Short Introduction to Time Series I
FINM 33170 and STAT 33910

Per Mykland

University of Chicago, Winter 2021

Outline

1. Time Series
   - Eliminate a trend in the absence of seasonality
   - Eliminate seasonal component.
   - Data example
2. Probability Models
   - Stationarity
   - Sample correlation function and sample covariance function
3. Basic Time Series Models
   - Moving Average Models
   - Autoregressive Models
   - Determine the order of AR(p)
   - ARMA and ARIMA Models

Time Series

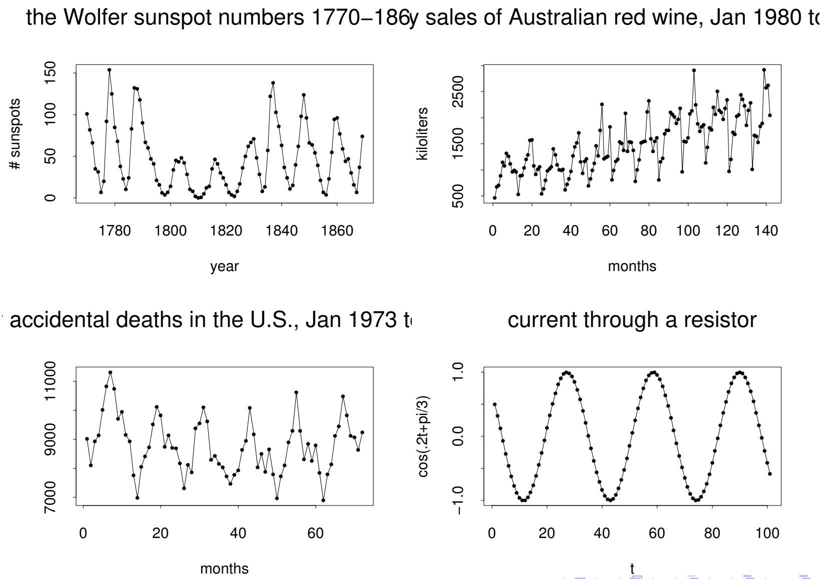
Introduction

Time Series

Unlike the classic linear regression, time series analysis focuses on time-dependent data. Often time series data are also state-dependent, for example, consider daily price series of s&p500 from 1990 to 2020, day t-th price inevitably has an impact on the price on the $t+1$-th or later day(s).

In both this and next lecture, we will discuss the probabilistic structure of basic time series models, the estimation and the diagnosis, as well as the forecasting in time series. Both simulation and data application will be covered.

Example Plots (from Brockwell and Davis)



Preliminary Screening of Time Series

From the time series plot, we can see several common features:

- trend (constant trend vs. sudden shift)
- seasonality
- outliers

The first step in the analysis of any time series (TS):

- Plot the data.
- By eyeballing the plot, one hopes to identify
  - Outliers (say, recording error in stock price at certain tick). Study the outliers separately
  - Sudden shift in level. Partition the time series into homogeneous segments

Representing the Data

After the preliminary screening, the data can be represented as a realization of the following process:

$$X_t = m_t + s_t + Y_t,$$

where m_t is a "trend component", s_t is a function with known period d referred to as a "seasonal component", and Y_t is a random stationary component. Most of the probabilistic and the statistical theories are developed on the basis of stationary process. A common task in TS is to transform the data X_t to a stationary Y_t, one can then continue with the modeling and forecasting for Y_t. At the end, the inferential results about Y are transformed back to those regarding X.
How do we transform X to Y? These involve different approaches to de-trend and/or de-season.

Eliminate a trend in the absence of seasonality

In the absence of seasonality, assuming we have a simple trend, for example, $m_t = \beta_1 + \beta_2 t$, there are three methods to remove the trend component:

- Least squares method
- Smoothing or filtering
- Differencing

Least squares method

Least squares method: This is similar to linear regression, treating Y_t as the noise term. Then one can find $\hat{\beta}_1$ and $\hat{\beta}_2$ so that $\sum_t (X_t - m_t)^2$ is minimized. The detrended version would be

$$X_t - \hat{m}_t = X_t - (\hat{\beta}_1 + \hat{\beta}_2 t).$$

Drawback: this approach can only deal with simple trend, and with fixed trend for the entire span of the data set.

Smoothing or filtering

Smoothing or filtering: This approach uses smoothing via a moving average. For example,

$$Sm(X_t) = \sum_{r=-q}^{s} a_r X_{t+r}, \text{ for } s, q > 0$$

The weights $\{a_r\}$ are usually assumed to be symmetric and normalized (i.e. $a_r = a_{-r}$ and $\sum_r a_r = 1$).
Special case:

$$Sm(X_t) = \frac{1}{2q+1} \sum_{r=-q}^{q} X_{t+r},$$

and the exponential smoothing

$$Sm(X_t) = \sum_{j=0}^{\infty} \alpha(1-\alpha)^j X_{t-j}, \text{ where } \alpha \in (0,1)$$

Drawback: how to choose the size of the smoothing window?

Differencing

Differencing: Given $X_t = m_t + Y_t$, without loss of generality assuming $m_t = \beta_1 + \beta_2 t$, then apply the difference operator Δ:

$$\begin{aligned}\Delta X_t &= X_t - X_{t-1} \\ &= m_t + Y_t - m_{t-1} - Y_{t-1} \\ &= \beta_2 + \Delta Y_t\end{aligned}$$

For higher-order polynomial functions of m_t, one only needs to invoke differencing a few more times, e.g. for $m_t = \sum_{j=0}^{p} a_j t^j$, then $\Delta^p X_t = p! a_p + \Delta^p Y_t$.

Eliminate seasonal component.

Moving average method: first estimate the trend:

- $d = 2q$, let $\hat{m}_t = \frac{1}{d}(\frac{1}{2}X_{t-q} + X_{t-q+1} + \ldots + X_{t+q-1} + \frac{1}{2}X_{t+q})$, $t = q+1, n-q$.
- $d = 2q+1$, let $\hat{m}_t = \frac{1}{d}\sum_{r=-q}^{q} X_{t+r}$, $t = q+1, n-q$.

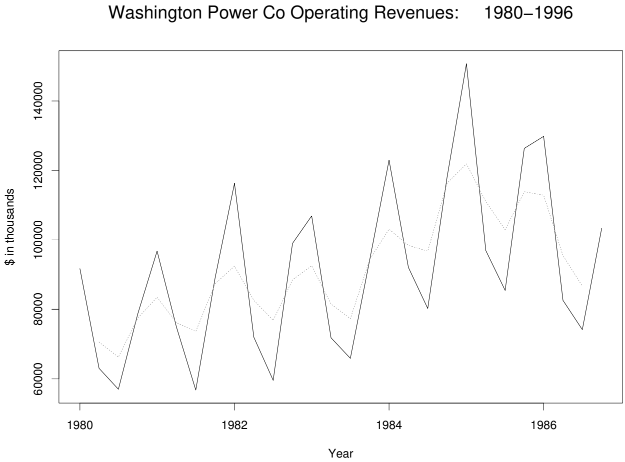
Then use moving average method on detrended data $X_t - \hat{m}_t$. See data example for more details.

Seasonal differencing: apply d-th differencing, $X_t - X_{t-d}$.

Other forms of transformation

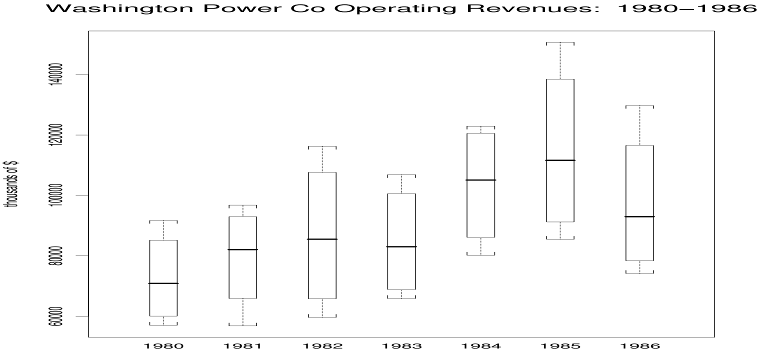
Similar to linear regression, one can use log transformation or Box-Cox transformation to handle the non-constant variance in Y_t.

Data example (from Chapter 1 in Chan (2010))



Increasing trend. Annual cycle: low in third quarter (July-September), high in first quarter (January-March). Grey line: Moving average.

Box Plot Representation



Variation within each year. One can see the increasing trend in median revenues across years.

R code: tsplot and boxplots

```
# ts-plot
wash<-ts(scan("washpower.dat"),start=1980,freq=4)
wash.ma<-filter(wash,c(1/3,1/3,1/3)) #method is implicitly convolution, i.e., MA
ts.plot(wash,wash.ma,lty=c(1,2),main="Washington Power Co Operating Revenues: 1980-1996",xlab=

# boxplots
wash.mat<-matrix(wash,nrow=4)
boxplot(as.data.frame(wash.mat),names=as.character(seq(1980,1986)),boxcol=1,medcol=1,main="Wa
```

De-seasoning the Data

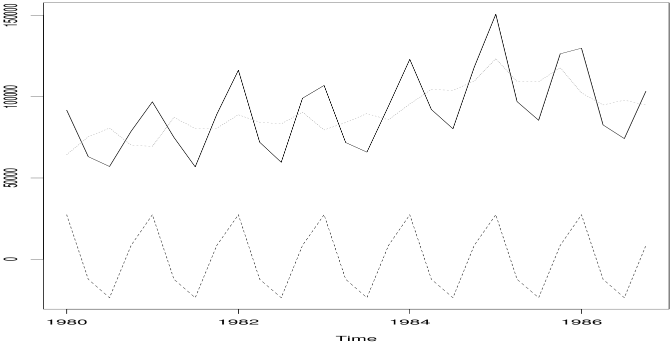
We use the moving average method to de-season the data:

- estimate trend through one compete cycle. Use $n = 28$, $d = 4$ and $q = 2$ to form $X_t - \hat{m}_t$, for $t = 3, \ldots, 26$
- compute the averages of $X_t - \hat{m}_t$ over the entire data series.
- estimate the seasonal component $\hat{s}_t$, $t = 1, 2, 3, 4$ by computing the de-meaned values of the averages. De-seasonalize data $X_t - \hat{m}_t$ is denoted by "wash.nosea" .
- re-estimate the trend from deseasonalized data $\hat{m}_t$.
- get residual $X_t - \hat{m}_t - \hat{s}_t$, denoted by "wash.res", analyze residuals to search for further structure.

R code: Data Decomposition

```
washsea.ma<-filter(wash,c(1/8,rep(1/4,3),1/8)) #hat-m_t, t=2,..26
wash.sea<-c(0,0,0,0)
for(i in 1:     2)
        {
        for(j in 1:     6){
                wash.sea[i]<-wash.sea[i]+(wash[i+4*j][[1]]-washsea.ma[i+4*j][[1]])
        }
        }
wash.sea<-(wash.sea-mean(wash.sea))/6
wash.sea1<-rep(wash.sea,7)
wash.nosea<-wash-wash.sea
wash.ma2<-filter(wash.nosea,c(1/8,rep(1/4,3),1/8))
wash.res<-wash-wash.ma2-wash.sea
write(wash.sea1,file="out.dat")
wash.seatime<-ts(scan("out.dat"),start=1980,freq=4)
ts.plot(wash,wash.nosea,wash.seatime)
wash.stl<-stl(wash,'periodic')
dwash<-diff(wash,4)
sea<-wash.stl$time.series[,1]
rem<-wash.stl$time.series[,3]
ts.plot(wash,sea,rem,dwash)
```

Data Decomposition



This figure summarizes the data decomposition. One can also use S/R's seasonal decomposition function stl

Probability Models

Probability Models

- probabilistic theory of stochastic processes
- focus on linear models
- special cases include AR, MA, ARMA, and ARIMA models.

Definition of Stochastic Process

Definition

A collection of random variables $\{X(t) : t \in \mathcal{R}\}$ is called a **stochastic process.** □

- The index t may indicate time or location. In time series analysis, t is usually a time index.
- X is defined on a given probability space $(\Omega, \mathcal{F}, P)$. For fixed t, X_t is a random variable:
$$X_t = X_t(\omega) : \Omega \to R \quad \text{for a fixed } t.$$
- For fixed ω, $X(\omega)$ is a function of t, which is called "a sample path" or a "realization" of the stochastic process.
- For example, the daily time series of Ford's stock price in the past decade is a sample path.
- All financial time series plots are based on a single sample path, time series analysis is about finding the probabilistic structure which governs the observed data series.

Depending on whether the index t takes value in an interval (say, $t \in [0, T]$, $t \in [0, \infty)$) or in a set of discrete values (say, $t = \{1, 2, \ldots, n\}$), $\{X_t\}$ is called **a continuous-time stochastic process** or **a discrete-time stochastic process**, respectively. A time series could refer to a discrete-time stochastic process that generates the data, it could also refer to the data series itself.

To understand the probabilistic model of the underlying process, one starts from the distribution of the process. A simple version would be the finite-dimensional distribution.

Definition

Let $\mathcal{T}$ be the set of all vectors $\{\mathbf{t} = (t_1, \ldots, t_n)' \in T^n : t_1 < \ldots < t_n, n = 1, 2, \ldots\}$. The **finite-dimensional distribution functions** of the stochastic process $\{X_t, t \in T\}$ are the functions $\{F_{\mathbf{t}}(\cdot), \mathbf{t} \in \mathcal{T}\}$ defined for $\mathbf{t} = (t_1, \ldots, t_n)'$ by

$$F_{\mathbf{t}}(\mathbf{x}) = P(X_{t_1} \le x_1, \ldots, X_{t_n} \le x_n), \mathbf{x} = (x_1, \ldots, x_n)' \in R^n.$$

□

One should be aware that the distribution function of X (or written as $\{X_t\}$) is concerned with the joint distribution of the entire path of X, it, in most situations, is NOT the same as the joint distribution of X at several discrete time points. The following theorem gives the condition for the existence of the distribution function for a process.

Theorem (Kolmogorov's Consistency Theorem)

The probability distribution functions $\{F_{\mathbf{t}}(\cdot), \mathbf{t} \in \mathcal{T}\}$ are the distribution functions of some stochastic process if and only if for any $n \in \{1, 2, \ldots\}$, $\mathbf{t} = (t_1, \ldots, t_n)' \in \mathcal{T}$ and $1 \leq i \leq n$,

$$\lim_{x_i \to \infty} F_{\mathbf{t}}(\mathbf{x}) = F_{\mathbf{t}(i)}(\mathbf{x}(i)),$$

where $\mathbf{t}(i)$ and $\mathbf{x}(i)$ are the (n-1)-component vectors obtained by deleting the i'th components of $\mathbf{t}$ and $\mathbf{x}$, respectively. $\square$

Stationarity

- Most techniques about handling stationary time series.
- Two types of stationarity:

Definition

$\{X_t\}$ is said to be **strictly stationary** if for all n, for all $(t_1, \dots, t_n)$, and for all s,

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{D}{=} (X_{t_1+s}, \dots, X_{t_n+s}),$$

where $\stackrel{D}{=}$ denotes "have the same distribution".

$\{X_t\}$ is said to be **weakly stationary** if

(1) $E(X_t) = \mu$, for all t,

(2) $Cov(X_t, X_{t+s})$ depends on s only.

- strict stationarity says that the joint probabilistic behavior of X remains the same after shifts in time
- weak stationarity: only restriction on first two moments of X
- strict stationary implies weakly stationary; the opposite is generally not true (one exception: Gaussian process)

Autocovariance Function

Not feasible to check strictly stationarity in practice. However:

Definition

$\{X_t : t \in T\}$ is stochastic process with $Var(X_t) < \infty$ for all $t \in T$. The **autocovariance function** $\gamma_X(\cdot, \cdot)$ of $\{X_t\}$ is defined by

$$\gamma_X(r, r+s) = Cov(X_r, X_{r+s}) = E[(X_r - EX_r)(X_{r+s} - EX_{r+s})], r, s \in T.$$

In particular, for a stationary process $\{X_t\}$, we have

$$\gamma_X(r, r+s) = \gamma_X(0, s)$$

So the autocovariance function for a stationary process can be denoted with a single index:

$$\gamma(s) = Cov(X_t, X_{t+s})$$

Also: $\rho(s) = \frac{\gamma(s)}{\gamma(0)}$ is called the **autocorrelation function** A few facts for stationary process:

- $\gamma(0) = Var(X_t) \geq 0$ for all t.
- $\gamma(s) = \gamma(-s)$.
- $|\gamma(s)| \leq \gamma(0)$ for all integer s. (why?)

Examples

Example

Suppose $\{X_t\}$ is a white noise process: $EX_t = 0, \forall t$, $Var(X_t) = \sigma^2$, and $Cov(X_t, X_s) = 0$ for $t \neq s$. Is X a stationary process?
Clearly, the expectation and the variance are time-invariant. Also the autocovariance doesn't depend on the time index. So, white noise process is stationary.

Example

Suppose $\{X_t\}$ is a random walk, that is, $X_t = \sum_{i=1}^{t} Z_i$, where $\{Z_i\}$ is a white noise process. Is X a stationary process?

- $E[X_t] = \sum_{i=1}^{t} E[Z_i] = 0$
- $Var[X_t] = \sum_{i=1}^{t} Var[Z_i] = \sigma^2 t$ depends on t
- so random walk is nonstationary!

Assumptions for rest of Lecture

- stationarity means weak stationarity
- X represents $\{X_t : t \in T\}$
- the moments $E|X|^k$ exist.

Sample correlation function and sample covariance function

In practice, $\gamma(k)$ and $\rho(k)$ are unknown, and have to be estimated from the data.

Definition

Let $\{X_t\}$ be a given time series and $\bar{X}$ is its sample mean. Then

- $\hat{\gamma}(k) = \sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})/n$ is known as the **sample autocovariance function** of X_t.
- $\hat{\rho}(k) = \hat{\gamma}(k)/\hat{\gamma}(0)$ is known as the **sample autocorrelation function** of X_t.

The plot of $\hat{\rho}(k)$ versus k is called a **correlogram.** For a stationary process (starting with its limiting distribution), the sample estimate $\hat{\rho}(k)$ approximates $\rho(k)$.

Sample Correlation and Covariance (cont'd)

Important facts:

- for an iid process, it can be shown that for each fixed k,

$$\hat{\rho}(k) \sim AN(0, 1/T) \quad \text{as } T \to \infty$$

- if $Y_t = Y$ (constant process), $\hat{\rho}(k) = 1$.
- a stationary process often exhibits short-memory behavior (short-term correlation, i.e. $\hat{\rho}(k)$ taper off for large k).
- in a nonstationary process, $\hat{\rho}(k)$ does not taper off for large values of k.
- by definition, $\hat{\rho}(0) = 1$.

Basic Time Series Models

Basic Time Series Models

We discuss four types of models in this section:

- the moving average model (MA)
- the autoregressive model (AR)
- the autoregressive moving average model (ARMA), and
- the autoregressive integrated moving average model (ARIMA).

The first three are stationary processes, while the last one is nonstationary.

Moving Average Models

Let $\{Z_t\}$ be a white noise sequence with mean zero and variance σ^2, denoted by $Z_t \sim WN(0, \sigma^2)$
(An iid process is a special case of white noise process.)
A moving average time series is formed by a weighted average of Z_t:

$$Y_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}, \quad Z_t \sim WN(0, \sigma^2). \tag{1}$$

We say that the process $\{Y_t\}$ follows a **moving average model of order** q, $MA(q)$.

MA(1)

Let $\{Y_t\}$ be a $MA(1)$ process as follows:

$$Y_t = Z_t + \theta_1 Z_{t-1}, Z_t \sim WN(0, \sigma^2).$$

Then, the expectation and the autocovariance of Y_t are:

$$\begin{aligned} E[Y_t] &= E[Z_t + \theta_1 Z_{t-1}] = E[Z_t] + \theta_1 E[Z_{t-1}] = 0 \text{ and} \\ Var[Y_t] &= Var[Z_t + \theta_1 Z_{t-1}] \\ &= Var[Z_t] + \theta_1^2 Var[Z_{t-1}] \quad \text{why the covariance is zero?} \\ &= (1 + \theta_1^2)\sigma^2 \end{aligned}$$

$$Cov(Y_t, Y_{t+1}) = Cov(Z_t + \theta_1 Z_{t-1}, Z_{t+1} + \theta_1 Z_t)$$

$$= Cov(Z_t, Z_{t+1}) + Cov(Z_t, \theta_1 Z_t) + Cov(\theta_1 Z_{t-1}, Z_{t+1}) + Cov(\theta_1 Z_{t-1}, \theta_1 Z_t)$$

$$= 0 + \theta_1 \sigma^2 + 0 + 0 = \theta_1 \sigma^2$$

and for $|k| > 1$:

$$Cov(Y_t, Y_{t+k}) = Cov(Z_t + \theta_1 Z_{t-1}, Z_{t+k} + \theta_1 Z_{t+k-1}) = 0$$

How about $Cov(Y_t, Y_{t-1})$?

MA(2)

How about $\{Y_t\} \sim MA(2)$ process? We still have $E[Y_t] = 0$, but the variance and the covariance become:

$$\begin{aligned} Var[Y_t] &= Var[Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}] \\ &= \sigma^2(1 + \theta_1^2 + \theta_2^2) \end{aligned}$$

and

$$\begin{aligned} Cov[Y_t, Y_{t+1}] &= Cov(Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}, Z_{t+1} + \theta_1 Z_t + \theta_2 Z_{t-1}) \\ &= \sigma^2(\theta_1 + \theta_1\theta_2) \\ Cov[Y_t, Y_{t+2}] &= Cov(Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}, Z_{t+2} + \theta_1 Z_{t+1} + \theta_2 Z_t) \\ &= \sigma^2\theta_2 \end{aligned}$$

General MA(q)

In general, we have the following properties for the $MA(q)$ process.

Proposition

Let $\{Y_t\}$ be the $MA(q)$ model given in (1). Then,

(i) $E[Y_t] = 0$

(ii) $\gamma(k) = Cov[Y_t, Y_{t+k}] = \begin{cases} \sigma^2 \sum_{i=0}^{q-|k|} \theta_i \theta_{i+|k|}, & |k| \leq q \\ 0, & |k| > q \end{cases}$

where $\theta_0 \triangleq 1$. In particular,

$$\gamma(0) = \sigma^2(1 + \theta_1^2 + \ldots + \theta_q^2).$$

$\square$

General MA(q) (cont'd)

Hence the autocorrelation function for a MA(q) is:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \begin{cases} \frac{\sum_{i=0}^{q-|k|} \theta_i \theta_{i+|k|}}{\sum_{i=0}^{q} \theta_i^2}, & |k| \leq q \\ 0, & |k| > q \end{cases}$$

In special case when $k = 0$, $\rho(k) = 1$.
It is worthy of mentioning that for an MA(q) model, its ACF vanishes after lag q. That is, if $\rho_Y(q) \neq 0$, but $\rho_Y(l) = 0$ for $l > q$, then the process follows a MA(q) model.
MA process has finite memory (why? Hint: ACF reflects a linear relation...)
One can see that MA model is stationary (why?). An MA(q) process has mean zero. If a constant trend μ is added such that $Y_t = \mu + Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$, then $E[Y_t] = \mu$, $\{Y_t\}$ is refered as "an MA(q) process with mean μ".

Invertibility

Example

Consider an MA(1) model $Y_t = Z_t - \theta_1 Z_{t-1}$. We have shown that its ACF satisfies

$$\rho_Y(k) = \begin{cases} 1, & k = 0 \\ \frac{-\theta_1}{1+\theta_1^2}, & |k| = 1 \\ 0, & |k| > 1 \end{cases}$$

Now if we express the residual Z_t by the Y's,

$$\begin{aligned} Z_t &= Y_t + \theta_1 Z_{t-1} \\ &= Y_t + \theta_1 (Y_{t-1} + \theta_1 Z_{t-2}) \\ &= Y_t + \theta_1 Y_{t-1} + \theta_1^2 Z_{t-2} \\ &\dots \\ &= Y_t + \theta_1 Y_{t-1} + \theta_1^2 Y_{t-2} + \theta_1^3 Y_{t-3} + \dots \end{aligned} \tag{2}$$

It is clear that if $|\theta_1| < 1$, equation (2) converges, in this case, we say $\{Y_t\}$ is invertible. From the viewpoint of interpreting the residuals, it is desirable to have an invertible process.

B: The Backshift Operator

$BZ_t = Z_{t-1}$, $B^2 Z_t = Z_{t-2}$ etc. We can then express an MA(q) model $\{Y_t\}$ as:

$$Y_t = \theta(B)Z_t, \quad \text{where } \theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q.$$

Theorem

An MA(q) model $\{Y_t\}$ is invertible if the roots of the equation $\theta(B) = 0$ all lie outside the unit circle.

Example

Consider a process $\{X_t\}$ follows the equation:

$$Z_t = X_t + 0.5Z_{t_1}.$$

Is X invertible? Since the polynomial

$$\begin{aligned} \theta(z) &= 1 - 0.5z \overset{\text{set}}{=} 0 \\ &\Rightarrow z = 2 > 1 \end{aligned}$$

Hence X is invertible.

Autoregressive Models

An $AR(p)$ **model** $\{Y_t\}$ satisfies

(i) $\phi(B)Y_t = Z_t$, where $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^q$, that is,

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + Z_t, \text{ for all } t$$

(ii) $\{Y_t\}$ is stationary

Also, $\{Y_t\}$ is said to be an $AR(p)$ process with mean μ if $\{Y_t - \mu\}$ is an $AR(p)$ process.

Causal Autoregressive Processes

Definition

A process $\{Y_t\}$ is said to be **causal** if there exists a sequence of constants $\{\psi_j\}$'s such that $Y_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ with $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

For an $AR(p)$ model $\phi(B)Y_t = Z_t$, one can write:

$$Y_t = \phi^{-1}(B)Z_t = \psi(B)Z_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i} \tag{3}$$

The following theorem gives the condition when an $AR(p)$ process is causal:

Theorem

An AR(p) process is causal if the roots of the characteristic polynomial

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$$

all lie outside the unit circle.

Example

Consider the following two processes:

$$X_t = 1.4X_{t_1} + Z_t,$$

and

$$Y_t + .4Y_{t-1} - .12Y_{t-2} = Z_t.$$

Is X or Y causal? Identify the models for X and Y.

Answer:

$\{X_t\}$ has characteristic polynomial

$$\phi(z) = 1 - 1.4z,$$

with root $z = 1/1.4 < 1$. So, $\{X_t\}$ is not causal.

$\{Y_t\}$ has characteristic polynomial

$$\phi(z) = 1 + .4z - .12z^2,$$

it has two roots, $z = -1.67, 5$, both lie outside the unit circle. Hence $\{Y_t\}$ is causal, it is an AR(2) model.

AR(1)

Causal AR(1) model,

$$Y_t = \phi_1 Y_{t-1} + Z_t, \quad Z_t \sim WN(0, \sigma^2).$$

Mean, variance and covariance:

$$E[Y_t] = E[\sum_{i=0}^{\infty} \psi_i Z_{t-i}] = 0.$$

and

$$Var[Y_t] = Var[\phi_1 Y_{t-1} + Z_t] = \phi^2 Var[Y_{t-1}] + \sigma^2 \text{ why?}$$

because Y is stationary, $Var[Y_t] = Var[Y_{t-1}]$, thus

$$Var[Y_t] = \frac{\sigma^2}{1 - \phi^2}.$$

AR(1): Autocovariance Function (ACF) at Lag 1

Notice: $Cov(Y_t, Z_{t+k}) = 0$ for $k > 0$ (why?). Also

$$Cov(Y_t, Z_t) = E[(\phi_1 Y_{t-1} + Z_t)Z_t] = \sigma^2$$

Hence:

$$\begin{aligned}\gamma_Y(1) = Cov(Y_t, Y_{t+1}) &= E[Y_t Y_{t+1}] \\ &= E[(\phi_1 Y_{t-1} + Z_t)(\phi_1 Y_t + Z_{t+1})] \\ &= \phi_1^2 E[Y_{t-1} Y_t] + 0 + \phi_1 E[Z_t Y_t] + 0 \\ &= \phi_1^2 \gamma_Y(1) + \phi_1 \sigma^2\end{aligned}$$

thus,

$$\gamma_Y(1) = \frac{\phi_1}{1-\phi_1^2}\sigma^2.$$

and

$$\rho_Y(1) = \phi_1.$$

AR(1): Autocovariance Function (ACF) at Lag k

$$
\begin{aligned}
\text{For } |k| > 1,\ \gamma_Y(k) &= E[Y_t Y_{t+k}] \\
&= E[Y_t(\phi_1 Y_{t+k-1} + Z_{t+k})] \\
&= \phi_1 \underbrace{E[Y_t Y_{t+k-1}]}_{\gamma(k-1)} + 0 \\
&= \phi_1(\phi_1 \gamma(k-2) + 0) \\
&= \phi_1^2 \gamma(k-2) \\
&\cdots \\
&= \phi_1^k \gamma(0)
\end{aligned}
$$

Hence for an AR(1) process, the ACF is

$$\rho(k) = \phi_1^k.$$

This suggests that **the ACF of an AR(1) process decays exponentially with rate ϕ_1 and starting value $\rho(0) = 1$ (or ρ_0).**

Example

From the two ACF plots, can you tell which series follows the model: $Y_t = 0.8Y_{t-1} + Z_t$? the sign of AR(1) coefficient?

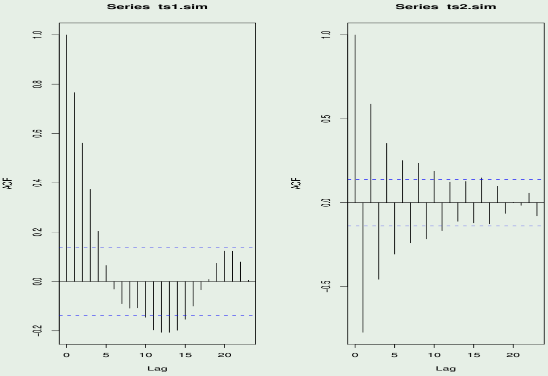


Figure: The autocorrelation function for an AR(1) model

The plots are generated from R code:

```
ts1.sim<-arima.sim(n=200,list(order=c(1,0,0), ar=.8))
ts2.sim<-arima.sim(n=200,list(order=c(1,0,0), ar=-.8))
acf(ts1.sim)
acf(ts2.sim)
```

AR(2)

For a causal AR(2) process,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t,$$

one can use similar approach to calculate the AVF and ACF, namely, for $k > 0$, multiply both sides by Y_{t-k}, and then take expectation:

$$\begin{aligned} Y_t Y_{t-k} &= (\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t) Y_{t-k} \\ &= \phi_1 Y_{t-1} Y_{t-k} + \phi_2 Y_{t-2} Y_{t-k} + Z_t Y_{t-k} \\ \underbrace{E[Y_t Y_{t-k}]}_{\gamma_k} &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} \end{aligned}$$

Equivalently, we have the equation

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \text{ for } k > 0$$

Specifically,

$$\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_{-1} = \phi_1 + \phi_2 \rho_1 \quad \Rightarrow \quad \rho_1 = \frac{\phi_1}{1 - \phi_2}$$

Simulate AR(2) Models

```
ts3.sim<-arima.sim(n=200,list(order=c(2,0,0),ar=c(0.2,0.35)))
ts4.sim<-arima.sim(n=200,list(order=c(2,0,0),ar=c(0.2,-0.35)))
ts5.sim<-arima.sim(n=200,list(order=c(2,0,0),ar=c(-0.2,0.35)))
ts6.sim<-arima.sim(n=200,list(order=c(2,0,0),ar=c(-0.2,-0.35)))
par(mfrow=c(2,2))
acf(ts3.sim)
acf(ts4.sim)
acf(ts5.sim)
acf(ts6.sim)
```

Identify AR(2) models

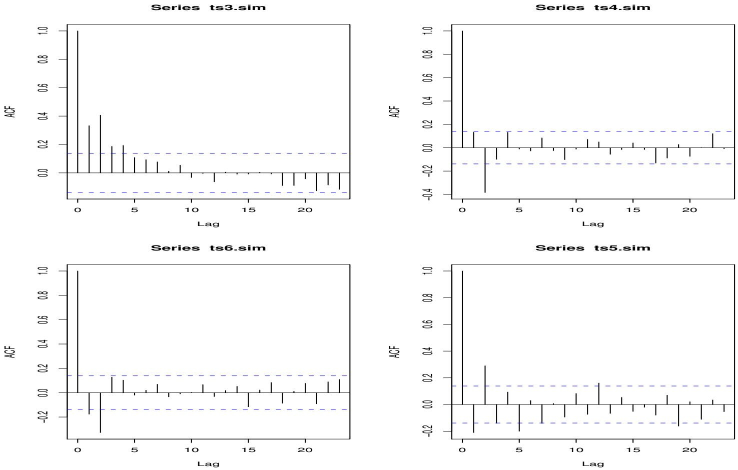


Figure: The autocorrelation function for an AR(2) model

Determine the order of AR(p)

Consider a sequence of AR models:

$$
\begin{aligned}
Y_t &= \phi_{1,1} Y_{t-1} + Z_{1t} \\
Y_t &= \phi_{1,2} Y_{t-1} + \phi_{2,2} Y_{t-2} + Z_{2t} \\
Y_t &= \phi_{1,3} Y_{t-1} + \phi_{2,3} Y_{t-2} + \phi_{3,3} Y_{t-3} + Z_{3t} \\
&\vdots
\end{aligned}
$$

where $\phi_{i,j}$ and Z_{jt} are the corresponding coefficient of Y_{t-i} and the white noise term in an $AR(j)$ model. Similar to the setup in multiple regression, we can consider how well each of the above models fits. The estimate $\hat{\phi}_{1,1}$ tells us the lag-1 impact (i.e. the impact of Y_{t-1} on Y_t); the estimate $\hat{\phi}_{2,2}$ tells us the additional contribution of lag-2 value Y_{t-2} to Y_t, $\hat{\phi}_{3,3}$ tells us the additional contribution of lag-3 value Y_{t-3} and so on.

Partial Autocorrelation Function (PACF)

In TS, $\hat{\phi}_{j,j}$ is called the lag-j sample PACF of Y_t. In an AR(p) model, Y_t clearly depends on $Y_{t-1}, \ldots, Y_{t-p}$, we expect $\hat{\phi}_{p,p}$ significantly different from zero. On the other hand, the PACF at greater lag $\hat{\phi}_{l,l}$'s, $l > p$, should be close to zero in an AR(p) model. Note that the expression of an AR(p) model states a linear relation between Y_t and its first p-lagged values $Y_{t-1}, \ldots, Y_{t-p}$.
The idea of PACF is supported by the asymptotic distribution of the estimated ϕ_j: subject to regularity conditions,

- $\hat{\phi}_{p,p}$ converges to ϕ_p as the sample size T goes to infinity.
- $\hat{\phi}_{l,l}$ converges to 0 as the sample size T goes to infinity, for all $l > p$.
- The asymptotic variance of $\hat{\phi}_{l,l}$ is $1/T$ for $l > p$.

Information Criteria

The most popular ones are AIC, AICC and BIC.
Akaike Information Criterion (AIC) is generally definded as

$$AIC = -2\log(\text{likelihood}) + \underbrace{2 \text{ number of parameters}}_{\text{penalty function}}.$$

For an AR(p) model, the number of paramers is $p+1$. In theory, for $Y_t \sim AR(p)$, $AIC(p)$ gives minimum value among AIC(l) for $l = 0, 1, \dots$. In practice, one never knows the true order, and should consider a clas of competing models with similar low AIC values.
AIC has the drawback of overfitting. In this respect, AICC and BIC are better criteria, we shall return to the topic later.

ARMA models

The ARMA(p,q) model is a combination of the AR and the MA models. $\{Y_t\}$ is ARMA(p,q) if

$$\phi(B)Y_t = \theta(B)Z_t,$$

$$\text{where} \quad \begin{aligned} \phi(B) &= 1 - \phi_1 B - \ldots - \phi_p B^p \\ \theta(B) &= 1 - \theta_1 B - \ldots - \theta_q B^q \end{aligned}$$

where $\phi(B)$ and $\theta(B)$ have no common roots. We also require that $\phi(B)$ be causal and that $\theta(B)$ invertible. $\{Y_t\}$ is then said to be causal and invertible. In this case:

$$Y_t = \psi(B)Z_t \text{ where } \psi(B) = \frac{\theta(B)}{\phi(B)}$$

$$Z_t = \pi(B)Y_t \text{ where } \pi(B) = \frac{\phi(B)}{\theta(B)}$$

Thus $\{Y_t\}$ has a representation both as an infinite order AR, and an infinite order MA process.

Example

Let $Y_t - \phi Y_{t-1} = Z_t - \theta Z_{t-1}$ be an ARMA(1,1) model with $\phi = 0.5$ and $\theta = 0.3$. Then

$$\begin{aligned}\psi(B) = \frac{\theta(B)}{\phi(B)} &= \frac{1-0.3B}{1-0.5B} \\ &= (1-0.3B)\left(1+0.5B+(0.5)^2B^2+...\right) \\ &= 1+0.2B+0.1B^2+0.05B^2+...\end{aligned}$$

In general, $\psi_i = 0.2 \times (0.5)^{i-1}$ for $i = 1, 2, ...,$ with $\psi_0 = 1$. It follows that

$$\begin{aligned}\rho(k) = \frac{Cov(Y_t, Y_{t+k})}{Var(Y_t)} &= \frac{Cov(\sum_{i=0}^{\infty}\psi_i Z_{t-i}, \sum_{j=0}^{\infty}\psi_j Z_{t+k-j})}{Var(\sum_{i=0}^{\infty}\psi_i Z_{t-i})} \\ = \frac{\sum_{i=0}^{\infty}\psi_i\psi_{k+i}}{\sum_{i=0}^{\infty}\psi_i^2} &= \sim (0.5)^k\end{aligned}$$

Main advantage of ARMA representation: parsimony.

ARIMA models

This is a generalization of ARMA models which permits detrending. $\{Y_t\}$ is said to be ARIMA(p,d,q) if $(1-B)^d Y_t = W_t$ and $\{W_t\}$ is ARMA(p,q). Usually, d is small, say $d \leq 3$.

Example

Let $Y_t = \alpha + \beta t + N_t$, so that $(1-B)Y_t = \beta + Z_t$, where $Z_t = N_t - N_{t-1}$. Thus, $(1-B)Y_t$ satisfies an MA(1) model, although a noninvertible one. Further, the original process $\{Y_t\}$ is an ARIMA(0,1,1) model.

Example

Consider an ARIMA(0,1,0), a random walk model:

$$Y_t = Y_{t-1} + Z_t.$$

If $Y_0 = 0$, then $Y_t = \sum_{i=1}^{t} Z_i$, which implies that $var(Y_t) = t\sigma^2$. Thus, in addition to being noncausal, this process is also nonstationary, as its variance changes with time.

Stock Price as ARIMA

A fundamental illustration of an ARIMA model is as follows. Let P_t denote the price of a stock at the end of day t. Define the return of the stock as $r_t = (P_t - P_{t-1})/P_{t-1}$. A simple Taylor expansion of the log function leads to the following expression:

$$\begin{aligned} r_t &= \frac{P_t - P_{t-1}}{P_{t-1}} \\ &\approx \log\left(1 + \frac{P_t - P_{t-1}}{P_{t-1}}\right) \\ &= \log\left(\frac{P_t}{P_{t-1}}\right) \\ &= \log P_t - \log P_{t-1} \end{aligned}$$

Now let $Y_t = \log P_t$. The return r_t on stock is often an white noise process, or $r_t = Z_t$. The above derivation therefore suggests that $\{Y_t\}$ follows an ARIMA(0,1,0) model. (Here ignoring an Itô correction.)

In practice, to model nonstationary data, use the following steps:

1. Look at the plot of the data, and apply differencing as many times as required, to make the diffenences series look stationary.
2. Then fit an ARMA(p, q) model to the differences series.

Example

Let $\{Y_t\}$ be an ARIMA(1,1,1) model that follows

$$(1-\phi B)(1-B)Y_t = Z_t - \theta Z_{t-1}.$$

Then $W_t = (1-B)Y_t = Y_t - Y_{t-1}$. Therefore,

$$\sum_{k=1}^{t} W_k = \sum_{k=1}^{t}(Y_k - Y_{k-1}) = Y_t - Y_0 = Y_t \quad \text{if} \quad Y_0 = 0.$$

Hence, Y_t is recovered from W_t by summing, hence integrated. The differneced process $\{W_t\}$ satisfies an ARMA(1,1) model.