49	-577	-13	83
5.2014	-45	131	-506	300	201	-280	-515	103	-49	-577	-13
6.2014	-578	-45	131	-506	300	201	-280	-515	103	-49	-577
7.2014	-80	-578	-45	131	-506	300	201	-280	-515	103	-49
8.2014	138	-80	-578	-45	131	-506	300	201	-280	-515	103
9.2014	-462	138	-80	-578	-45	131	-506	300	201	-280	-515
10.2014	-361	-462	138	-80	-578	-45	131	-506	300	201	-280
11.2014	201	-361	-462	138	-80	-578	-45	131	-506	300	201
12.2014	-211	201	-361	-462	138	-80	-578	-45	131	-506	300
1.2015	-554	-211	201	-361	-462	138	-80	-578	-45	131	-506
2.2015	32	-554	-211	201	-361	-462	138	-80	-578	-45	131
3.2015	74	32	-554	-211	201	-361	-462	138	-80	-578	-45
4.2015	-533	74	32	-554	-211	201	-361	-462	138	-80	-578
5.2015	-235	-533	74	32	-554	-211	201	-361	-462	138	-80
6.2015	187	-235	-533	74	32	-554	-211	201	-361	-462	138
7.2015	-372	187	-235	-533	74	32	-554	-211	201	-361	-462
8.2015	-442	-372	187	-235	-533	74	32	-554	-211	201	-361
9.2015	182	-442	-372	187	-235	-533	74	32	-554	-211	201
10.2015	-147	182	-442	-372	187	-235	-533	74	32	-554	-211
11.2015	-566	-147	182	-442	-372	187	-235	-533	74	32	-554
12.2015	25	-566	-147	182	-442	-372	187	-235	-533	74	32
1.2016	68	25	-566	-147	182	-442	-372	187	-235	-533	74
2.2016	-535	68	25	-566	-147	182	-442	-372	187	-235	-533
3.2016	-244	-535	68	25	-566	-147	182	-442	-372	187	-235
4.2016	194	-244	-535	68	25	-566	-147	182	-442	-372	187
5.2016	-351	194	-244	-535	68	25	-566	-147	182	-442	-372
6.2016	-463	-351	194	-244	-535	68	25	-566	-147	182	-442
7.2016	174	-463	-351	194	-244	-535	68	25	-566	-147	182
8.2016	-125	174	-463	-351	194	-244	-535	68	25	-566	-147
9.2016	-570	-125	174	-463	-351	194	-244	-535	68	25	-566
10.2016	15	-570	-125	174	-463	-351	194	-244	-535	68	25
11.2016	72	15	-570	-125	174	-463	-351	194	-244	-535	68
12.2016	-550	72	15	-570	-125	174	-463	-351	194	-244	-535
1.2017	-190	-550	72	15	-570	-125	174	-463	-351	194	-244
2.2017	172	-190	-550	72	15	-570	-125	174	-463	-351	194
3.2017	-424	172	-190	-550	72	15	-570	-125	174	-463	-351
4.2017	-385	-424	172	-190	-550	72	15	-570	-125	174	-463
5.2017	198	-385	-424	172	-190	-550	72	15	-570	-125	174
6.2017	-218	198	-385	-424	172	-190	-550	72	15	-570	-125
7.2017	-536	-218	198	-385	-424	172	-190	-550	72	15	-570
8.2017	96	-536	-218	198	-385	-424	172	-190	-550	72	15

Regression Statistics	Results
Multiple R	0.972056
R Square	0.944892
Adjusted R Square	0.941813
Standard Error	67.17728
Observations	190

	Coefficients	Standard Error	t Stat	P-value
Intercept	-301.643	57.0362	-5.28862	3.56E-07
lag 1	-0.27739	0.074602	-3.71825	0.000268
lag 2	-0.55811	0.076787	-7.2683	1.09E-11
lag 3	0.356679	0.086535	4.12177	5.74E-05
lag 4	**0.013823**	**0.090524**	**0.152697**	**0.878809**
lag 5	-0.14453	0.089015	-1.62366	0.106209
lag 6	-0.02001	0.078551	-0.25476	0.7992
lag 7	0.048186	0.074732	0.644783	0.519894
lag 8	-0.01276	0.072079	-0.17699	0.859713
lag 9	-0.04807	0.057523	-0.83569	0.404441
lag 10	-0.04469	0.057081	-0.78286	0.434743

ANOVA

SV	df	SS	MS	F-value	P-value
Regression	10	13850518	1385052	306.9172	5.2E-107
Residual	179	807788.9	4512.787		
Total	189	14658307			

Regression Statistics	Results
Multiple R	0.965633
R Square	0.932447
Adjusted R Square	0.93066
Standard Error	73.23764
Observations	195

ANOVA

SV	df	SS	MS	F-value	P-value
Regression	5	13992985	2798597	521.7611	1.6E-108
Residual	189	1013749	5363.752		
Total	194	15006735			

	Coefficients	Standard Error	t Stat	P-value
Intercept	-288.389	34.1692	-8.44002	8.17E-15
lag 1	-0.11143	0.06172	-1.80543	0.072598
lag 2	-0.66093	0.061539	-10.7399	2.62E-21
lag 3	0.388821	0.072508	5.362419	2.38E-07
lag 5	-0.12262	0.061361	-1.99838	0.047109

$Y_t = b_0+b_1X_{t-1}+b_2X_{t-2}+b_3X_{t-3}+b_5X_{t-5}$

$Y_t = -288.389 - 0.11143 * X_{t-1} - 0.66093 * X_{t-2} + 0.388821 * X_{t-3} - 0.12262 * X_{t-5}$

Regression Statistics	Results
Multiple R	0.965633
R Square	0.932447
Adjusted R Square	0.93066
Standard Error	73.23764
Observations	195

ANOVA

SV	df	SS	MS	F-value	P-value
Regression	5	13992985	2798597	521.7611	1.6E-108
Residual	189	1013749	5363.752		
Total	194	15006735			

	Coefficients	Standard Error	t Stat	P-value
Intercept	-288.389	34.1692	-8.44002	8.17E-15
Lag 1	-288.389	34.1692	-8.44002	8.17E-15
Lag 2	-0.11143	0.06172	-1.80543	0.072598
Lag 3	-0.66093	0.061539	-10.7399	2.62E-21
Lag 4	0.388821	0.072508	5.362419	2.38E-07
Lag 5	-0.11281	0.061297	-1.84034	0.067286

$Y_t = b_0+b_1X_{t-1}+b_2X_{t-2}+b_3X_{t-3}+b_4X_{t-4}+b_5X_{t-5}$

$Y_t = -288.389 - 0.11143 *X_{t-1} - 0.66093*X_{t-2} + 0.388821*X_{t-3} - 0.11281*X_{t-4} - 0.12262*X_{t-5}$

Period (t)	Sales (Y_t)	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7
1	5.30							
2	4.40	5.30						
3	5.40	4.40	5.30					
4	5.80	5.40	4.40	5.30				
5	5.60	5.80	5.40	4.40	5.30			
6	4.80	5.60	5.80	5.40	4.40	5.30		
7	5.60	4.80	5.60	5.80	5.40	4.40	5.30	
8	5.60	5.60	4.80	5.60	5.80	5.40	4.40	5.30
9	5.40	5.60	5.60	4.80	5.60	5.80	5.40	4.40
10	6.50	5.40	5.60	5.60	4.80	5.60	5.80	5.40
11	5.10	6.50	5.40	5.60	5.60	4.80	5.60	5.80
12	5.80	5.10	6.50	5.40	5.60	5.60	4.80	5.60
13	5.00	5.80	5.10	6.50	5.40	5.60	5.60	4.80
14	6.20	5.00	5.80	5.10	6.50	5.40	5.60	5.60
15	5.60	6.20	5.00	5.80	5.10	6.50	5.40	5.60
16	6.70	5.60	6.20	5.00	5.80	5.10	6.50	5.40
17	5.20	6.70	5.60	6.20	5.00	5.80	5.10	6.50
18	5.50	5.20	6.70	5.60	6.20	5.00	5.80	5.10
19	5.80	5.50	5.20	6.70	5.60	6.20	5.00	5.80
20	5.10	5.80	5.50	5.20	6.70	5.60	6.20	5.00
21	5.80	5.10	5.80	5.50	5.20	6.70	5.60	6.20
22	6.70	5.80	5.10	5.80	5.50	5.20	6.70	5.60
23	5.20	6.70	5.80	5.10	5.80	5.50	5.20	6.70
24	6.00	5.20	6.70	5.80	5.10	5.80	5.50	5.20
25	5.80	6.00	5.20	6.70	5.80	5.10	5.80	5.50
		5.80	6.00	5.20	6.70	5.80	5.10	5.80
			5.80	6.00	5.20	6.70	5.80	5.10
				5.80	6.00	5.20	6.70	5.80
					5.80	6.00	5.20	6.70
						5.80	6.00	5.20
							5.80	6.00
								5.80

Lag	Intercept	Coefficient
1	8.502	-0.486
2	4.814	0.161
3	6.408	-0.121
4	6.617	-0.159
5	6.246	-0.094
6	3.381	0.421
7	5.816	-0.017

Time Series ⇨ Test for stationarity using Dickey-Fuller test

- **Consider the first order autoregressive model**

$Y_t = \mu + \phi_1 Y_{t-1} + \epsilon_t$ ---------- (1)

where μ is constant, ϕ_1 is the regression coefficient and ϵ_t is the random error with zero mean and constant variance

If Y_t is a non-stationary process, then it is essential to test ϕ_1 whether it has unitary root or not.

Hence, the hypothesis to be tested is

H_0: ϕ_1=1 (Unitary root) against H_1: ϕ_1 < 1 (No unitary root)

Time Series ⇨ Test for stationarity using Dickey-Fuller test

- **By rewriting the model (1)**

$Y_t - Y_{t-1} = \mu + (\phi_1 - 1)\, Y_{t-1} + \epsilon_t$. ---------(2)

Denoting $\Delta Yt = Y_t - Y_{t-1}$ the model (2) may be expressed as

$\Delta Y_t = \mu + \delta Y_{t-1} + \epsilon_t$ ----------- (3)

Hence, the hypothesis to be tested is

H_0: $\delta = 0$ against H_1: $\delta < 0$

$$\boldsymbol{The\ test - statistic\colon\ t_{\hat{\delta}} = \frac{\hat{\delta}}{SE(\hat{\delta})}}$$

$$\boldsymbol{The\ test - statistic\colon\ t_{\hat{\delta}} = \frac{\hat{\delta}}{SE(\hat{\delta})}}$$

Will follow Dickey – Fuller distribution

Reject H0, if $t_{\hat{\delta}}$ < DF – CV (P ≤ 0.05)
and

Fail to reject H0, if $t_{\hat{\delta}}$ > DF – CV (P > 0.05)

- **Consider the 'p' order autoregressive model**

$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \cdots + \phi_p Y_{t-p} + \epsilon_t$ ---------- (1)

where μ is constant, ϕ_1, ϕ_2 , ϕ_3 , …, ϕ_p are the regression coefficients and ϵ_t is the random error with zero mean and constant variance

If Y_t is a non-stationary process, then it is essential to test ϕ_1, ϕ_2 , ϕ_3 , …, ϕ_p whether they haveunitary root or not.

Hence, the hypothesis to be tested is

H_0: ϕ_i=1 (Unitary root) against H_1: ϕ_i < 1 (No unitary root) for all i = 1, 2, 3, …, p

Time Series → Test for stationarity using Dickey-Fuller test

- **By rewriting the model (1)**

$$Y_t - Y_{t-1} = \mu + (\phi_1 - 1)\, Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \cdots + \phi_p Y_{t-p} + \epsilon_t \text{ -------(2)}$$

Denoting $\Delta Yt = Y_t - Y_{t-1}$ the model (2) may be expressed as

$$\Delta Y_t = \mu + \delta Y_{t-1} + \sum_{i=2}^{p} \beta_i Y_{t-i} + \epsilon_t \text{ --------}$$

Hence, the hypothesis to be tested is

H$_0$: δ = 0 against H$_1$: δ < 0

$$\boldsymbol{The\ test-statistic:\ t_{\hat{\delta}} = \frac{\hat{\delta}}{SE(\hat{\delta})}}$$

$$\boldsymbol{The\ test-statistic:\ t_{\hat{\delta}} = \frac{\hat{\delta}}{SE(\hat{\delta})}}$$

Will follow Dickey – Fuller distribution

Reject H0, if $t_{\hat{\delta}}$ < DF (CV) (P ≤ 0.05)
and
Fail to reject H0, if $t_{\hat{\delta}}$ > DF (CV) (P > 0.05)

innovate achieve lead

Time Series ⇨ Test for stationarity using Dickey-Fuller test

$$\Delta Y_t = \mu + \delta Y_{t-1} + \sum_{i=2}^{p} \beta_i Y_{t-i} + \epsilon_t \text{ ----------- (3)}$$

Like testing for δ, there is a need to test other parameters β_i

H_0: $\beta_i = 0$ against H_1: $\beta_i < 0$ for all i = 1, 2, 3, …, p

$$The\ test - statistic: t_{\widehat{\beta}_i} = \frac{\widehat{\beta}_i}{SE(\widehat{\beta}_i)}$$

Hence, the hypothesis to be tested is

H_0: $\delta = 0$ against H_1: $\delta < 0$

$$The\ test - statistic: t_{\widehat{\beta}_i} = \frac{\widehat{\beta}_i}{SE(\widehat{\beta}_i)}$$

Will follow Student's t – distribution with p-1 degrees of freedom

Reject H0, if $t_{\widehat{\beta}_i}$ > t (CV) (P ≤ 0.05) and

Fail to reject H0, if $t_{\widehat{\delta}}$ < t (CV) (P > 0.05)

Time Series ⇨ Moving Average (MA) Model

- **Moving-average model**: The **moving-average model** specifies that the output variable depends linearly on the **current and various past** values of a stochastic term.

- Moving average model of order q (MA(q)):

$$Y_t = w_t + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \cdots + \theta_q Y_{t-q} + \varepsilon_t$$

- **Moving-average model** of order q (MA(q)):

$$X_t = w_t + \sum_{i=1}^{q} \theta_i w_{t-q} + \varepsilon_t$$

where: $\theta_1, \theta_2, \ldots, \theta_q$ are constants with $\theta_q \neq 0$; and w_t is **Gaussian white noise** $w_t\,(0, \sigma^2{}_w)$.

Note: Gaussian noise, named after Carl Friedrich **Gauss**, is **statistical noise** having a probability density function (PDF) equal to that of the normal distribution, which is also known as the **Gaussian** distribution. In other words, the values that the **noise** can take on are **Gaussian** distributed.

Time Series ⇨ Autoregressive Moving Average (ARMA) Model

- **Autoregressive–moving-average model**. In the statistical analysis of time series Auto-Regressive Moving Average (ARMA) models provide a parsimonious description of a (weakly) stationary stochastic process in terms of two polynomials, one for the **autoregression** (AR) and the second for the **moving average** (MA).

Time Series ⇨ Autoregressive Moving Average (ARMA) Model

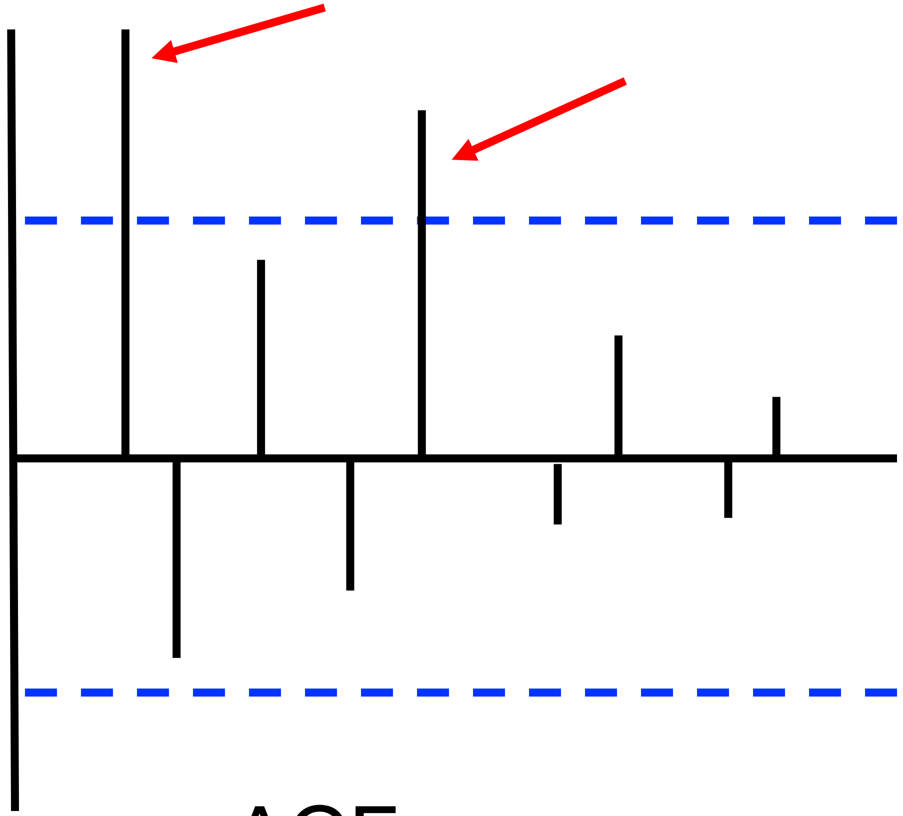
- The AR and MA models dynamics can be combined into what is called an *autoregressive moving-average (ARMA) model*.
- The ARMA (1, 1) is $Y_t = \phi_0 + \phi_1 Y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$
    - ϕ_0 – Constant term in AR model
    - ϕ_1 – Coefficient associated with Y_{t-1} in AR model
    - θ_1 – Coefficient associated with ϵ_{t-1} in MA model
    - ϵ_{t-1} – Error while measuring Y_{t-1} in MA model
    - ϵ_t – Error while measuring Y_t in MA model

- The predicted value for the ARMA (1, 1) is

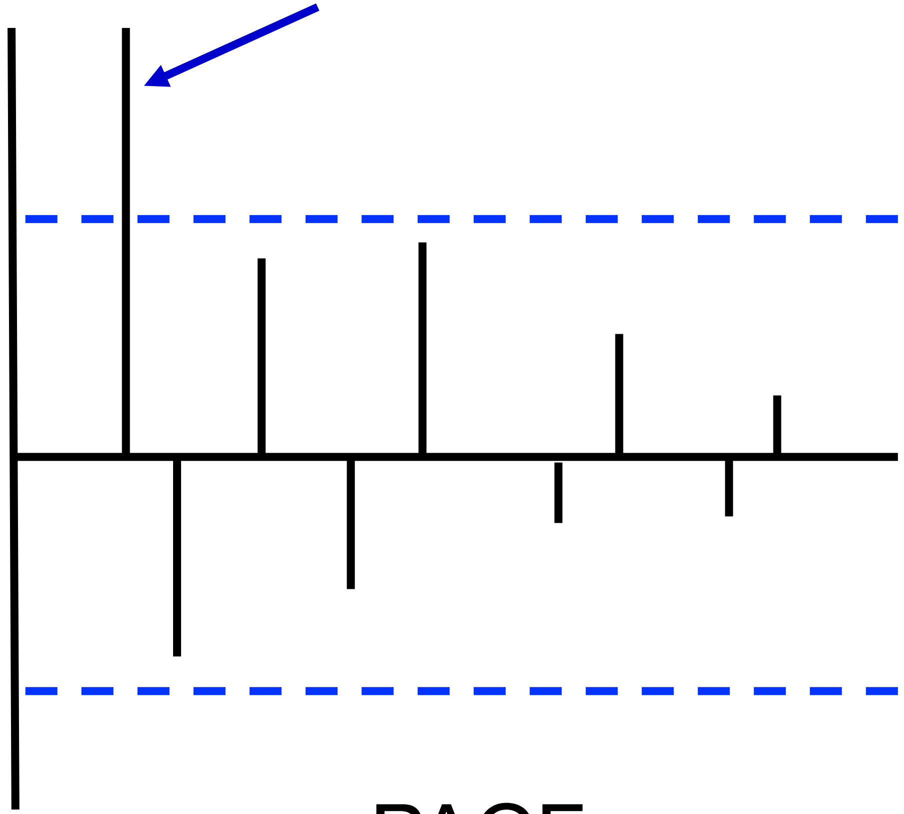
$$\hat{Y}_t = \phi_0 + \phi_1 Y_{t-1} + \theta_1 \epsilon_{t-1}$$

- Here ϕ_0, ϕ_1, and ϵ_{t-1} are all known except ϵ_t – the error of the current time
- With the help of ACF which helps to identify the MA order and PCAF helps to identify the AR model the ARMA (p, q) model is given by

- The ACF and PACF are one of the many ways to use to decide the order of ARMA model



ACF
MA order is 2



PACF
AR order is 1

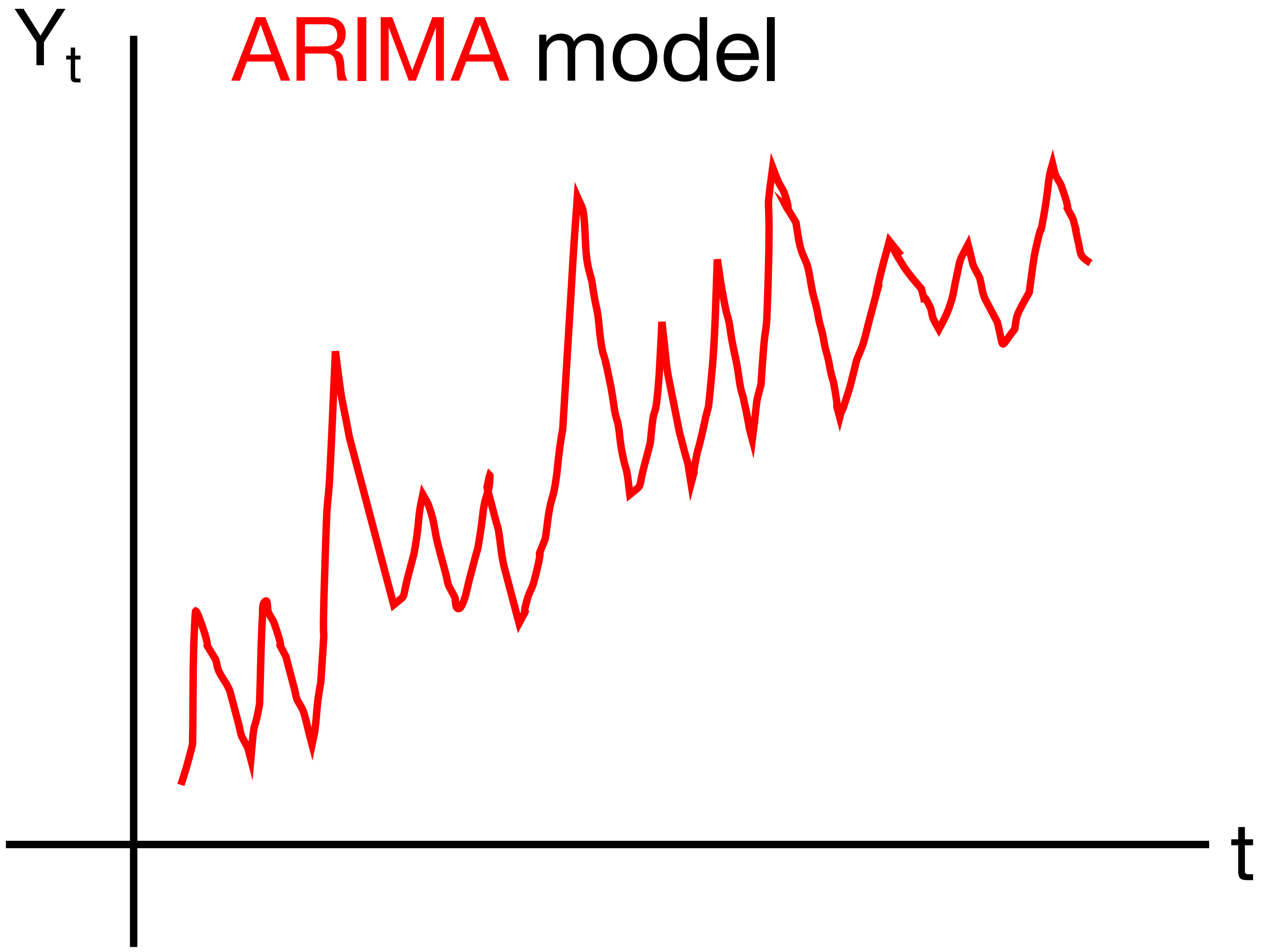
- Higher order ARMA processes involve additional lags of X and epsilon.

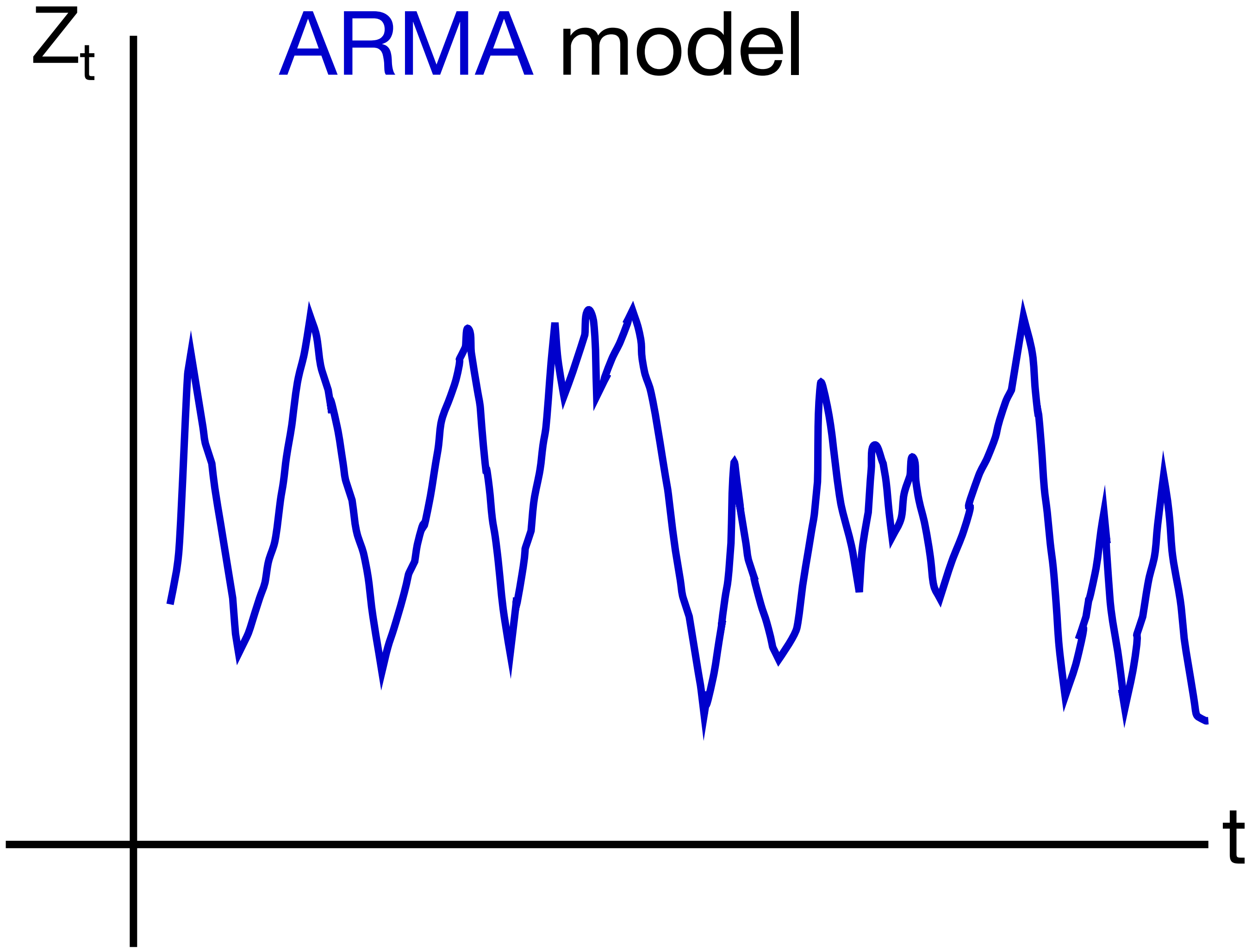
- The ARMA (p, q) is

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p}$$
$$+ \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}, \quad or$$

$$\Phi(L) y_t = \Theta(L) \varepsilon_t$$

- ARIMA is an acronym that stands for Auto-Regressive Integrated Moving Average. Specifically,
    - AR ➡ Auto-regression: A model that uses the dependent relationship an observation and some number of **lagged observations**
    - I ➡ Integrated: The use of **differencing** of raw observations in order to make the time series stationary
    - MA ➡ Moving Average: A model that uses the dependency between an observation and **residual error** from a moving average model applied to lagged observations





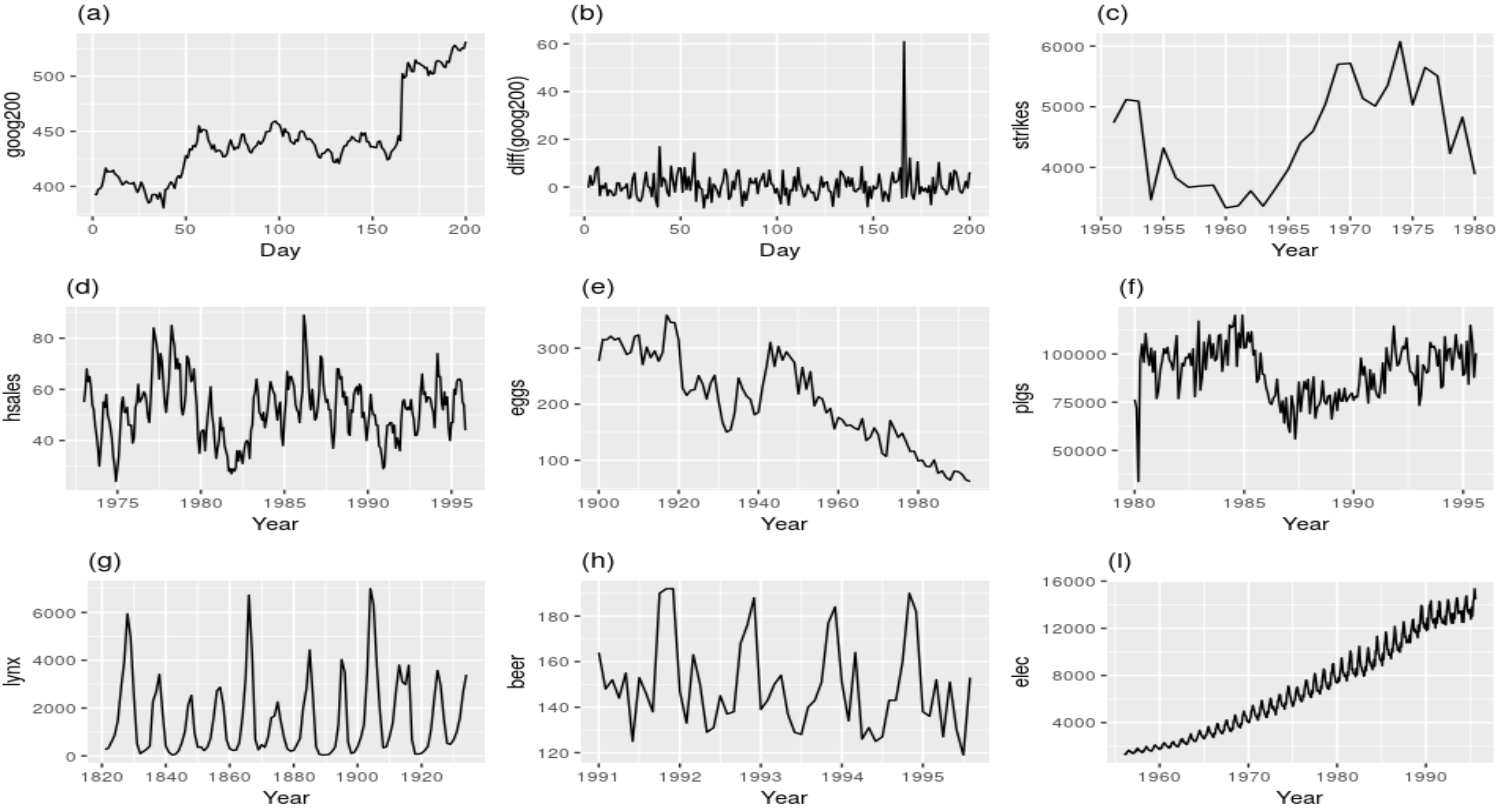
Non-stationary – Mean not same

Stationary – Mean same

The Stationary model now is $Y_k = \sum_{i=1}^{k-l} Z_{k-i} + a_l$

- Let Y_t, Y_{t-1}, Y_{t-2}, …, Y_{t-p} are the ‘p’ non-stationary time series data.
- Define $Z_t = Y_{t+1} – Y_t$ be the first order differencing between two successive values.
- The Z_t will now be transformed into stationary data

- Each of these are explicitly specified in the model as a parameter.
- Note that AR and MA are two widely used **linear models** that work on stationary time series and I is a **pre-processing to stationarize** time series if needed.


(a)
goog200
500
450
400
0
50
100
150
200
Day
(b)
diff(goog200)
60
40
20
0
0
50
100
150
200
Day
(c)
strikes
6000
5000
4000
1950
1955
1960
1965
1970
1975
1980
Year
(d)
hsales
80
60
40
1975
1980
1985
1990
1995
Year
(e)
eggs
300
200
100
1900
1920
1940
1960
1980
Year
(f)
pigs
100000
75000
50000
1980
1985
1990
1995
Year
(g)
lynx
6000
4000
2000
0
1820
1840
1860
1880
1900
1920
Year
(h)
beer
180
160
140
120
1991
1992
1993
1994
1995
Year
(i)
elec
16000
12000
8000
4000
1960
1970
1980
1990
Year

- **Rationale** - The first task is to provide a reason why we're interested in a particular model, as quants. Why are we introducing the time series model? What effects can it capture? What do we gain (or lose) by adding in extra complexity?
- **Definition** - We need to provide the full mathematical definition (and associated notation) of the time series model in order to minimise any ambiguity

Time Series ⇨ Choice of p, d and q

- **Second Order Properties** - We will discuss (and in some cases derive) the second order properties of the time series model, which includes its mean, its variance and its autocorrelation function
- **Correlogram** - We will use the second order properties to plot a correlogram of a realisation of the time series model in order to visualise its behaviour.

Time Series ➡ Choice of p, d and q

- **Simulation** - We will simulate realisations of the time series model and then fit the model to these simulations to ensure we have accurate implementations and understand the fitting process.

Time Series ⇨ Choice of p, d and q

- Look at autocorrelation graph of data (will help if MA model is appropriate).
- Look at partial autocorrelation graph of data (will help if MR model is appropriate).
- Look at extended autocorrelation chart of data (will help if a combination of MA and AR models is needed).

Time Series ➡ Choice of p, d and q

- Try Akaike's information criterion (AIC) on a set of models and investigate the models with the lowest AIC values
- Try the Schwartz Bayesian information criterion (BIC) and investigate the models with the lowest BIC values
- All the above criterion to choose p, d and q are available as package in R.

Time Series ➡ SARIMA model

- Autoregressive Integrated Moving Average, or ARIMA, is one of the most widely used forecasting methods for univariate time series data forecasting.
- Although the method can handle data with a trend, it does not support time series with a seasonal component.
- An extension to ARIMA that supports the direct modeling of the seasonal component of the series is called Seasonal Autoregressive Integrated Moving Average model (SARIMA).

Time Series ➡ SARIMAX model

- SARIMAX(Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors) is **an updated version of the ARIMA model**. ARIMA includes an autoregressive integrated moving average, while SARIMAX includes seasonal effects and eXogenous factors with the autoregressive and moving average component in the model.

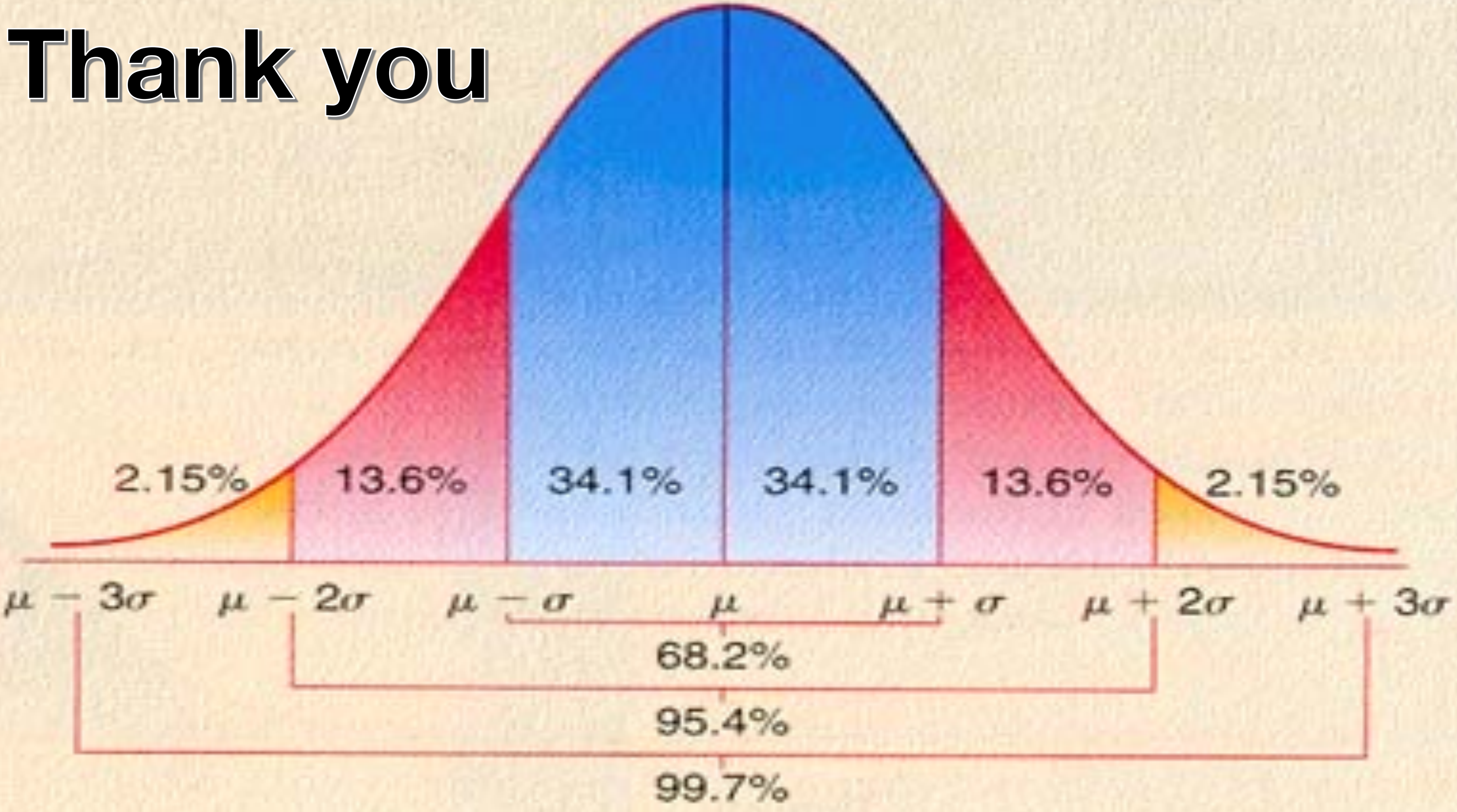
Time Series ➡ VAR and VARMAX models

- VAR models (vector autoregressive models) are used for multivariate time series. The structure is that each variable is a linear function of past lags of itself and past lags of the other variables in the system.
- The VARMAX procedure **enables you to model the dynamic relationship both between the dependent variables and also between the dependent and independent variables**. VARMAX models are defined in terms of the orders of the autoregressive or moving-average process (or both).

Reference

- Peter J Brockwell and Richard A Davis. Introduction to Time Series and Forecasting, 2/e, Springer
- Douglas C. Montgomery, and Cheryl L. Jennings, Murat Kulahcin. *Introduction to* Time Series Analysis and Forecasting, 2/e, Wiley

Thank you


2.15%
13.6%
34.1%
34.1%
13.6%
2.15%
$\mu - 3\sigma$
$\mu - 2\sigma$
$\mu - \sigma$
μ
$\mu + \sigma$
$\mu + 2\sigma$
$\mu + 3\sigma$
68.2%
95.4%
99.7%