Time Series Analysis: Part 1

Fan Liu
liufan@nju.edu.cn
School of Management & Engineering

Forecasting and Time Series Analysis

Forecasting is an important problem that spans many fields including business and industry, government, economics, medicine, finance and so on.

Forecasting is typically based on identifying, modeling, and extrapolating the patterns found in historical data.

Most forecasting problems involve the use of time series data. Time series data, or time series, usually refers to a set of observations collected sequentially in time.

An intrinsic feature of time series data is that, typically, adjacent observations are dependent. Time series analysis is concerned with techniques for the analysis of dependence.

Time Series Data

Time series data, or time series, usually refers to a set of observations collected sequentially in time.

Is this a time series data?

R	Name	Team	Pos	Apps	Goal	A	Yel	Red	SpG	PS%	AW	MoM	Rt
1	Lionel Messi	Barcelona	AM(R),FW	20(1)	33	8	1	-	5.3	85.3	0.1	14	9.02
2	Franck Ribéry	Bayern Munich	AM(L)	14	4	7	1	-	2.9	86.8	0.4	7	8.65
3	Cristiano Ronaldo	Real Madrid	AM(LR),FW	20	21	4	5	-	7.4	77.1	1.5	5	8.12
4	Andrea Pirlo	Juventus	DM(C)	18	5	3	6	-	2.1	87.2	0.6	4	8.05
5	Diego	Wolfsburg	AM(C)	18	5	4	2	-	2.4	82.9	0.5	7	7.98
6	Stefan Kießling	Bayer Leverkusen	FW	19	13	3	1	-	3.5	63.7	5.3	6	7.94
7	Zlatan Ibrahimovic	Paris Saint G...	FW	19	19	4	4	1	5.1	77.2	1.9	8	7.93
8	Edinson Cavani	Napoli	FW	19	18	1	3	-	4.4	76.3	0.8	5	7.88
9	Thomas Müller	Bayern Munich	AM(CLR),FW	17(1)	10	8	-	-	2.8	78.2	1.7	4	7.87
10	Luis Suárez	Liverpool	AM(CL),FW	23	17	3	7	-	5.8	77.1	0.4	5	7.84
11	Andrea Ranoc...	Inter	D(C)	19	1	1	5	-	0.7	86.5	3.8	1	7.83
12	Giorgio Chiellini	Juventus	D(CL)	13	-	2	3	-	0.5	90.4	3.2	-	7.83
13	Erik Lamela	Roma	AM(CR),FW	15(2)	10	1	4	-	3.1	81.5	0.4	5	7.77
14	Hernanes	Lazio	AM(C)	18(3)	8	3	2	1	3.2	85.2	1.5	7	7.75

It is a cross sectional data.

Time Series Data

Is this a time series data?

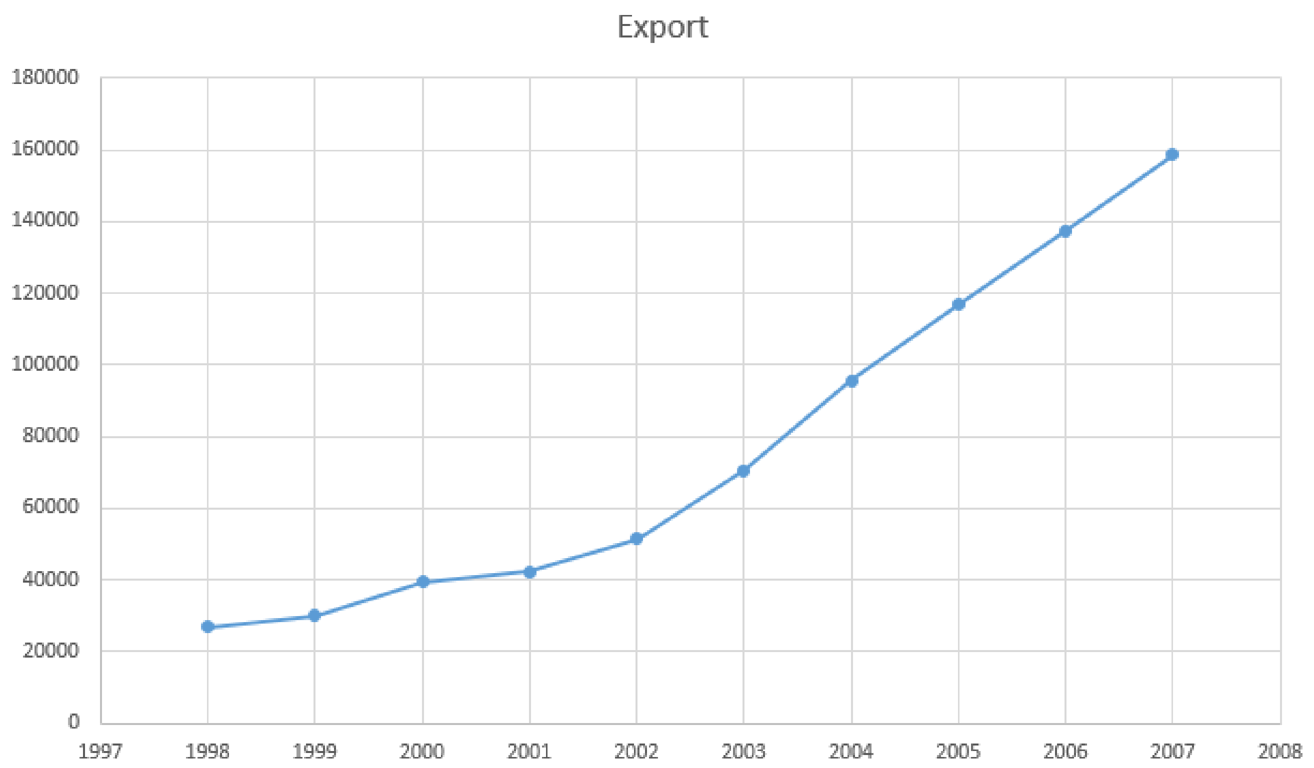
◇	A	B	C	D	E	F	G	H	I
1	COUNTRY_NAME	IND1_DESC	1995	1996	1997	1998	1999	2000	2001
2	Sweden	Commercial service imports (current US$)	1.71E+10	1.87E+10	1.94E+10	2.16E+10	2.25E+10	2.34E+10	2.29E+10
3	Sweden	Domestic credit to private sector (% of GDP)	102.7035	101.1959	102.6803	103.4841	104.6534	45.68565	..
4	Sweden	GDP (current LCU)	1.71E+12	1.76E+12	1.82E+12	1.91E+12	2.00E+12	2.10E+12	2.17E+12
5	Sweden	Population, total	8831000	8843000	8849440	8851800	8857400	8869000	8894000
6	Switzerland	Commercial service imports (current US$)	1.49E+10	1.56E+10	1.40E+10	1.50E+10	1.58E+10	1.55E+10	1.52E+10
7	Switzerland	Domestic credit to private sector (% of GDP)	168.3593	166.2425	168.4685	167.2875	174.0978	165.4197	158.5018
8	Switzerland	GDP (current LCU)	3.63E+11	3.66E+11	3.71E+11	3.80E+11	3.89E+11	4.04E+11	4.17E+11
9	Switzerland	Population, total	7041000	7074000	7088000	7110000	7140000	7180000	7231000
10	United Kingdom	Commercial service imports (current US$)	6.23E+10	6.84E+10	7.42E+10	8.38E+10	9.10E+10	9.59E+10	9.18E+10
11	United Kingdom	Domestic credit to private sector (% of GDP)	115.3449	119.6871	120.0104	119.0575	122.0526	133.6535	138.8419
12	United Kingdom	GDP (current LCU)	7.19E+11	7.62E+11	8.11E+11	8.60E+11	9.01E+11	9.45E+11	9.89E+11
13	United Kingdom	Population, total	5.83E+07	5.83E+07	5.84E+07	5.85E+07	5.86E+07	5.87E+07	5.88E+07
14	United States	Commercial service imports (current US$)	1.29E+11	1.38E+11	1.52E+11	1.68E+11	1.74E+11	2.03E+11	1.93E+11
15	United States	Domestic credit to private sector (% of GDP)	104.1564	111.8284	121.4929	132.2526	146.1111	145.6414	145.7946
16	United States	GDP (current LCU)	7.34E+12	7.75E+12	8.26E+12	8.72E+12	9.21E+12	9.81E+12	1.01E+13
17	United States	Population, total	2.65E+08	2.68E+08	2.72E+08	2.75E+08	2.79E+08	2.82E+08	2.85E+08

It is a panel data. Or you can treat it as a multivariate time series data.

Time Series Data

Is this a time series data?

Year	Export
1998	26854.1
1999	29896.3
2000	39273.2
2001	42183.6
2002	51378.2
2003	70483.5
2004	95539.1
2005	116921.8
2006	137487.8
2007	158787.4



It is the export value of China from 1998 to 2007.

Discrete VS Continuous

In the previous time series, the observations are taken at predetermined, equal interval time points. So that one might get hourly, daily, weekly, quarterly, or yearly readings. Such data form a *discrete* time series. Most time series we face are in such format.

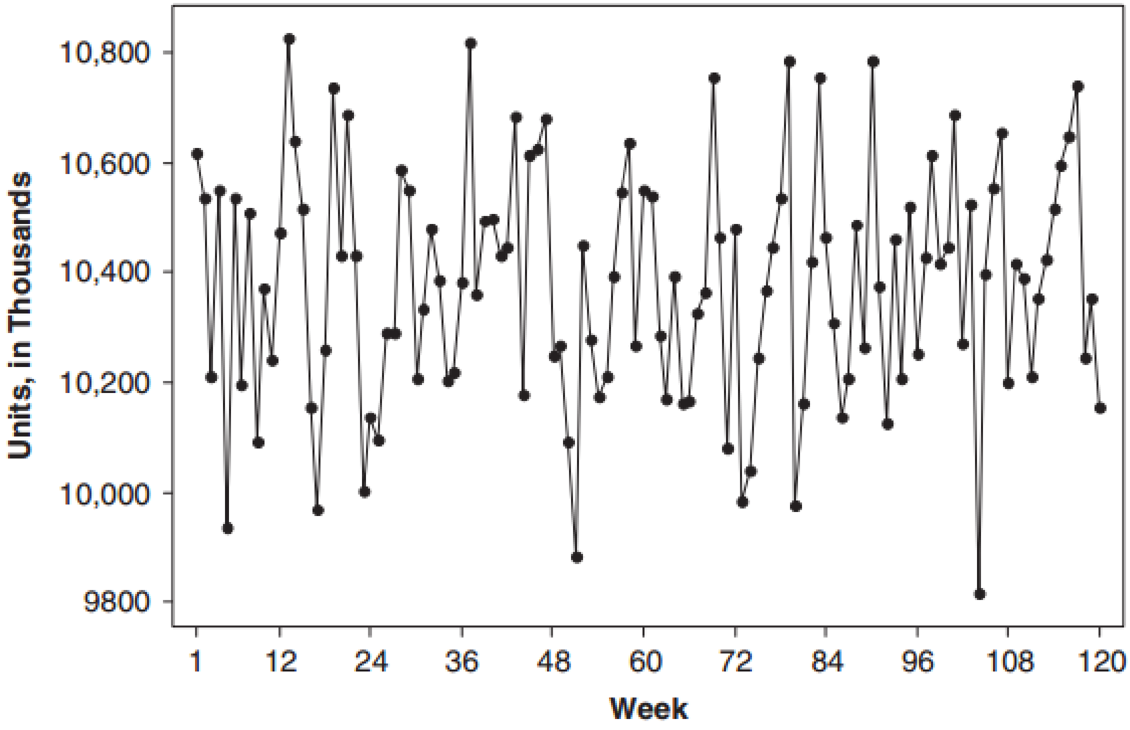
For some series it is possible to take measurements at every moment of time, so that a trace results. Such data are said to form a *continuous* time series. A continuous series can always be approximated by a discrete one by selecting smaller and smaller sampling interval.

Sometimes the data may be instantons, such as the viscosity of a chemical product at the time when it is measured.

Examples of Time Series

The sales of a mature pharmaceutical product may remain relatively flat in the situation of unchanged marketing or manufacturing strategies.

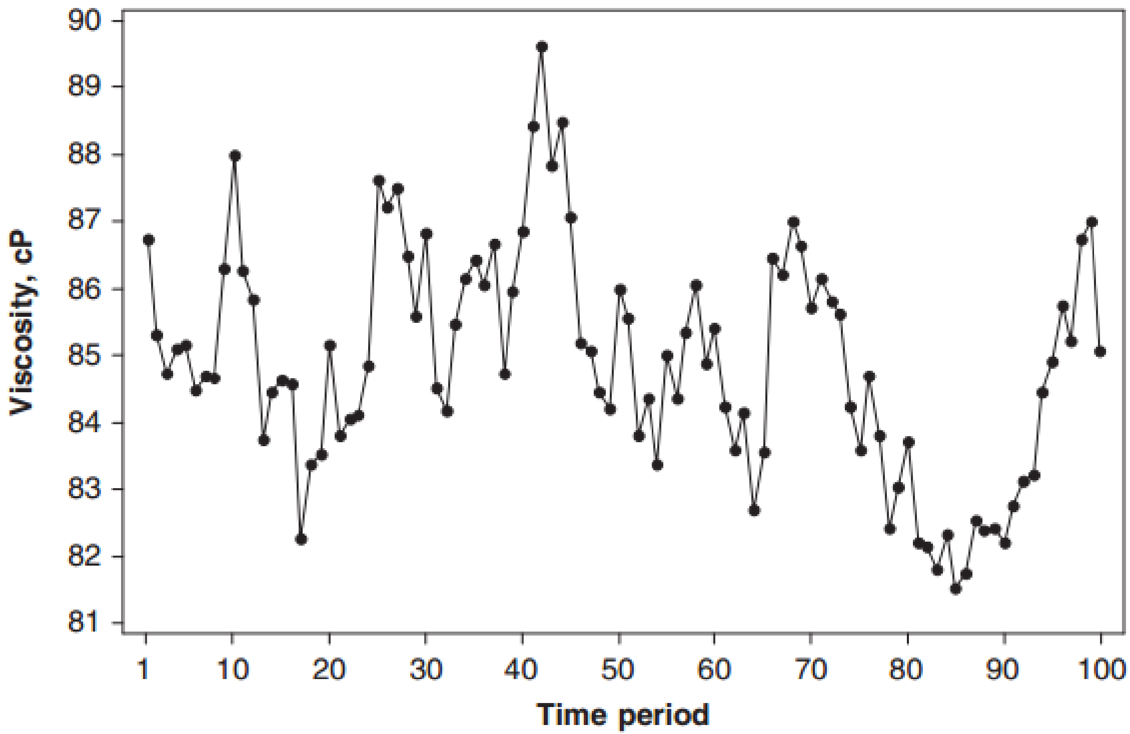
Weekly sales of a generic pharmaceutical product are shown below.



Examples of Time Series

In chemical industry, due to the continuous nature of chemical manufacturing processes, output properties often are positively autocorrelated.

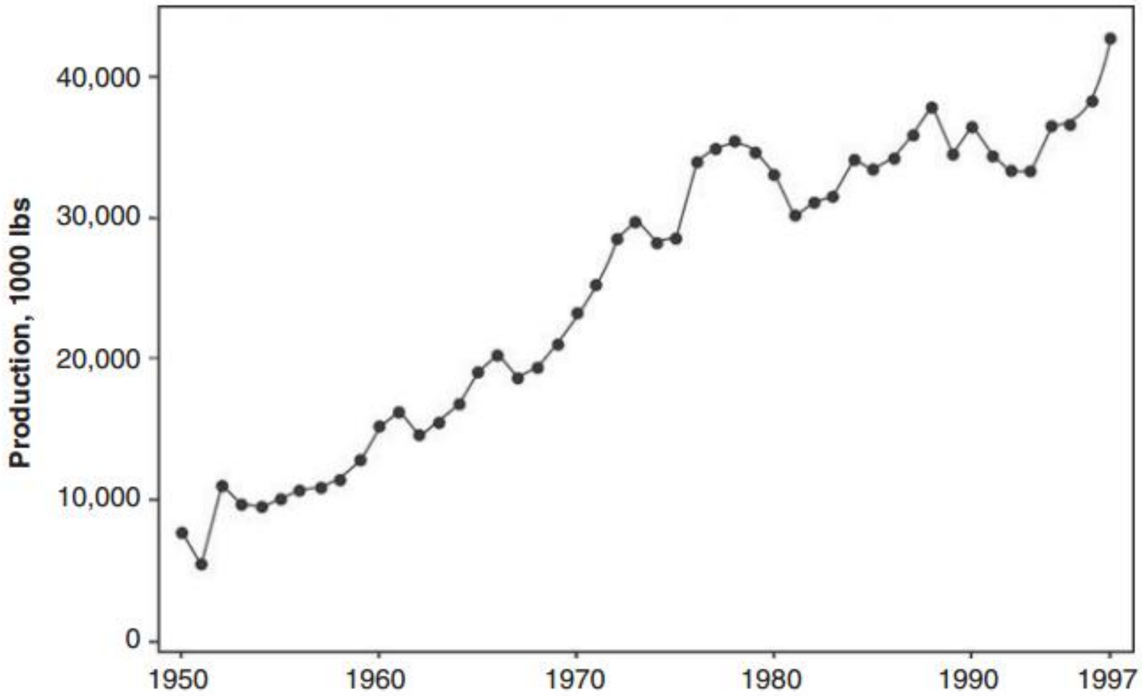
The viscosity readings exhibit autocorrelated behavior.



Examples of Time Series

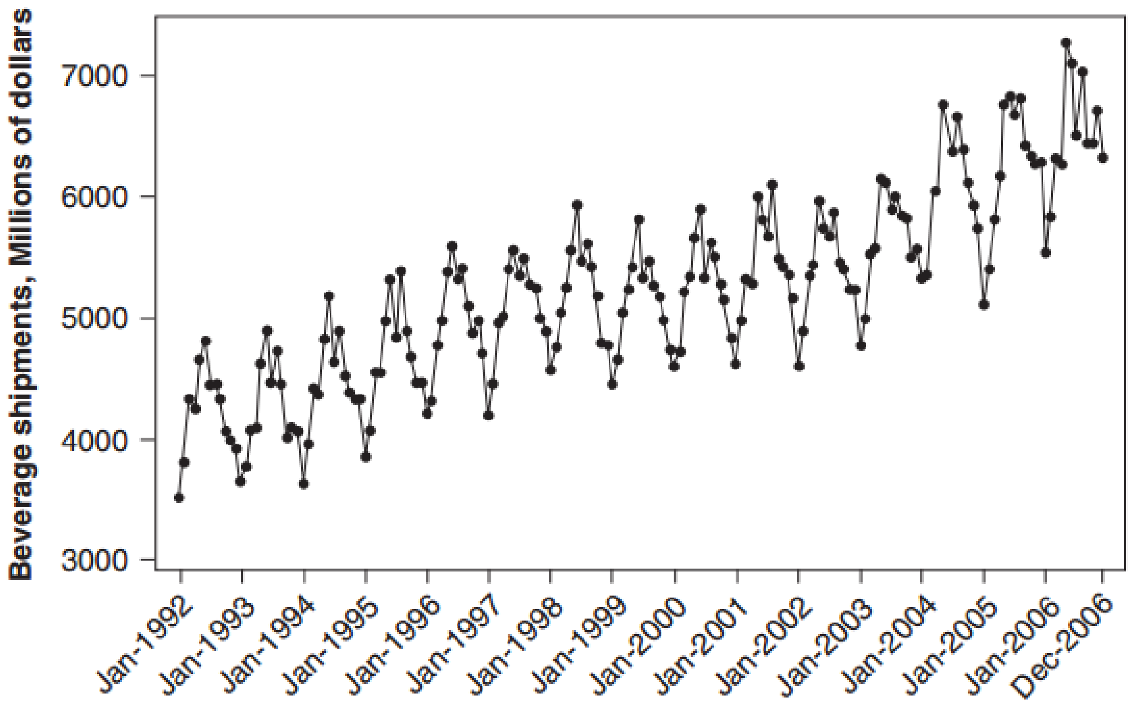
The annual production of blue cheese is published by the USDA National Agricultural Statistics Service.

The data are from 1950 to 1997, and the linear trend has a constant positive slope with random, year-to-year variation.



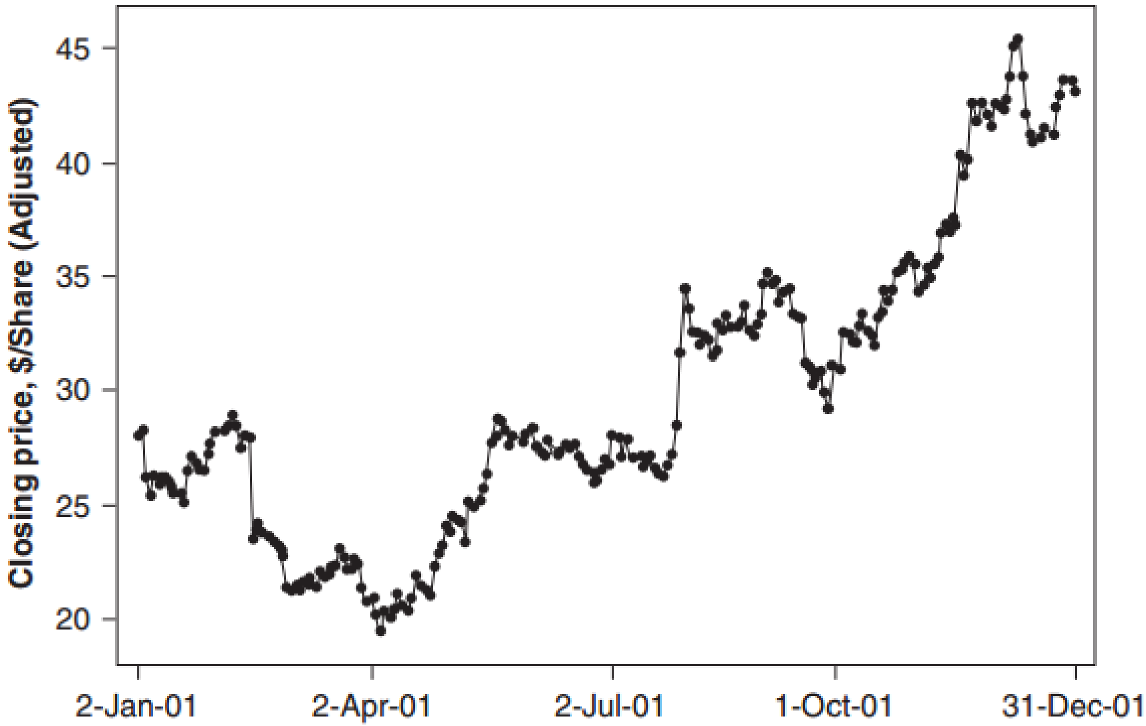
Examples of Time Series

The manufacture of beverage products is reported by the US Census Bureau. The plot of monthly beverage product shipments reveals an overall increasing trend, with a distinct seasonal pattern that is repeated within each year.



Examples of Time Series

Business data such as stock prices and interest rates often exhibit nonstationary behavior, such as the following data which record the daily closing price adjusted for stock splits of Whole Foods Market (WFMI) stock in 2001.



Basic Statistic

1. Moments
2. Covariance

   Now we are studying a time series $\{X_t\}$.
   The **mean function** of $\{X_t\}$ is $\mu_X(t) = E(X_t)$.

   The **covariance function** of $\{X_t\}$ is
   $\gamma_X(r, s) = Cov(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$
   It is named as **autocovariance function**.

   The **autocorrelation function (ACF)** is given by $\rho_X(r, x) = \frac{\gamma_X(r,s)}{\sqrt{\gamma_X(r,r)\gamma_X(s,s)}}$.
3. Least square estimation
4. Maximum likelihood estimation

Stationary

A very important type of time series is a stationary time series.

$\{X_t\}$ is (**weakly**) **stationary** if

(i) $\mu_X(t)$ is independent of t,

and

(ii) $\gamma_X(t+h, t)$ is independent of t for each h.

The second condition means that the autocovariance $\gamma_X(r, s)$ has nothing to do with either r or s, but is determined by the time interval $|r - s|$.

A time process $\{X_t\}$ is called a **white noise process** if a sequence of uncorrelated random variables, each with a constant mean $E(X_t) = \mu_X$, usually assumed to be 0, and constant variance $Var(X_t) = \sigma_X^2$. It is the simplest but important stationary time series.

Time Series Model

We can use x_t, $t = 1, 2, .., n$ to denote the observed discrete time series.

A **time series model** for the observed data $\{x_t\}$ is a specification of the joint distributions (or possibly only the means and covariances) of a sequence of random variables $\{X_t\}$, of which $\{x_t\}$ is postulated to be a realization.
We call the stochastic process $\{X_t\}$ as a time series process.

For example, if we toss a coin every week, then we can define a series of random variables $\{X_t\}$, let $X_t = 1$ if head in shown in week t, or 0 otherwise.

Suppose in the 1st week, head is observed, then the observation of X_1 is $x_1 = 1$. In the 2nd week, tail is observed, then the observation of X_2 is $x_2 = 0$. After 5 weeks, 5 observations are generated for n random variables.

Then $\{X_t\}$ denotes a time series model which models a binary process. And the 5 observations are one possible realizations of such process.

One major task is to build a time series model which can best explain the time series data.

Autoregressive Process

The **first-order autoregressive process** $\{X_t\}$, denoted by AR(1), is given by $X_t = c + \phi_1 X_{t-1} + \varepsilon_t$, where c and ϕ_1 are constants and $\{\varepsilon_t\}$ is a white noise process. $E(\varepsilon_t) = 0$ $Var(\varepsilon_t) = \sigma^2$

Let B denote the backward operator. It means $BX_t = X_{t-1}$.

Then AR(1) can be written as $(1 - \phi_1 B)X_t = c + \varepsilon_t$.

When the root of $1 - \phi_1 z = 0$ is outside of the unit circle, which means $|\phi_1| < 1$, the first-order autoregressive process $\{X_t\}$ is stationary.

For example, $X_t = 0.9X_{t-1} + \varepsilon_t$ is stationary, but $X_t = -1.1X_{t-1} + \varepsilon_t$ is not.

Autoregressive Process

For the stationary AR(1),

The autocovariance function is $\gamma_k = \phi_1 \gamma_{k-1}, k \geq 1$.

The variance function is $\gamma_0 = \frac{\sigma^2}{1-\phi_1^2}$.

The mean is $\mu = \frac{c}{1-\phi_1}$.

The ACF is $\rho_k = \phi_1 \rho_{k-1} = \phi_1^k$ for $k \geq 1$.

The PACF is $P_k = \begin{cases} \rho_1 & k = 1 \\ 0 & k \geq 2 \end{cases}$

Partial Autocorrelation Function

In addition to the autocorrelation between X_t and X_{t+h} , we also investigate the correlation between them after their mutual linear dependency on the intervening variables X_{t+1}, $X_{t+2} \ldots$ and X_{t+h-1} has been removed. In general, partial correlation is a conditional correlation which is

$$P_h = Corr(X_t, X_{t+h} | X_{t+1}, X_{t+2}, ..., X_{t+h-1})$$

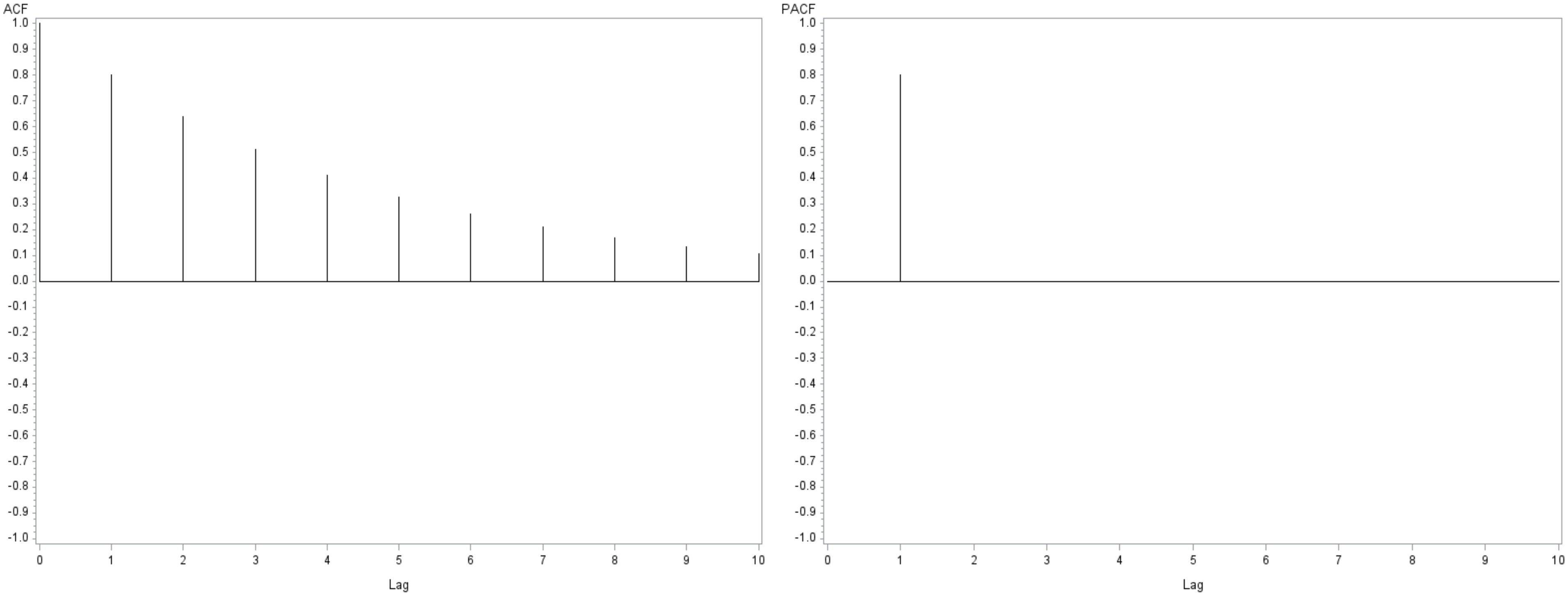
The calculation of the partial autocorrelation function (PACF) is complex.

$$P_1 = \rho_1 \qquad P_2 = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} \qquad P_k = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

Together with ACF, PACF also helps us to identify model.

Autoregressive Process

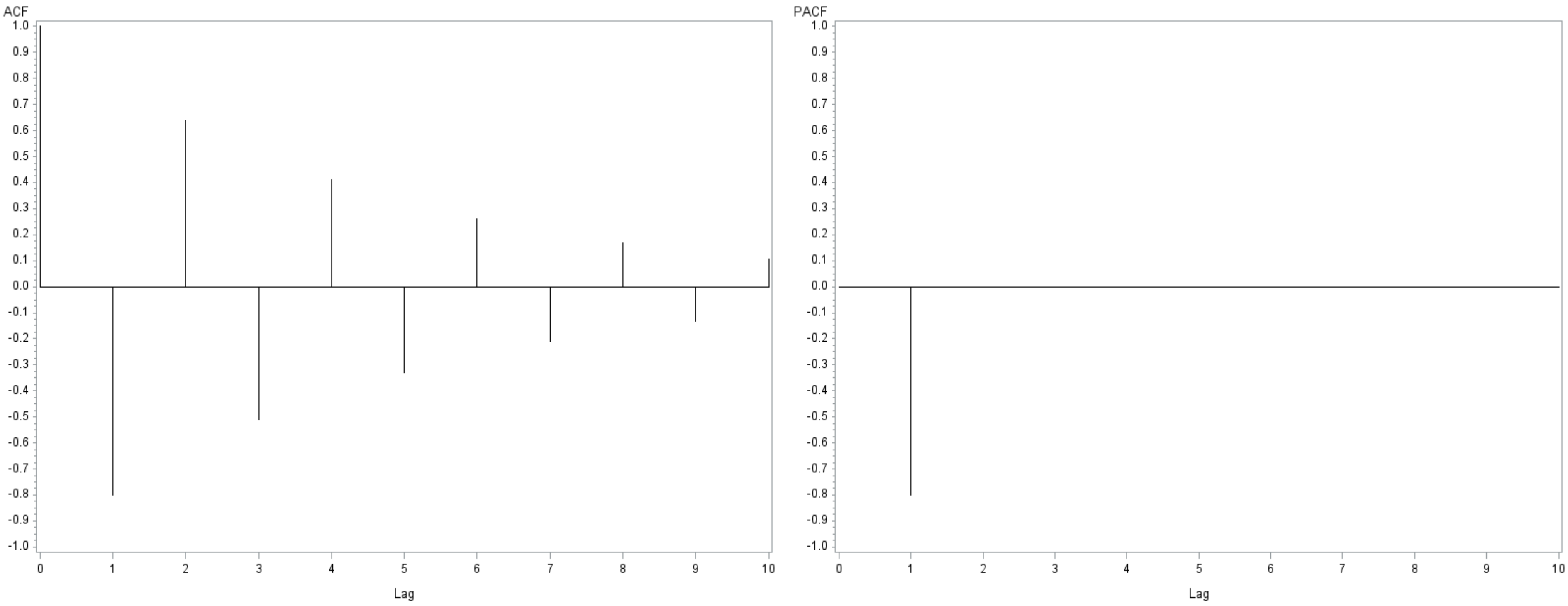
We can get the theoretical ACF and PACF of the AR(1) process with $\phi_1 = 0.8$.



The ACF decays exponentially, while the PACF cuts off after lag 1. So if the sample ACF and sample PACF (especially the later one) derived from one data is similar to these graphs, then we can try AR(1) model on the data.

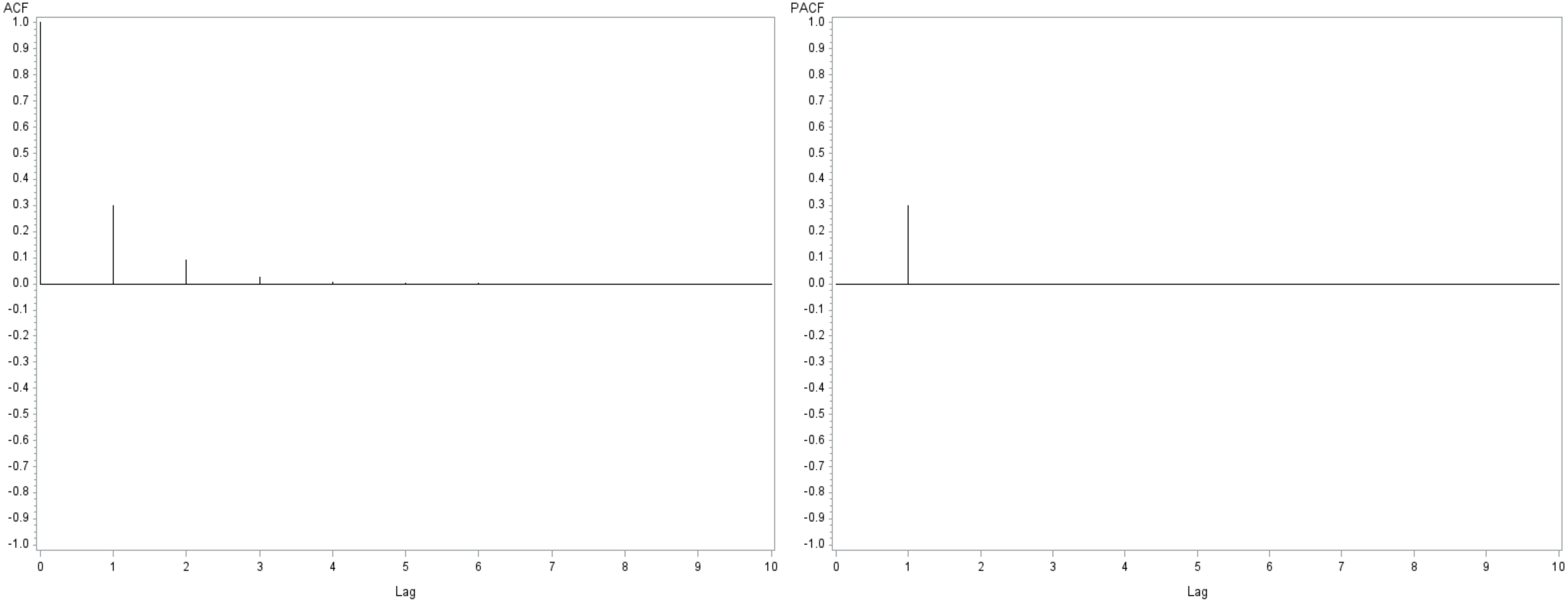
Autoregressive Process

If $\phi_1 = -0.8$, the theoretical ACF and PACF of AR(1) is as follows.



Autoregressive Process

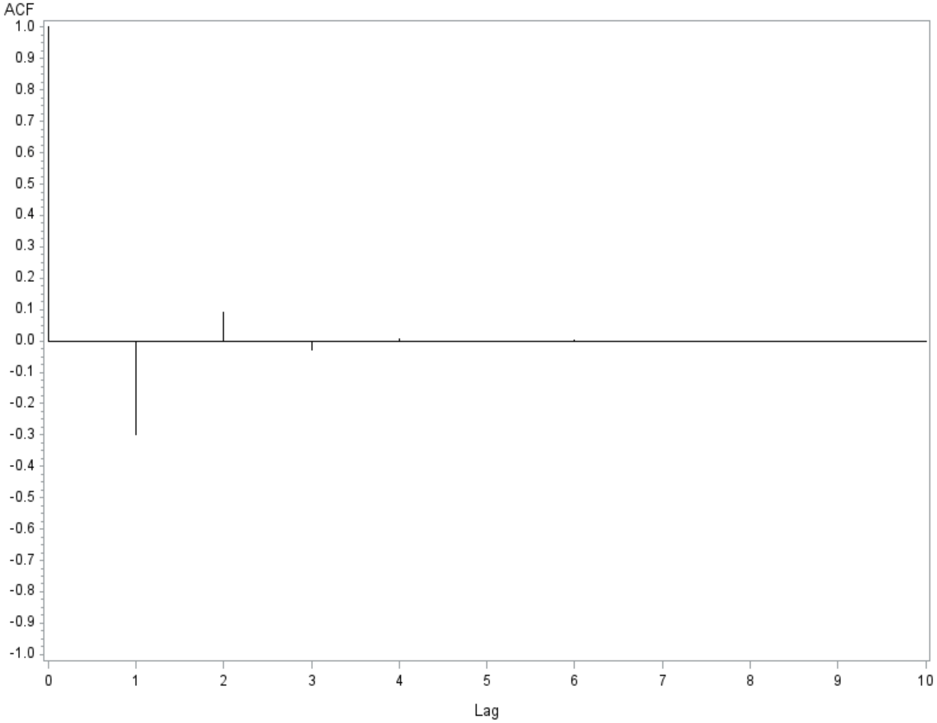
If $\phi_1 = 0.3$, the theoretical ACF and PACF of AR(1) is as follows.

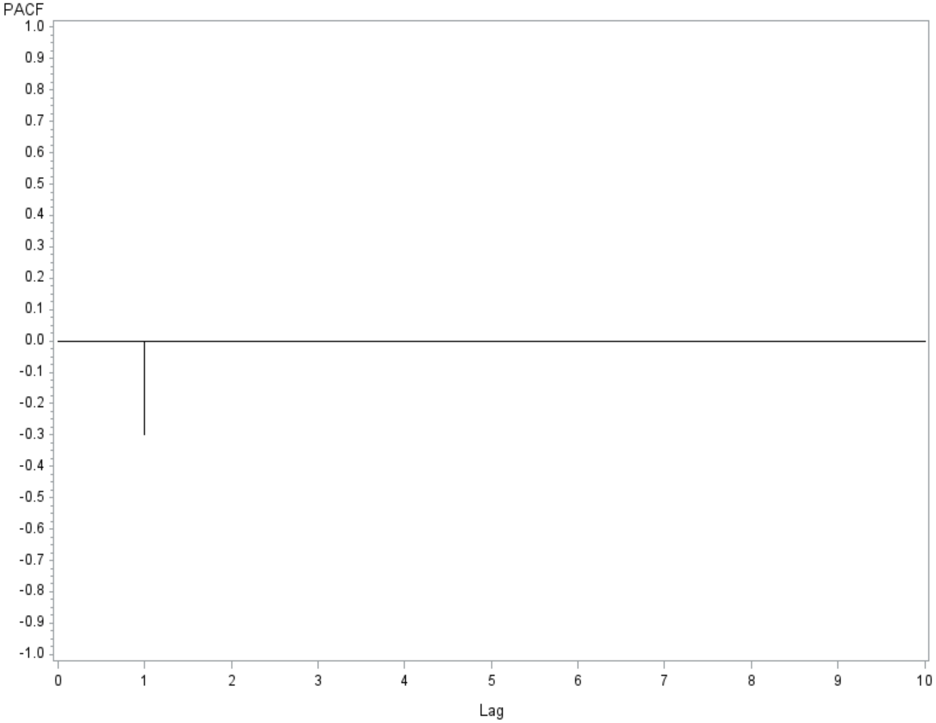


When the absolute value of ϕ_1 is small, then the ACF decays quickly.

Autoregressive Process

If $\phi_1 = -0.3$, the theoretical ACF and PACF of AR(1) is as follows.





Autoregressive Process

The **second-order autoregressive AR(2)** process is defined by
$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$ or $(1 - \phi_1 B - \phi_2 B^2) X_t = c + \varepsilon_t$.

To be stationary, the roots of $1 - \phi_1 z - \phi_2 z^2 = 0$ must lie outside of the unit circle.

For example, is $X_t = c - 1.5X_{t-1} + X_{t-2} + \varepsilon_t$ stationary?

No. The roots of $1 + 1.5z - z^2 = 0$ are $z_1 = 2$ $z_2 = -0.5$. The second root is not outside of the unit circle.

For example, is $X_t = c + 1.5X_{t-1} - 0.56X_{t-2} + \varepsilon_t$ stationary?

The roots of $1 - 1.5z + 0.56z^2 = 0$ are $z_1 = 1/0.7$, $z_2 = 1/0.8$. Both of the roots are outside of the unit circle. So it is stationary.

Autoregressive Process

For the stationary AR(2) process, we have $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}, k \geq 1$. With a little algebra,

$$\rho_1 = \frac{\phi_1}{1-\phi_2}, \rho_2 = \frac{\phi_1^2+\phi_2-\phi_2^2}{1-\phi_2}$$

The PACF is given by

$$P_1 = \rho_1 = \frac{\phi_1}{1-\phi_2}, P_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2} = \phi_2, \text{ and } P_k = 0, \text{ for } k \geq 3$$

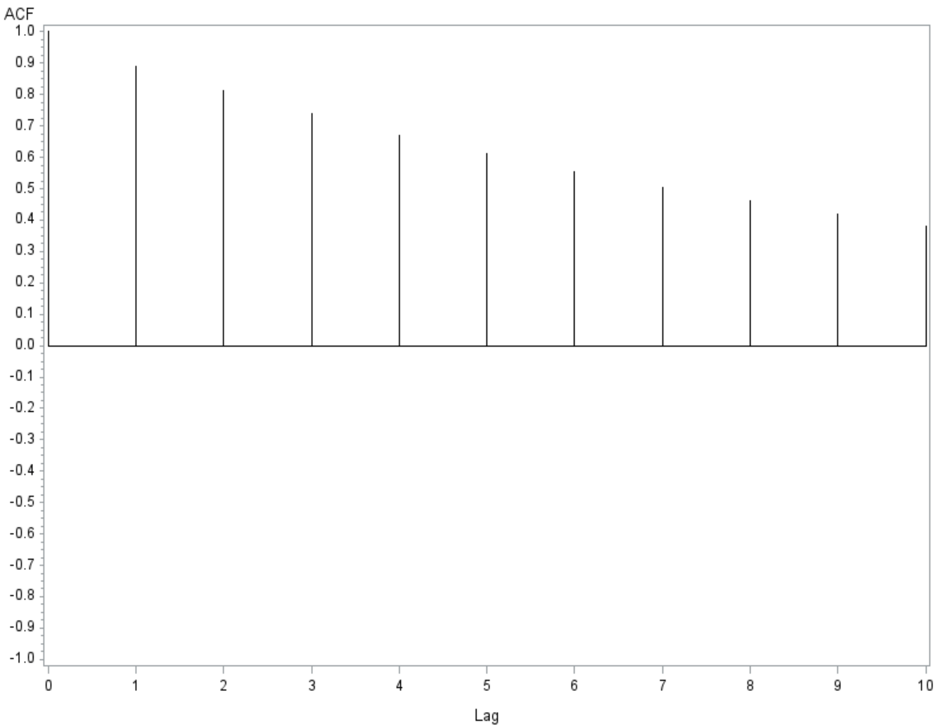
The autocovariance of zero order, which is also the variance, of X_t is

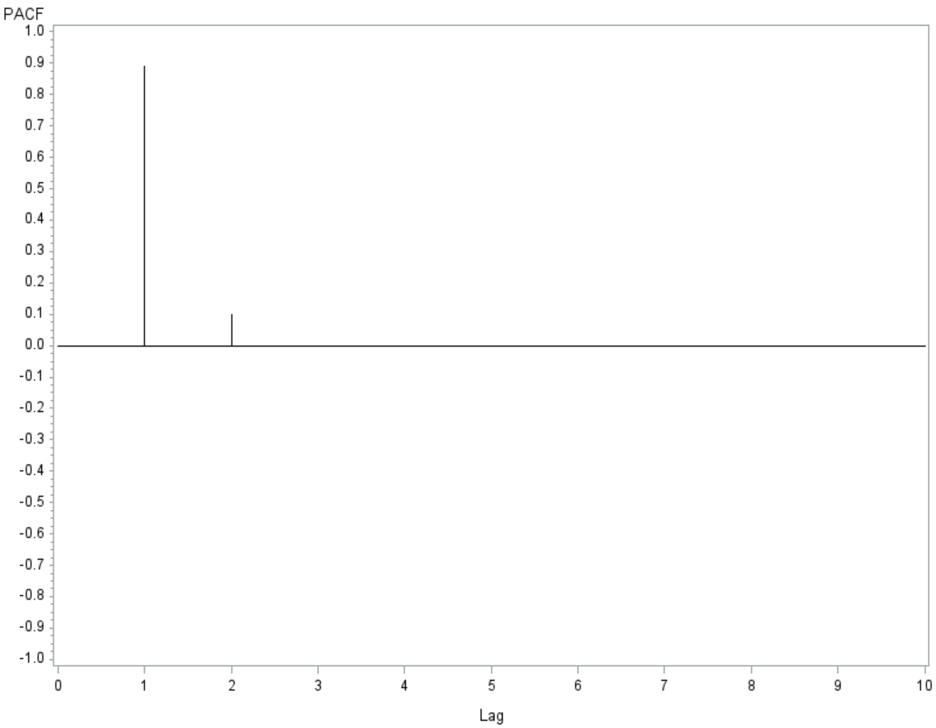
$$\gamma_0 = \frac{(1-\phi_2)\sigma^2}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]}$$

The mean of AR(2) process is $\mu = \frac{c}{1-\phi_1-\phi_2}$

Autoregressive Process

The theoretical ACF and PACF of AR(2) process with $\phi_1 = 0.8$ and $\phi_2 = 0.1$.

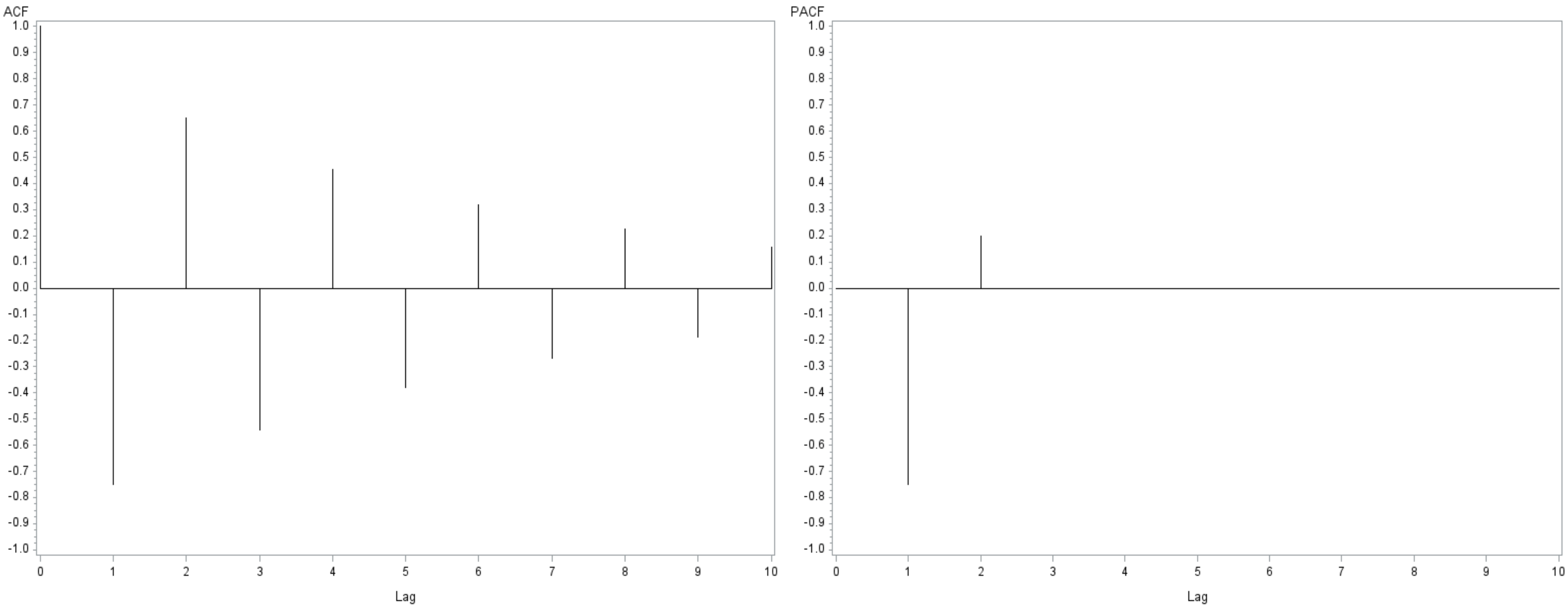




We can see that the ACF decays exponentially but slower than AR(1). The PACF cuts off at lag 2.

Autoregressive Process

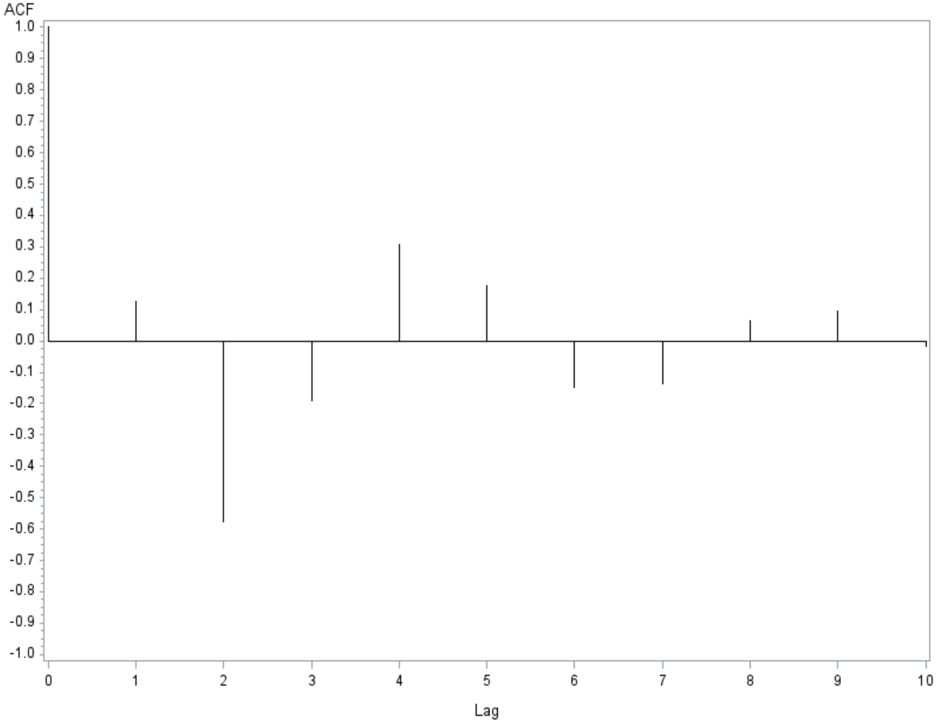
The theoretical ACF and PACF of AR(2) process with $\phi_1 = -0.6$ and $\phi_2 = 0.2$.

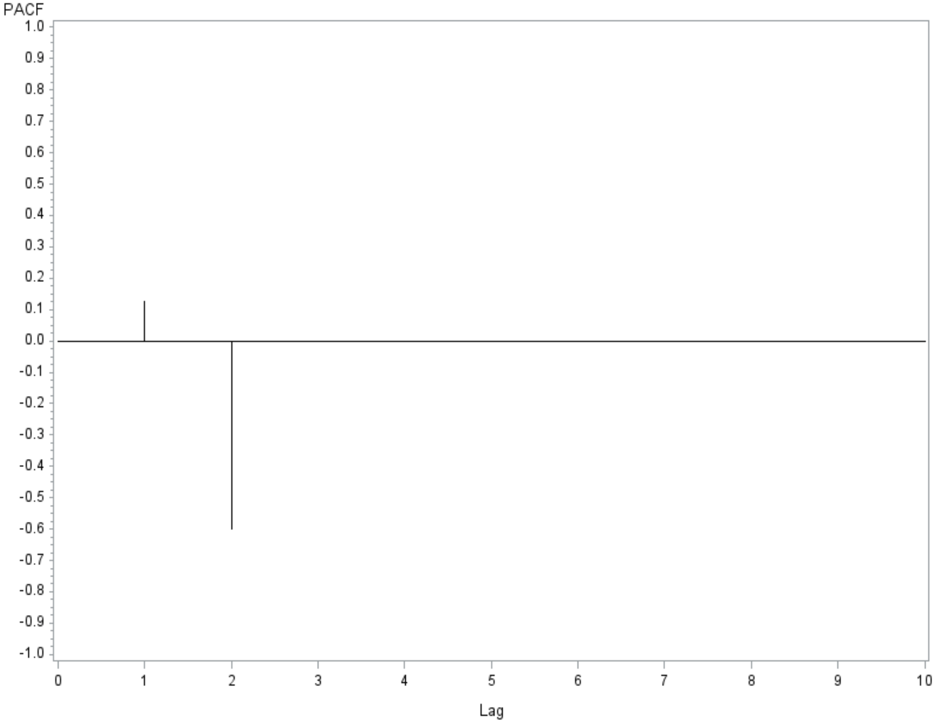


We can see that the absolute value of ACF decays exponentially. The PACF cuts off at lag 2.

Autoregressive Process

The theoretical ACF and PACF of AR(2) process with $\phi_1 = 0.2$ and $\phi_2 = -0.6$.





We can see that after excluding the relationship between X_t and X_{t-1}, the correlation between X_t and X_{t-2} is just the parameter $\phi_2 = -0.6$. This property can be used to estimate the parameters.

Autoregressive

The general **pth-order autoregressive AR(P)** is

$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t$, or

$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) X_t = c + \varepsilon_t$

To make it stationary, the roots of $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$ should be outside of the unit circle.

The autocovariance function $\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p}, k > 0$.

$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \cdots + \phi_p \gamma_p + \sigma^2$

The autocorrelation function $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}, k > 0$.

About PACF, remember that it will vanish after lag p. That is, $P_k = 0$, when $k > p$. We can use this property to identify the AR process and its order.

Moving Average

The general **first-order moving average MA(1)** is defined as $X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$. The term 'moving average' comes from the fact that X_t is constructed from a weighted sum, akin to an average, of the two most recent values of noises.

Finite moving average process are all stationary.

The variance of X_t is $\gamma_0 = (1 + \theta_1^2)\sigma^2$.

The ACF of the MA(1) is $\rho_1 = \frac{\theta_1}{1+\theta_1^2}$, $\rho_k = 0$, when $k \geq 2$.

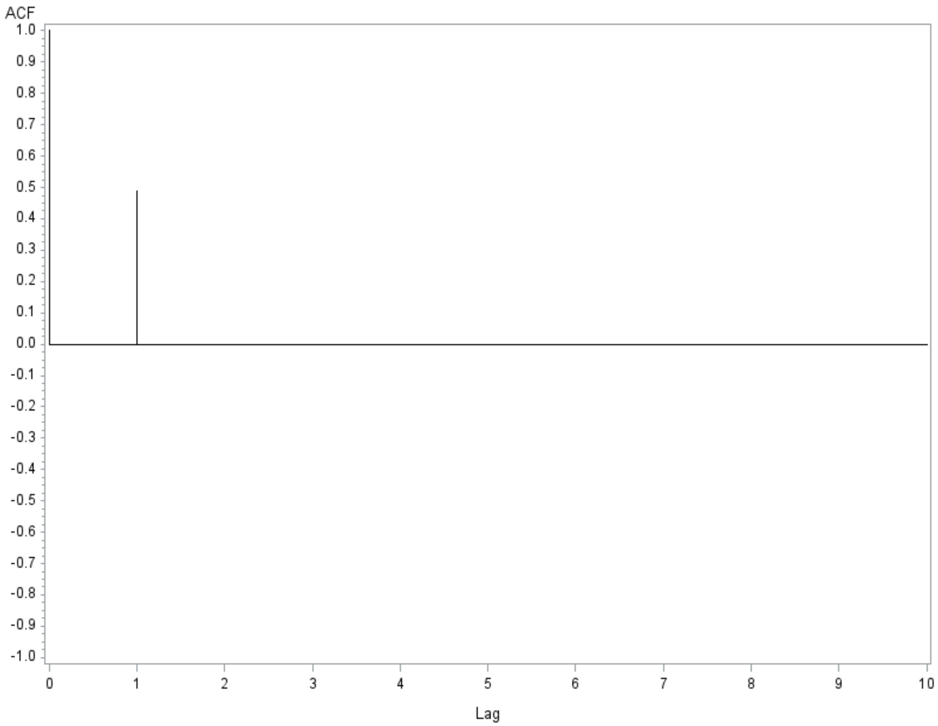
The PACF is calculated as

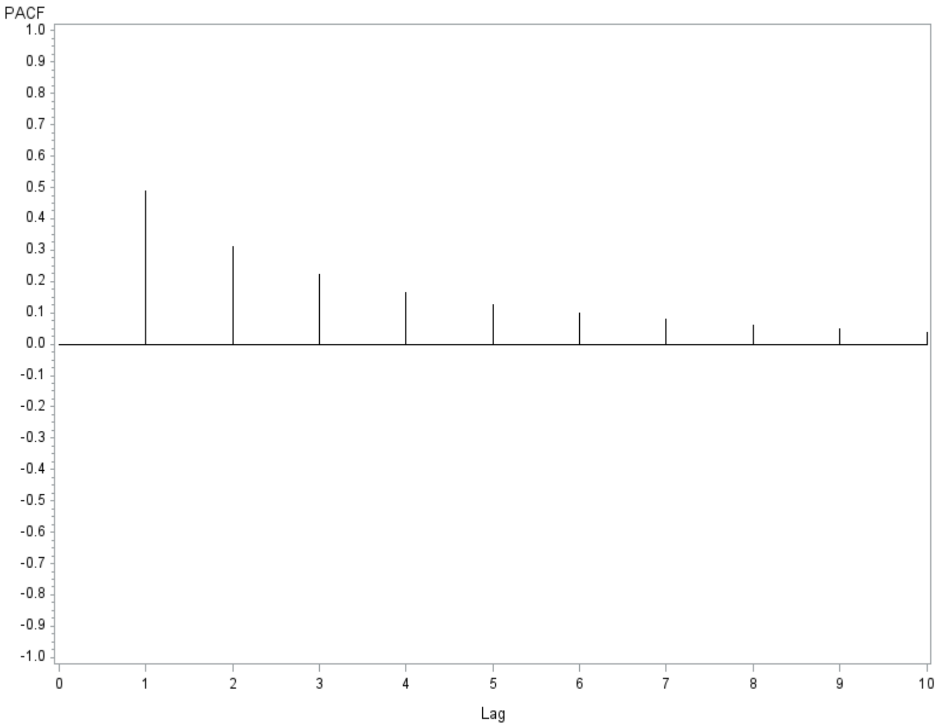
$$P_k = \frac{\theta_1^k(1-\theta_1^2)}{1-\theta_1^{2(k+1)}}$$

We can see that, different from AR(1), the ACF of MA(1) cuts off after lag 1, while the PACF tails off.

Moving Average

We can get the theoretical ACF and PACF of the MA(1) process with $\theta_1 = 0.8$.

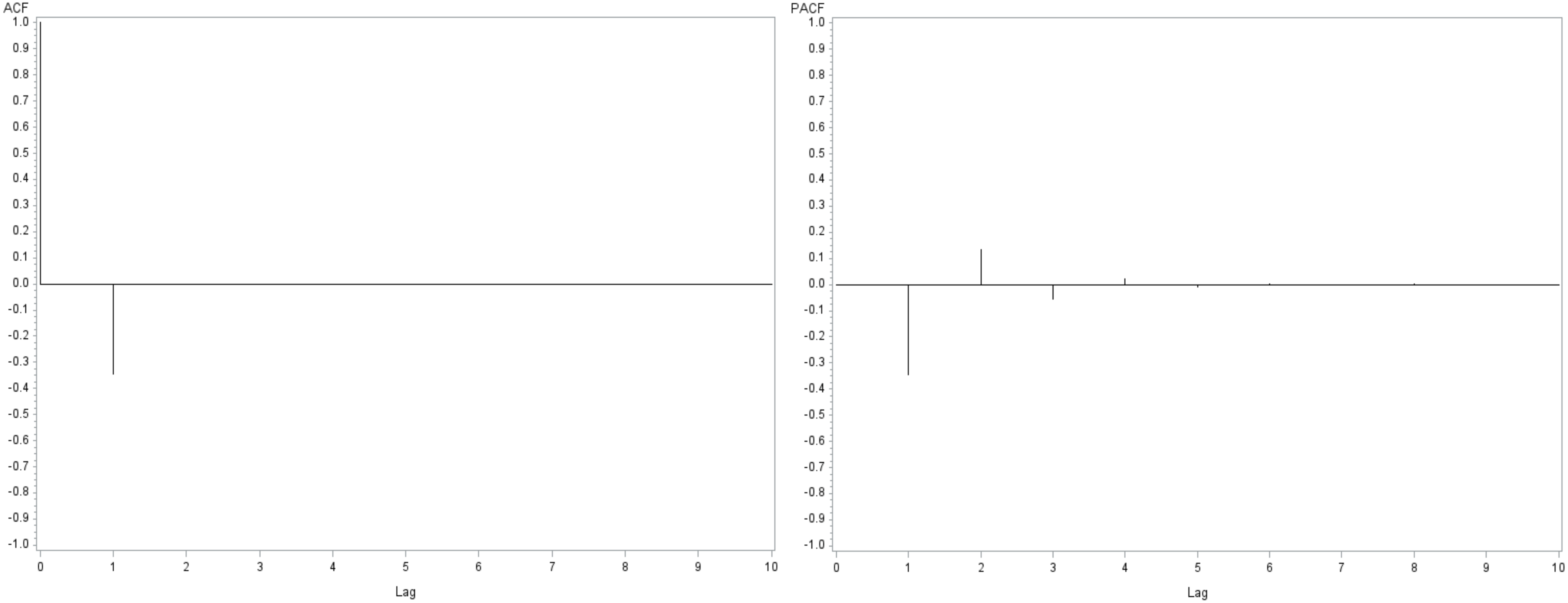




Conversely, MA(1) cuts off after lag 1.

Moving Average

We can get the theoretical ACF and PACF of the MA(1) process with $\theta_1 = -0.4$



Conversely, MA(1) cuts off after lag 1.

Moving Average

The **general qth-order moving average process** is defined as
$X_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$. It is stationary.

The variance of X_t is $\gamma_0 = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma^2$.

The autocovariance function of X_t is

$\gamma_k = (\theta_k + \theta_{k+1}\theta_1 + \theta_{k+2}\theta_2 + \cdots + \theta_q\theta_{q-k})\sigma^2$, for $k = 1, 2, ..., q$

It cuts off after lag q.

The ACF is calculated by $\rho_k = \gamma_k/\gamma_0$

When the order q approaches infinite, we have the infinite-order moving average process.

$$X_t = \mu + \sum_{j=0}^{\infty} \theta_j\varepsilon_{t-j} = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots$$

MA(∞) is stationary provided that $\sum_{j=0}^{\infty} \theta_j^2 < \infty$.

Then it is also stationary when $\sum_{j=0}^{\infty} |\theta_j| < \infty$.

ARMA(p,q) process

The autoregressive moving average (ARMA) process includes the autoregressive and moving average process as special cases.

The process $\{X_t\}$ is said to be an ARMA(p,q) process if $\{X_t\}$ is stationary and if for every t,

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

It can be written in the more compact form

$\phi(B)X_t = \theta(B)\varepsilon_t$, where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, and $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$.

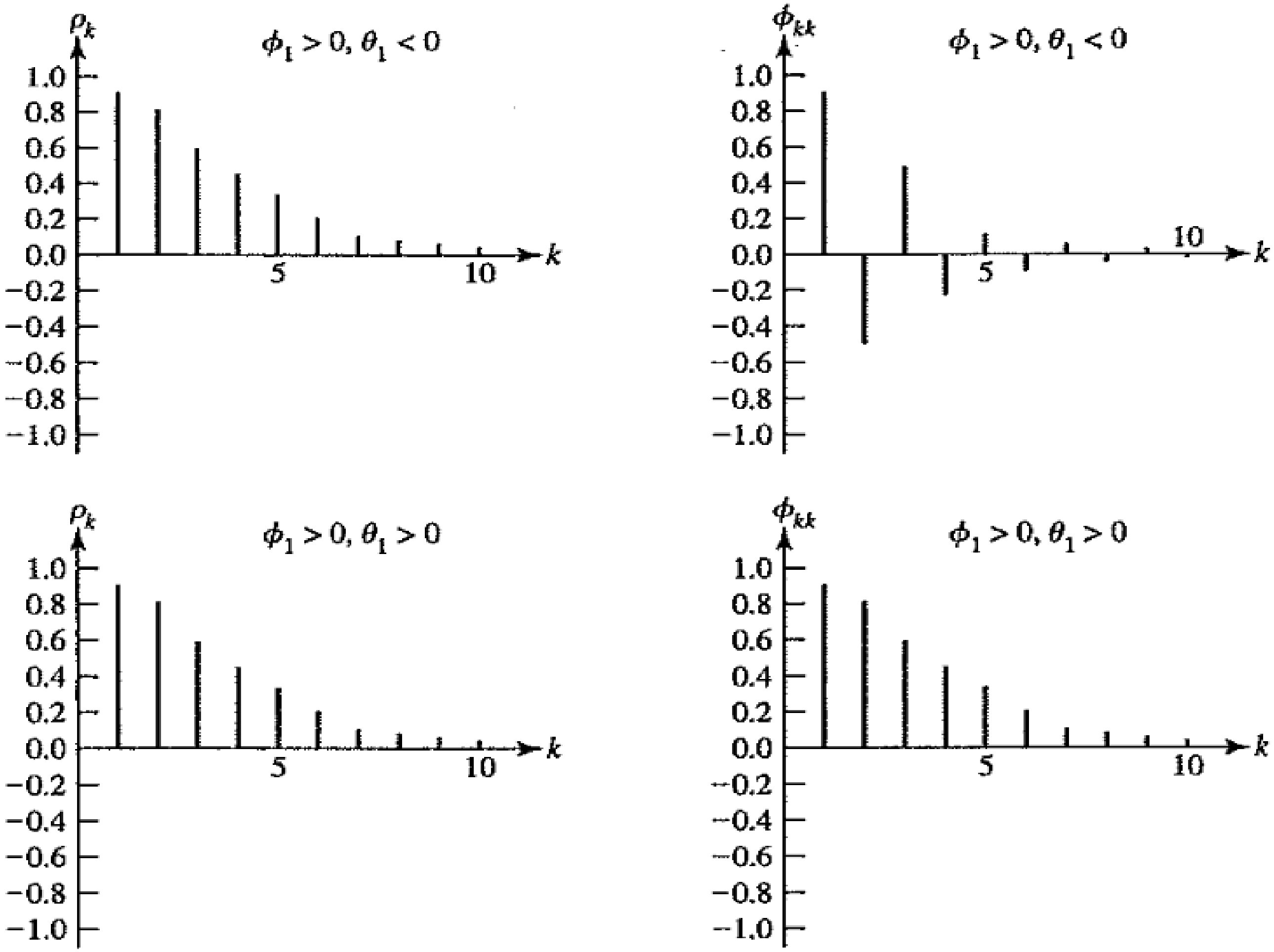
If $\phi(B) = 1$, then it becomes a MA(q) process. If $\theta(B) = 1$, then it becomes an AR process. p and q are used to indicate the orders of the associated autoregressive and moving average polynomials, respectively.

The calculation of the ACF will be the same with the previous ones. But it is complex, so we don't illustrate here.

Neither the ACF nor the PACF cuts off for ARMA process.

ARMA(p,q) process

Theoretical ACF and PACF of ARMA(1,1).



Summary

The AR(p) model is $X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t$, so the current value X_t can be found from past values plus a random effect. Like a multiple regression model with the independent variables are the lag values of the dependent variable.

The MA(q) model is $X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$, so the current value X_t is regarded as the moving average of the random residual series.

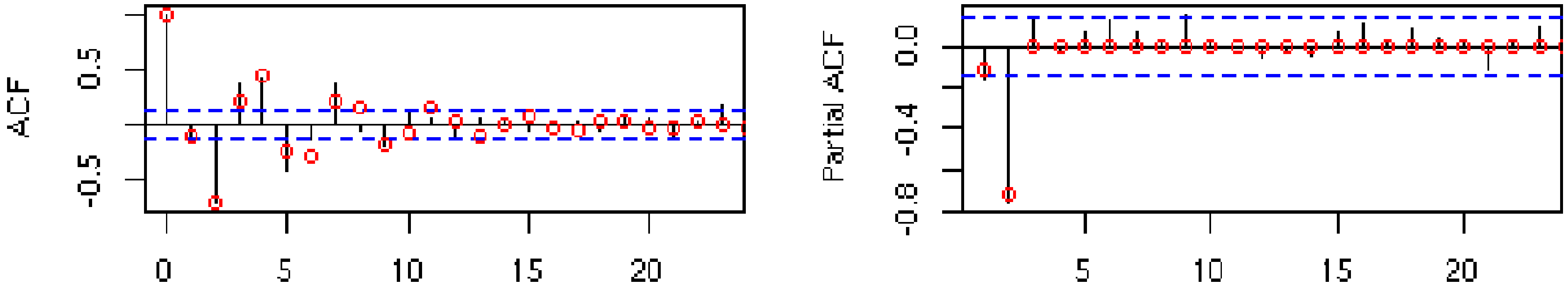
ARMA(p, q) process mixes the AR(p) and MA(q) processes, which is defined as $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$, or briefly as $\phi(B) X_t = \theta(B) \varepsilon_t$.

ARMA(p, q) can only be used to deal with the stationary time series.

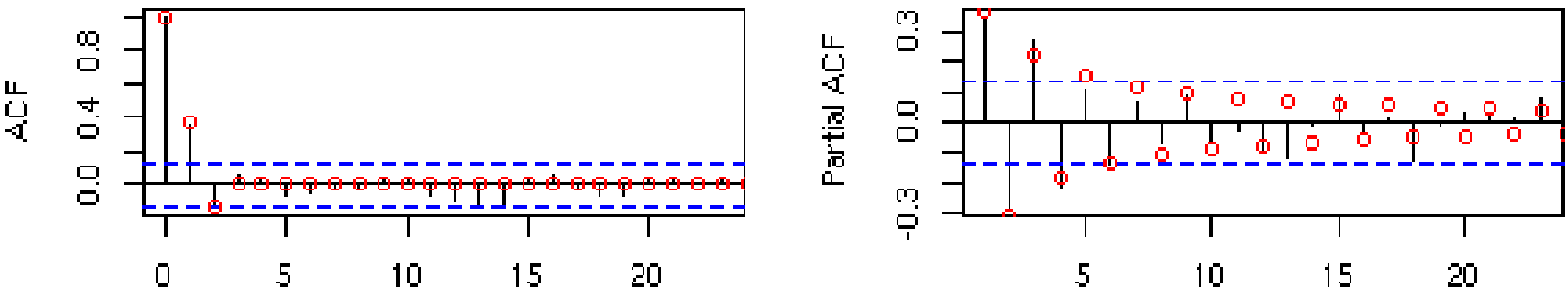
Summary

Combined ACF and PACF Pattern		Possible Model
Sample ACF	Sample PACF	
Tapering or sinusoidal pattern that converges to 0, possibly alternating negative and positive signs	Significant values at the first p lags, then non-significant value	AR of order p
Significant values at the first q lags, then non-significant values	Tapering or sinusoidal pattern that converges to 0, possibly alternating negative and positive signs	MA of order q
Tapering or sinusoidal pattern that converges to 0, possibly alternating negative and positive signs	Tapering or sinusoidal pattern that converges to 0, possibly alternating negative and positive signs	ARMA with both AR and MA terms – identifying the order involves some guess work

Your Turn

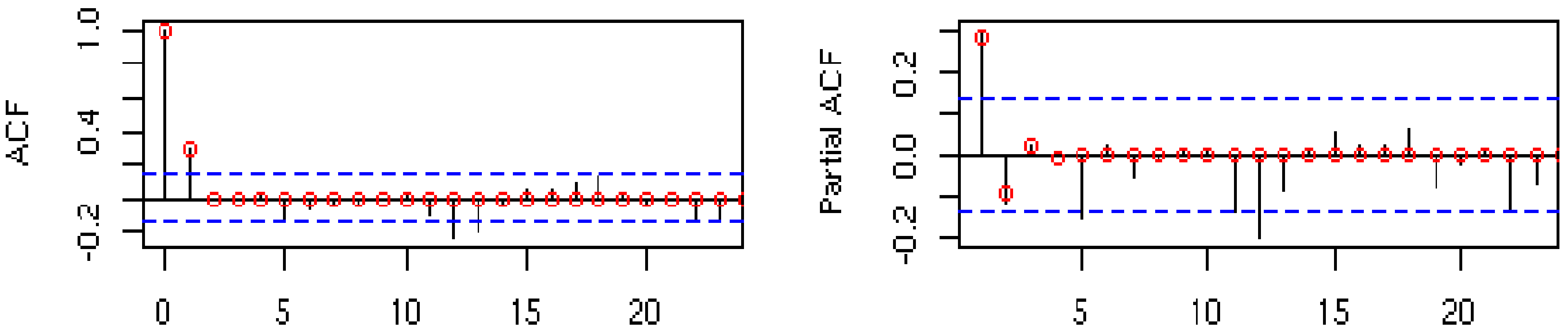


AR(2) shall be used.

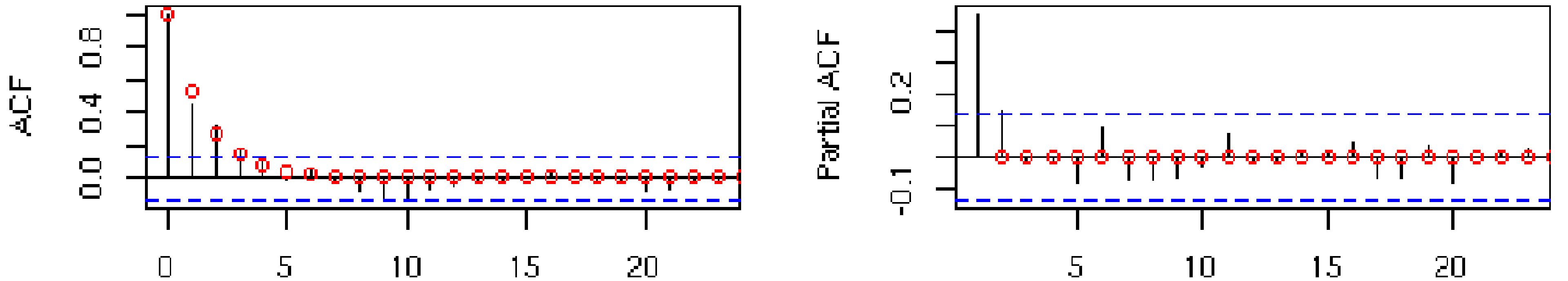


MA(2) shall be used.

Your Turn



MA(1) shall be used.



AR(1) shall be used. Sometimes, the ACF and the PACF may indicate that several possible models can be used. Then we need more tools to select the orders.

Difference to deal with the Trend Component

Differencing the time series is always an effective method to eliminate the trend component.

The first difference is determined by

$$\triangle X_t = X_t - X_{t-1} = (1-B)X_t$$

Accordingly, the second difference is defined by

$$\triangle^2 X_t = \triangle X_t - \triangle X_{t-1} = (1-B)^2 X_t$$

Usually, first differencing, and occasionally second differencing would be enough to achieve stationary.

If after d differencing, the time series becomes an ARMA(p,q) process, then we say the original time series is an ARIMA(p,d,q) process. It is written as

$$X_t = (1-B)^d Y_t \text{ , } X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

Or, write the formula in a brief way

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)(1-B)^d Y_t = \boxed{c} + (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q)\varepsilon_t$$

Usually c=0

Difference

The level of differencing is estimated by considering the autocorrelation plots. When the autocorrelations die out quickly, the appropriate value of *d* has been found.

The value of *p* is determined from the partial autocorrelations of the appropriately differenced series. If the partial autocorrelations cut off after a few lags, the last lag with a large value would be the estimated value of *p*.

The value of *q* is found from the autocorrelations of the appropriately differenced series. If the autocorrelations cut off after a few lags, the last lag with a large value would be the estimated value of *q*.

And even we choose AR(p), it does not mean that the formula should contain all the orders of AR(p). For example, if from the sample PACF, AR(2) is suggested, the resulting model could be $X_t = c + \phi_2 X_{t-2} + \varepsilon_t$ if the t-test of the ϕ_1 suggest that it is not significant.

Seasonal Time Series

Many business and economic time series contain a seasonal phenomenon that repeats itself after a regular period of time. The smallest time period for this repetitive phenomenon is called the seasonal period.

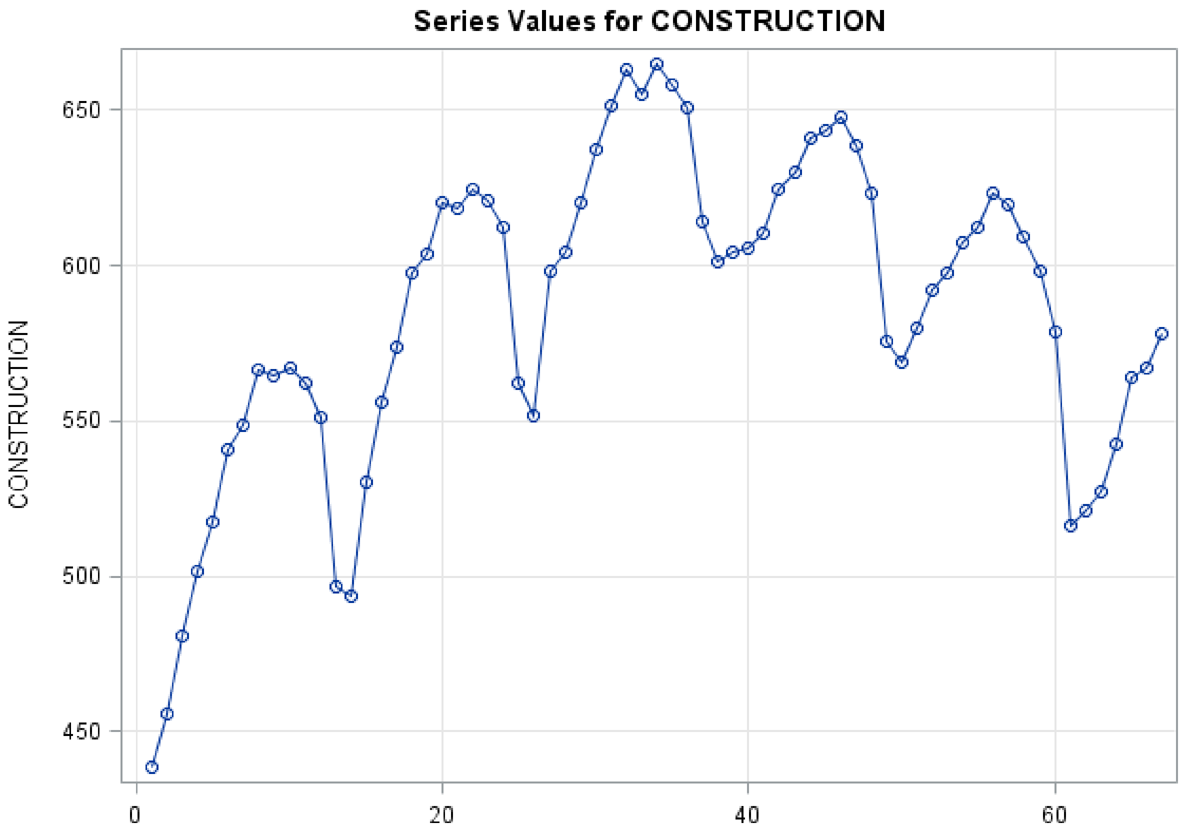
For example, the quarterly series of ice-cream sales is high each summer, and the series repeats this phenomenon each year, giving a seasonal period of 4.

For example, the monthly auto sales tend to decrease during August and September every year because of the changeover to new models. The seasonal period in this case is 12.

The number of calls every half hour in a credit card call center is increasing from 9 am every day, reaching the top around 11 am and then decreasing to the lowest. The seasonal period in this case is 48.

Example of Seasonal Time Series

Construction.xls records the the monthly number of U.S. masonry and electrical construction workers in thousands. It is very clear that it is not stationary. The trend is first increasing and then decreasing. Totally 67 observations.



Summary

It is provided by Box and Jenkins (1972) that any nonstationary time series can be stationary after several differencing. (Of course, it is not true.)

But differencing can be an effective method to eliminate the trend and seasonal component.

If there is no seasonal component, we just need to make d differencing, which is $(1-B)^d Y_t = X_t$. We expect X_t to be stationary.

If there exists seasonal component, then we can consider to make $(1-B^s)^D Y_t = X_t$. We expect X_t to be stationary.

One differencing will exhaust one observation. One seasonal differencing will exhaust s observations, where s is the length of the seasonality.

Summary

Sometimes, there are both trend and seasonal components, we can transform the time series by $(1-B)^d(1-B^s)^D Y_t = X_t$.

The resulting X_t may be seasonal stationary, which means that the series $\{X_{i+ks}\}$, $k = 0, 1, 2, \ldots$, are stationary, but the overall $\{X_t\}$ is not.

And sometimes, the moving average part of the stationary time series which is derived from differencing a nonstationary one may have seasonal lags.

If d and D are nonnegative integers, then $\{X_t\}$ is a **seasonal ARIMA$(p, d, q) \times (P, D, Q)_s$ process with period** s if the differenced series $Y_t = (1-B)^d(1-B^s)^D X_t$ is a causal ARMA process defined by

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t, \qquad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2),$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$, $\Phi(z) = 1 - \Phi_1 z - \cdots - \Phi_P z^P$, $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$, and $\Theta(z) = 1 + \Theta_1 z + \cdots + \Theta_Q z^Q$.

Q & A!

- Thanks!
- Read books.