Lecture 2: Linear Time Series (TS) Models

Instructor: Lin Juan

SOE & WISE, Xiamen University

Outline

In this chapter,

- discuss basic theories of linear time series analysis, include stationarity, dynamic dependence, autocorrelation function, modeling, and forecasting;
- introduce some simple econometric models useful for analyzing financial time series;
- apply the models to asset returns;

Time Series

- Financial TS: collection of financial observations generated sequentially through time.
- Example: log return r_t of a stock. Denoted by $\{r_1, r_2, \cdots, r_T\}$ (T data points)
- Data are ordered with respect to time;
- Successive observations are usually expected to be dependent.
- Time series analysis studies the dependence among adjacent observations.

Strong Stationarity

Definition 1

A time series $\{r_t\}$ is said to be strictly stationary if the joint distribution of $(r_{t_1}, \cdots, r_{t_k})$ is identical to that of $(r_{t_1+h}, ..., r_{t_k+h})$ for $\forall t_1, \cdots, t_k$ and h. That is,

$$\Pr(r_{t_1} \leq c_1, \cdots, r_{t_k} \leq c_k) = \Pr(r_{t_1+h} \leq c_1, \cdots, r_{t_k+h} \leq c_k)$$

for $\forall t_1, \cdots, t_k, h$ and $(c_1, \cdots, c_k)$ and k.

Strong Stationarity

In other words, strict stationarity requires that the joint distribution of $(r_{t_1}, \cdots, r_{t_k})$ is invariant under time shift.

This is a very strong condition that is hard to verify empirically.

i.i.d. (independent and identically distributed) sequence $\{\varepsilon_t\}$ is strictly stationary.

Weak Stationarity

A weaker version of stationarity is often assumed.

Definition 2

A time series $\{r_t\}$ is weakly stationary (or second order stationary or covariance stationary) if $E(r_t^2) < \infty$ and both $E(r_t)$ and $cov(r_t, r_{t-j})$ for any integer j, do not depend on t.
More specifically, $\{r_t\}$ is weakly stationary if
(1) $\mu = E(r_t)$, which is a constant;
(2) $Cov(r_t, r_{t-j}) = \gamma_j$, which only depends on j.

Weak Stationarity

Definition 3

The lag-j autocovariance function (ACVF) of a covariance stationary scalar process $\{r_t\}$ is defined,

$$\gamma_j = Cov(r_t, r_{t-j}) = E[(r_t - \mu)(r_{t-j} - \mu)],$$

where $\mu = E[r_t]$ and $j = 0, \pm 1, \pm 2 \ldots$

It has two important properties:

- $\gamma_0 = \text{Var}(r_t)$
- $\gamma_{-j} = \gamma_j$

Weak Stationarity

The second property ($\gamma_{-j} = \gamma_j$) holds because

$$
\begin{aligned}
Cov(r_t, r_{t-(-j)}) &= Cov(r_{t-(-j)}, r_t) \\
&= Cov(r_{t+j}, r_t) \\
&= Cov(r_{t_1}, r_{t_1-j}),
\end{aligned}
$$

where $t_1 = t + j$. Note that $Cov(r_1, r_2) = Cov(r_2, r_1)$ in the first equality.

Weak Stationarity

What does weak stationarity mean in practice?

Past: time plot of $\{r_t\}$ fluctuate with constant variation around a constant level.

Future: the first 2 moments of future r_t are the same as those of the data so that meaningful inferences can be made.

Weak Stationarity

- Implicitly in the condition of weak stationarity, we assume that the first two moments of r_t are finite.
- From the definitions, if r_t is strictly stationary and its first two moments are finite, then r_t is also weakly stationary. The converse is not true in general.
- However, if the time series r_t is normally distributed, then weak stationarity is equivalent to strict stationarity.
- In the literature, usually stationarity means weak stationarity, unless otherwise specified.

Unconditional moments of the returns

- Unconditional mean (or expectation) of returns:

$$\mu = E(r_t)$$

- Unconditional variance (variability) of returns:

$$\sigma^2 = \mathrm{Var}(r_t) = E[(r_t - \mu)^2]$$

Weak Stationarity

- Covariance stationarity requires that both the unconditional mean and unconditional variance are finite and do not change with time.
- Note that covariance stationarity only applies to unconditional moments and not conditional moments, and so a covariance stationary process may have a varying conditional mean, for example, ARMA model.

Weak Stationarity

- In the finance literature, it is common to assume that an asset return series is weakly stationary.
- This assumption can be checked empirically provided that a sufficient number of historical returns are available.

White Noise

Definition 4

A process $\{e_t\}_{t=-\infty}^{\infty}$ is called a white noise, written as $e_t \sim WN(0, \sigma^2)$, if it satisfies the conditions,

$$E(e_t) = 0, \quad E(e_t^2) = \sigma^2, \; E(e_t e_s) = 0 \text{ for } t \neq s$$

- White noise processes are serially uncorrelated, linearly unforecastable.
- White noise processes are not generally independent. It's possible that $Cov(e_t^2, e_s^2) \neq 0$, although e_t and e_s are uncorrelated.
- There are some stronger versions of definitions.

Independent (Strong) White Noise

Definition 5

A process $\{e_t\}_{t=-\infty}^{\infty}$ is called an independent (strong) white noise if it satisfies the conditions,

$$E(e_t) = 0$$
$$E(e_t^2) = \sigma^2$$
$$e_t, e_s \text{ independent for } t \neq s$$

Conditional moments of Strong White Noise

Let $\Omega_{t-1} = e_{t-1}, e_{t-2}, \ldots$

- Conditional mean

$$E(e_t|\Omega_{t-1}) = 0$$

- Conditional variance

$$var(e_t|\Omega_{t-1}) = E[(e_t - E(e_t|\Omega_{t-1}))^2|\Omega_{t-1}] = \sigma^2$$

Gaussian White Noise

Definition 6

A process $\{e_t\}_{t=-\infty}^{\infty}$ is called a Gaussian white noise process if it satisfies the conditions,

$$
\begin{aligned}
&E(e_t) = 0 \\
&E(e_t^2) = \sigma^2 \\
&e_t, e_s \text{ independent for } t \neq s \\
&e_t \sim N(0, \sigma^2)
\end{aligned}
$$

Serial (or auto-) correlations

For an asset return r_t, simple models attempt to capture the linear relationship between r_t and information available prior to time t.

In particular, correlations between the variable of interest and its past values become the focus of linear time series analysis. These correlations are referred to as serial correlations or autocorrelations. They are the basic tool for studying a stationary time series.

Serial (or auto-) correlations

The lag-jth autocorrelation function (ACF) of a covariance stationary scalar process $\{r_t\}$ is defined:

$$\rho_j = Corr(r_t, r_{t-j}) = \frac{Cov(r_t, r_{t-j})}{\sqrt{Var(r_t)Var(r_{t-j})}} = \frac{\gamma_j}{\gamma_0}.$$

where $j = 0, \pm 1, \pm 2....$ It has two important properties:

- $\rho_0 = 1$
- $\rho_j = \rho_{-j}$.

Autocorrelation function (ACF)

$$\rho_j = \frac{\gamma_j}{\gamma_0}, \quad j = 0, 1, 2, \ldots$$

- It shows us how the dependence pattern alters with the lag;
- $\rho_j \in [-1, 1]$;
- Existence of serial correlations implies that the return is predictable, indicating market inefficiency.

Sample autocorrelation

- sample autocorrelation

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^{T}(r_t - \bar{r})(r_{t-l} - \bar{r})}{\sum_{t=1}^{T}(r_t - \bar{r})^2}, \quad 0 \leq l \leq T - 1$$

  where $\bar{r}$ is the sample mean and T is the sample size.

- Correlogram: a graphical representation of the sample autocorrelation, a plot of $\hat{\rho}_l$ against l.

Sample autocorrelation

- If $\{r_t\}$ is an iid sequence satisfying $E(r_t^2) < \infty$, then $\hat{\rho}_l$ is asymptotically normal with mean zero and variance $1/T$ for any fixed positive integer l.
- More generally, if $\{r_t\}$ is a weakly stationary time series satisfying $r_t = \mu + \sum_{i=0}^{q} \theta_i e_{t-i}$, where $\theta_0 = 1$ and $\{e_j\}$ is a sequence of iid random variables with mean zero, then $\hat{\rho}_l$ is asymptotically normal with mean zero and variance $(1 + 2\sum_{i=1}^{q} \rho_i^2)/T$ for $l > q$.

Testing individual ACF

For a given positive integer l, we test $H_0 : \rho_l = 0$ vs $H_a : \rho_l \neq 0$

$$t = \frac{\hat{\rho}_l}{\sqrt{(1 + 2\sum_{i=1}^{l-1} \hat{\rho}_i^2)/T}}.$$

If $\{r_t\}$ is a stationary Gaussian series satisfying $\rho_j = 0$ for $j > l$, the t ratio is asymptotically distributed as a standard normal random variable.

Decision rule: Reject H_0 if $|t| > Z_{\alpha/2}$ or $p-$value less than α, where $Z_{\alpha/2}$ is the $100(1 - \alpha/2)$th percentile of the standard normal distribution.

Portmanteau Test

Jointly test that several autocorrelations of r_t are zero.

$H_0: \ \rho_1 = \cdots = \rho_m = 0 \ vs \ H_1: \ \rho_i \neq 0$ for some $i \in \{1, ..., m\}$

Ljung-Box statistics:

$$Q(m) = T(T+2) \sum_{j=1}^{m} \frac{\hat{\rho}_j^2}{T-j}.$$

Asym. chi-squared dist with m degrees of freedom.

Portmanteau Test

Ljung-Box statistics:

$$Q(m) = T(T+2)\sum_{j=1}^{m}\frac{\hat{\rho}_j^2}{T-j}.$$

Asym. chi-squared dist with m degrees of freedom.

Decision rule: Reject H_0 if $Q(m) > \chi_m^2(\alpha)$ or p-value is less than α, where $\chi_m^2(\alpha)$ denotes the $100(1-\alpha)$th percentile of a chi-squared distribution with m degrees of freedom.

Monthly returns of IBM stock

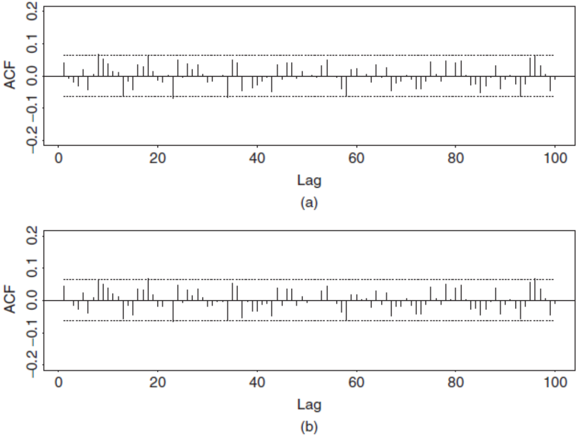


Figure 2.1 Sample autocorrelation functions of monthly (a) simple returns and (b) log returns of IBM stock from January 1926 to December 2008. In each plot, two horizontal dashed lines denote two standard error limits of sample ACF.

Monthly returns of IBM stock

$$H_0: \ \rho_1 = \cdots = \rho_m = 0 \ vs \ H_1: \ \rho_i \neq 0$$

for some $i \in \{1, ..., m\}$.

Monthly returns of IBM stock from 1926 to 2008.

- R_t: $Q(5) = 3.37(0.64)$ and $Q(10) = 13.99(0.17)$.
- r_t: $Q(5) = 3.52(0.62)$ and $Q(10) = 13.39(0.20)$.

Implication: Monthly IBM stock returns do not have significant serial correlations.

Monthly returns of value-weighted index

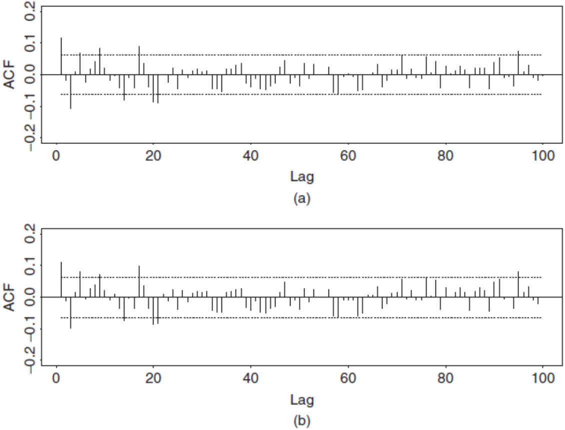


Figure 2.2 Sample autocorrelation functions of monthly (a) simple returns and (b) log returns of value-weighted index of U.S. markets from January 1926 to December 2008. In each plot, two horizontal dashed lines denote two standard error limits of sample ACF.

Monthly returns of value-weighted index

$$H_0: \ \rho_1 = \cdots = \rho_m = 0 \ vs \ H_a: \ \rho_i \neq 0$$

for some $i \in \{1, ..., m\}$.

Monthly returns of CRSP value-weighted index from 1926 to 2008.

- R_t: $Q(5) = 29.71$ and $Q(10) = 39.55$.
- r_t: $Q(5) = 28.38$ and $Q(10) = 36.16$.

The p values of these four test statistics are all less than 0.001. Therefore, monthly returns of the value-weighted index are serially correlated.

Monthly returns of value-weighted index

- Nonsynchronous trading might explain the existence of the serial correlations, among other reasons.
- Assets are exchanged in different time periods and with different frequency, and therefore they incorporate the new information at different times.
- This fact can induce positive serial correlation in stock index returns (in particular equally weighted.)

Univariate Linear Processes

- The behaviour of sample autocorrelations of the value-weighted index returns indicates that for some asset returns it is necessary to model the serial dependence.
- Linear models can be extremely useful for quantifying the association between variables, even though we don't believe the model.
- All models are wrong but some are useful — Box, G. E. P.; Draper, N. R. (1987), Empirical Model-Building and Response Surfaces, John Wiley & Sons.

A proper perspective

At a time point t,

- Available data: $\{r_1, r_2, \cdots, r_{t-1}\} \equiv \Omega_{t-1}$
- What are the important statistics in practice? Conditional quantities, not unconditional

A proper perspective

In other words, given information Ω_{t-1}

$$\begin{aligned} r_t &= \mu_t + e_t \\ &= E(r_t|\Omega_{t-1}) + \sigma_t \varepsilon_t \end{aligned}$$

- μ_t: conditional mean of r_t
- e_t: shock or innovation at time t
- ε_t: an *i.i.d.* sequence with mean zero and variance 1
- σ_t: conditional standard deviation (commonly called volatility in finance)

A proper perspective

Conditional mean:

$$\mu_t = E(r_t|\Omega_{t-1})$$

It is the best predictor of r_t in mean square error, i.e. for any $g_t \in \Omega_{t-1}$, we have

$$E(r_t - g_t)^2 > E(r_t - \mu_t)^2 \text{ if } g_t \neq \mu_t$$

A proper perspective

Given information Ω_{t-1},

$$\begin{aligned} r_t &= \mu_t + e_t \\ &= E(r_t|\Omega_{t-1}) + \sigma_t \varepsilon_t \end{aligned}$$

- Traditional TS modeling is concerned with μ_t
- Model for μ_t: mean equation
- Model for σ_t^2: volatility equation

Wold's Theorem

- If r_t is a zero-mean covariance stationary process, then it can be written as an infinite order moving average, also known as a general linear process

$$r_t = \sum_{i=0}^{\infty} \theta_i e_{t-i},$$

where $e_t \sim WN(0, \sigma^2)$.

Wold's Theorem

$$r_t = \sum_{i=0}^{\infty} \theta_i e_{t-i},$$

- Normalization: $\theta_0 = 1$
- Square summability
  Because $Var(r_t) < \infty$, $\sum_{i=0}^{\infty} \theta_i^2 < \infty$.
  $\{\theta_i^2\}$ must be a convergent sequence, implying that $\theta_i^2 \to 0$ as $i \to \infty$. It means for a stationary series, impact of the remote shock e_{t-i} on the return r_t vanishes as i increases.

Innovations

- Time series models are constructed as linear functions of fundamental forecasting errors e_t, also called innovations or shocks.
- These basic building blocks are white noise.
- In general, if you see an error e_t, it should be interpreted as white noise.

Interpretation of Wold's Theorem

- This implies that if a process is stationary we immediately know how to write a model for it.
- Problem: we might need to estimate a lot of parameters (in most cases, an infinite number of them!)
- ARMA models: they are an approximation to the Wold representation. This approximation is more parsimonious (=less parameters)

Univariate Linear Processes

- 3 basic time series models
  - Moving-average (MA) models;
  - Autoregressive (AR) models;
  - Autoregressive moving average (ARMA)
- By Wold's Theorem, they are all in the general linear process framework.
- They are linear functions of stochastic errors (white noise).

Important properties of a model

- Stationarity condition
- Basic properties: mean, variance, serial dependence
- Empirical model building: specification, estimation, & checking
- Forecasting

Outline

1 Basic concepts

2 Univariate Linear Processes

- Moving average processes
- Simple AR models
- Simple ARMA model
- Forecasting using an ARMA model

Moving-average (MA) model

- Model with finite time lags of memory!
- Some daily stock returns have minor serial correlations. Can be modeled as MA or AR models.

MA(1) process

- The general form of a first-order moving average, or MA(1) process is,

$$r_t = \mu + e_t - \theta e_{t-1},$$
$$e_t \sim WN(0, \sigma^2).$$

- The MA coefficient θ controls the degree of serial correlation. It may be positive or negative;
- The innovations e_t impact r_t over two periods
  - An contemporaneous (same period) impact
  - A one-period delayed impact

Back-shift (lag) operator

A useful notation in time series analysis,

- Definition: $Br_t = r_{t-1}$ or $Lr_t = r_{t-1}$.
- $B^2 r_t = B(Br_t) = Br_{t-1} = r_{t-2}$.

B (or L) means time shift! Lr_t is the value of the series at time $t-1$.

Back-shift (lag) operator

Suppose that the daily log returns are

Day	1	2	3	4
r_t	0.017	-0.005	-0.014	0.021

Answer the following questions:

- $r_2 = -0.005.$
- $Lr_3 = -0.005.$
- $L^2 r_5 = -0.014.$

Lag Operator Notation

- We can write the MA(1) as

$$r_t = \mu + e_t - \theta e_{t-1} = \mu + (1 - \theta L)e_t$$

or

$$r_t = \mu + \theta(L)e_t$$

where $\theta(L) = 1 - \theta L$ is a function of the lag operator.

MA(1) process is weakly stationary

- Unconditional mean: $E(r_t) = \mu$;
- Unconditional variance: $Var(r_t) = (1 + \theta^2)\sigma^2$;
- Unconditional autocovariance: $\gamma_1 = -\theta\sigma^2$ and $\gamma_j = 0$ for $j > 1$.

Unconditional Mean of MA(1)

Consider a MA(1) process,

$$r_t = \mu + e_t - \theta e_{t-1}, \;\; e_t \sim WN(0, \sigma^2)$$

- Unconditional mean

$$E(r_t) = \mu + E(e_t) - \theta E(e_{t-1}) = \mu$$

Unconditional Variance of MA(1)

Consider a MA(1) process,

$$r_t = \mu + e_t - \theta e_{t-1}, \quad e_t \sim WN(0, \sigma^2)$$

- Unconditional variance

$$Var(r_t) = Var(e_t) + \theta^2 Var(e_{t-1}) = (1 + \theta^2)\sigma^2,$$

where $E(e_{t-1}e_t) = 0$ follows from the white noise assumption.

Autocovariances of MA(1)

- The first autocovariance is

$$\begin{aligned}\gamma_1 &= Cov(r_t, r_{t-1}) \\ &= E[(r_t - \mu)(r_{t-1} - \mu)] \\ &= E[(e_t - \theta e_{t-1})(e_{t-1} - \theta e_{t-2})] \\ &= -\theta E(e_{t-1}^2) = -\theta\sigma^2\end{aligned}$$

- The autocovariances for $j > 1$ are

$$\gamma_j = 0$$

- Thus, r_t is not related to r_{t-2}, r_{t-3},...

Autocorrelations of MA(1)

Since

$$\gamma_0 = Var(r_t) = (1+\theta^2)\sigma^2$$
$$\gamma_1 = -\theta\sigma^2$$
$$\gamma_j = 0, \quad \text{for } j > 1$$

We have

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = -\frac{\theta}{1+\theta^2}$$
$$\rho_j = \frac{\gamma_j}{\gamma_0} = 0, \quad \text{for } j > 1$$

Autocorrelations of MA(1)

The autocorrelation function of an MA(1):

$$\rho_0 = 1$$
$$\rho_1 = \frac{\gamma_1}{\gamma_0} = -\frac{\theta}{1+\theta^2}$$
$$\rho_j = \frac{\gamma_j}{\gamma_0} = 0, \quad \text{for } j > 1$$

- The autocorrelation function of an MA(1) is zero after the first lag.
- Finite memory! MA(1) models do not remember what happen two time periods ago.

First autocorrelation

- The first autocorrelation has the opposite sign as θ,

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = -\frac{\theta}{1+\theta^2}$$

- As θ ranges from -1 to 1 (required by the invertability), ρ_1 ranges from $-1/2$ to $1/2$.

Linearly unforecatable and unforecastable

- Linearly uncorrelated
  The innovations e_t are linearly uncorrelated with past realization $(r_{t-1}, r_{t-2}, ...)$ if
  $$E[e_t r_{t-j}] = 0$$
  for any $j \geq 1$.
- White noise processes are linearly unforecastable.
- A stronger condition is unforecastable.
- The innovations e_t are unforecastable if
  $$E[e_t | \Omega_{t-1}] = 0$$
  It means that the optimal forecast for e_t is zero.

Results from probability theory

- Suppose $E(Y|X) = E(Y)$. Show $Cov(X, Y) = 0$ (Use law of iterated expectations).
  Note that

$$
\begin{aligned}
Cov[X, Y] &= E[XY] - E[X]E[Y] \\
E[XY] &= E[E(XY|X)] = E[XE(Y|X)] \\
&= E[XE(Y)] = E(Y)E(X)
\end{aligned}
$$

  Therefore, $Cov(X, Y) = 0$.

Conditional mean of MA(1)

- Assume the error $\{e_t\}$ is unforecastable: $E(e_t|\Omega_{t-1}) = 0$.
- Conditional mean (moving average)

$$\begin{aligned} E(r_t|\Omega_{t-1}) &= \mu + E(e_t|\Omega_{t-1}) - \theta E(e_{t-1}|\Omega_{t-1}) \\ &= \mu - \theta e_{t-1} \end{aligned}$$

Note that e_{t-1} is in the information set Ω_{t-1}.

Conditional variance of MA(1)

- Assume the error $\{e_t\}$ is unforecastable: $E(e_t|\Omega_{t-1}) = 0$.
- Conditional variance

$$\begin{aligned} Var(r_t|\Omega_{t-1}) &= E((r_t - E(r_t|\Omega_{t-1}))^2|\Omega_{t-1}) \\ &= E((r_t - \mu + \theta e_{t-1})^2|\Omega_{t-1}) \\ &= E(e_t^2|\Omega_{t-1}) = E(e_t^2) = \sigma^2 \end{aligned}$$

- The conditional variance, the forecast variance, and the innovation variance are all the same thing.

Invertibility

- Consider a zero-mean MA(1) model,

$$r_t = e_t - \theta e_{t-1}$$

- Rewrite as

$$e_t = r_t + \theta e_{t-1}$$

- Then lag this equation one period

$$e_{t-1} = r_{t-1} + \theta e_{t-2}$$

  Then combine

$$\begin{aligned} e_t &= r_t + \theta e_{t-1} \\ &= r_t + \theta(r_{t-1} + \theta e_{t-2}) \\ &= r_t + \theta r_{t-1} + \theta^2 e_{t-2} \end{aligned}$$

Invertibility

- Do this again

$$
\begin{aligned}
e_{t-2} &= r_{t-2} + \theta e_{t-3} \\
e_t &= r_t + \theta r_{t-1} + \theta^2 r_{t-2} + \theta^3 e_{t-3}
\end{aligned}
$$

- Repeat to infinity

$$e_t = r_t + \theta r_{t-1} + \theta^2 r_{t-2} + \theta^3 r_{t-3} + \ldots$$

Then

$$e_t = \sum_{i=0}^{\infty} \theta^i r_{t-i},$$

Invertibility

$$e_t = \sum_{i=0}^{\infty} \theta^i r_{t-i}.$$

- The current shock e_t is a linear combination of the present and past returns $\{r_t, r_{t-1}, r_{t-2}, ...\}$.
- This series converges (and the inversion exists) if $|\theta| < 1$. Such an MA(1) model is said to be invertible.
- If $|\theta| = 1$, then the MA(1) model is noninvertible.

Invertibility: Lag operator

- Recall the lag operator expression

$$r_t = (1 - \theta L)e_t$$

- We can write this as

$$(1 - \theta L)^{-1} r_t = e_t$$

- This inversion is valid if $|\theta| < 1$.

Inversion of Lag Polynomial

- What does this mean? $(1-\theta L)^{-1} r_t = e_t$.
- By taking a power series expansion (from calculus)

$$(1-\theta L)^{-1} = 1 + \theta L + \theta^2 L^2 + \theta^3 L^3 + \cdots$$

- This expansion converges if $|\theta| < 1$
- Applying this expression

$$\begin{aligned}(1-\theta L)^{-1} r_t &= (1 + \theta L + \theta^2 L^2 + \theta^3 L^3 + \cdots) r_t \\ &= r_t + \theta r_{t-1} + \theta^2 r_{t-2} + \theta^3 r_{t-3} + \ldots\end{aligned}$$

as needed

MA(2) model

- Consider an MA(2) model,

$$r_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

- The autocorrelation coefficients are,

$$\begin{aligned}\gamma_j &= Cov(r_t, r_{t-j}) \\ &= E[(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2})(e_{t-j} - \theta_1 e_{t-j-1} - \theta_2 e_{t-j-2})]\end{aligned}$$

- When $j = 0$,

$$\begin{aligned}\gamma_0 &= E[(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2})(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2})] \\ &= (1 + \theta_1^2 + \theta_2^2)\sigma^2\end{aligned}$$

MA(2) model

- When $j = 1$,

$$\begin{aligned}\gamma_1 &= E[(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2})(e_{t-1} - \theta_1 e_{t-2} - \theta_2 e_{t-3})] \\ &= (-\theta_1 + \theta_1\theta_2)\sigma^2\end{aligned}$$

- When $j = 2$,

$$\begin{aligned}\gamma_2 &= E[(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2})(e_{t-2} - \theta_1 e_{t-3} - \theta_2 e_{t-4})] \\ &= -\theta_2\sigma^2\end{aligned}$$

- Therefore,

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}, \ \rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, \ \rho_j = 0, \ \text{for } j > 2$$

The ACF of MA(2) model cuts off at lag 2.

MA(q) Process

- The moving average process of order q, or MA(q), is

$$r_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q},$$

  where $e_t \sim WN(0, \sigma^2)$.
- We can write the equation as

$$\begin{aligned} r_t &= \mu + (1 - \theta_1 L - \theta_2 L^2 - \ldots - \theta_q L^q) e_t \\ &= \mu + \Theta(L) e_t \end{aligned}$$

  where $\Theta(L)$ is a q'th order polynomial in L.

Autocorrelations of MA(q) Process

The first q autocorrelations of MA(q) are non-zero, the autocorrelations above q are zero because:

$$E(r_t) = \mu$$
$$\gamma_0 = Var(r_t) = \sigma^2(1 + \theta_1^2 + ... + \theta_q^2)$$
$$\gamma_k = \sigma^2(-\theta_k + \theta_{k+1}\theta_1 + ... + \theta_q\theta_{q-k}) \text{ for } k = 1, 2, ..., q$$
$$\gamma_k = Cov(r_t, r_{t-k}) = 0 \text{ if } k > q$$

Conditional moments of MA(q) Process

The moving average process of order q, or MA(q), is

$$r_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q},$$

The conditional moments are

$$\begin{aligned} E(r_t|\Omega_{t-1}) &= \mu + E(e_t|\Omega_{t-1}) - \theta_1 E(e_{t-1}|\Omega_{t-1}) \\ &\quad - ... - \theta_q E(e_{t-q}|\Omega_{t-1}) \\ &= \mu - \theta_1 e_{t-1} - ... - \theta_q e_{t-q} \\ Var(r_t|\Omega_{t-1}) &= E((r_t - E(r_t|\Omega_{t-1}))^2|\Omega_{t-1}) \\ &= E(e_t^2|\Omega_{t-1}) = \sigma^2 \end{aligned}$$

Invertibility of MA(q) model

Theorem 7

If $r_t \sim MA(q)$, r_t is an invertible process if and only if the roots of the polynomial

$$\Theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \ldots - \theta_q z^q$$

have all modulus greater than one.

The modulus of a real number x is its absolute value $|x|$.

The modulus of a complex number $a + bi$ is $\sqrt{a^2 + b^2}$.

Example

Is $r_t = e_t + \frac{5}{2}e_{t-1} - \frac{3}{2}e_{t-2}$ invertible?

Note that $\Theta(z) = 1 + \frac{5}{2}z - \frac{3}{2}z^2$. By the quadratic formula, the roots of $\Theta(z)$ are 2 and $-1/3$.

It's not invertible.

Example

Is $r_t = e_t - \frac{1}{6}e_{t-1} - \frac{1}{6}e_{t-2}$ invertible?

Note that $\Theta(z) = 1 - \frac{1}{6}z - \frac{1}{6}z^2$. By the quadratic formula, the roots of $\Theta(z)$ are 2 and -3.

It's invertible.

Identifying MA Order

The ACF is useful in identifying the order of an MA model.

For a time series r_t with ACF ρ_l, if $\rho_q \neq 0$, but $\rho_l = 0$ for $l > q$, then r_t follows an MA(q) model.

Identifying MA Order

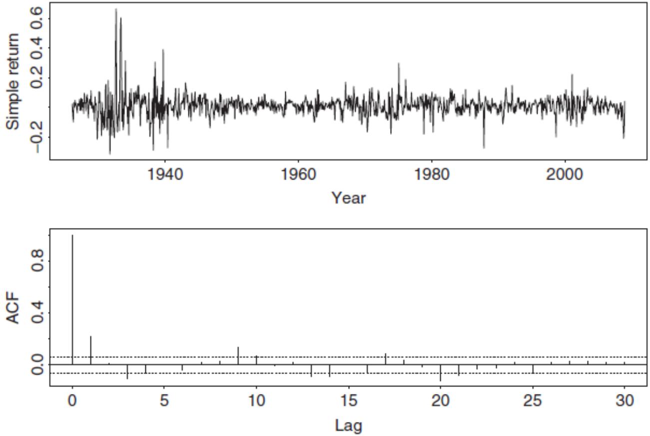


Figure 2.8 Time plot and sample autocorrelation function of monthly simple returns of CRSP equal-weighted index from January 1926 to December 2008.

Identifying MA Order

- It is seen that the series has significant ACF at lags 1, 3, and 9.
- There are some marginally significant ACF at higher lags, but we do not consider them here.
- Based on the sample ACF, the following MA(9) model,

$$r_t = \mu + e_t - \theta_1 e_{t-1} - \theta_3 e_{t-3} - \theta_9 e_{t-9}$$

is identified for the series.

Maximum Likelihood Estimation in a Nutshell

The likelihood function (often simply the likelihood) is a function of the parameters, θ, of a statistical model

$$L(Y|\theta) = f(Y_1, ..., Y_n|\theta),$$

where $Y_1, ..., Y_n$ is the sample of data, $f(\cdots|\theta)$ is a known probability density function (pdf) parameterized by the unknown vector of parameters θ.

Maximum Likelihood Estimation in a Nutshell

- The maximum likelihood estimator (MLE) is

$$\hat{\theta} = \arg\max_{\theta} L(Y|\theta) = \arg\max_{\theta} \log L(Y|\theta)$$

  That is, the MLE maximizes a conditional probability function considered as a function of θ, with its first argument – the data Y – held fixed.
- MLE answers the question "What is the most likely value of θ given the sample we have observed? "

MLE of Independent Data

- If the data are independent, identically distributed (iid) we have

$$f(Y_1, ..., Y_n|\theta) = f(Y_1|\theta) \times f(Y_2|\theta) \times \cdots \times f(Y_n|\theta)$$

  in the same way as P(A,B) = P(A)P(B) if and only if A and B are independent.

- The likelihood can then be written as the product of n probability densities

$$L(Y|\theta) = \prod_{i=1}^{n} f(Y_i|\theta) \Rightarrow \log L(Y|\theta) = \sum_{i=1}^{n} \log f(Y_i|\theta)$$

Example: Estimating mean and variance

Recall the Gaussian distribution $N(\mu, \sigma^2)$ has pdf

$$f(Y_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2}\frac{(X_i - \mu)^2}{\sigma^2}\}$$

The corresponding likelihood is

$$\begin{aligned} L(Y_i|\mu, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(Y_i - \mu)^2}{\sigma^2}\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\frac{\sum_{i=1}^{n}(Y_i - \mu)^2}{\sigma^2}\right\} \end{aligned}$$

Example: Estimating mean and variance

We have

$$\log L(Y_i|\mu,\sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\frac{\sum_{i=1}^n (Y_i-\mu)^2}{\sigma^2}$$

Taking the FOC for a maximum we have

$$\frac{\partial \log L(Y_i|\mu,\sigma^2)}{\partial \mu} = -\frac{\sum_{i=1}^n (Y_i-\mu)}{\sigma^2} = 0$$

$$\frac{\partial \log L(Y_i|\mu,\sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2}\frac{\sum_{i=1}^n (Y_i-\mu)^2}{\sigma^4} = 0$$

Therefore, $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n Y_i$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (Y_i-\hat{\mu})^2$.

Likelihood function for Time Series Models

- The standard approach to MLE we have seen so far is to obtain the likelihood function by
  - writing the density for each observation and then
  - since the observations are independent, write the likelihood as the product of these densities.
- This standard approach will not work in time series since the observations are generally dependent.
- But: a joint density can be always factored into a conditional density times a marginal density.

$$f(r_1, r_2) = f(r_1|r_2)f(r_2).$$

Likelihood function for Time Series Models

Example:

- If we have two observations,

$$f(r_2, r_1) = f(r_2|r_1)f(r_1).$$

- If you have three observations,

$$\begin{aligned} f(r_3, r_2, r_1) &= f(r_3|r_2, r_1)f(r_2, r_1) \\ &= f(r_3|r_2, r_1)f(r_2|r_1)f(r_1). \end{aligned}$$

Likelihood function for Time Series Models

In general, we have

$$
\begin{aligned}
& f(r_T, r_{T-1}, \cdots, r_2, r_1) \\
& = f(r_T | r_{T-1}, \cdots, r_1) f(r_{T-1}, \cdots, r_1) \\
& = f(r_T | r_{T-1}, \cdots, r_1) f(r_{T-1} | r_{T-2}, \cdots, r_1) f(r_{T-2}, \cdots, r_1) \\
& = \cdots \\
& = \prod_{t=2}^{T} f(r_t | r_{t-1}, \cdots, r_1) f(r_1),
\end{aligned}
$$

where $\prod_{t=2}^{T}$ denotes product.

Likelihood function for Time Series Models

The corresponding likelihood is

$$
\begin{aligned}
L &= \log f(r_T, r_{T-1}, \cdots, r_2, r_1) \\
&= \sum_{t=2}^{T} \log f(r_t | r_{t-1}, \cdots, r_1) + \log f(r_1)
\end{aligned}
$$

The MLE for MA(1) model

For MA(1) model,

$$r_t = \mu + e_t + \theta_1 e_{t-1}, \quad e_t \sim \text{i.i.d.} N(0, \sigma^2), \quad t = 1, ..., T$$

We set $e_0 = 0$. The parameter set $\Theta = (\mu, \theta_1, \sigma^2)'$. We would like to estimate Θ.

Note that $r_t | e_{t-1} \sim N(\mu + \theta_1 e_{t-1}, \sigma^2)$. Given $e_0 = 0$, we can compute

$$\begin{aligned}
e_1 &= r_1 - \mu - \theta_1 e_0 = r_1 - \mu \\
e_2 &= r_2 - \mu - \theta_1 e_1 = r_2 - \mu - \theta_1 (r_1 - \mu) \\
e_T &= r_T - \mu - \theta_1 e_{T-1} = g_\Theta(r_1, r_2, ..., r_T)
\end{aligned}$$

The MLE for MA(1) model

We have $(r_t|r_{t-1}, ..., r_1, e_0 = 0) \sim N(\mu + \theta_1 e_{t-1}, \sigma^2)$. It implies

$$
\begin{aligned}
& f(r_t|r_{t-1}, ..., r_1, e_0 = 0, \Theta) \\
& \qquad = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(r_t - \mu - \theta_1 e_{t-1})^2}{2\sigma^2}\right)
\end{aligned}
$$

The MLE for MA(1) model

The likelihood function is given by,

$$
\begin{aligned}
L(\Theta) &= f(r_1, ..., r_T | e_0 = 0, \Theta) \\
&= \prod_{t=1}^{T} f(r_t | r_{t-1}, ..., r_1, e_0 = 0, \Theta) \\
&= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(r_t - \mu - \theta_1 e_{t-1})^2}{2\sigma^2} \right)
\end{aligned}
$$

Now we can use

$$
\begin{aligned}
e_0 &= 0 \\
e_1 &= r_1 - \mu \\
e_2 &= r_2 - \mu - \theta_1 (r_1 - \mu)
\end{aligned}
$$

MLE for MA model

There are two approaches for evaluating the likelihood function of an MA model.

- Conditional-likelihood method
  This approach assumes that the initial shocks (i.e., e_t for $t \leq 0$) are zero.
- Exact-likelihood method
  This approach treats the initial shocks e_t, $t \leq 0$, as additional parameters of the model and estimate them jointly with other parameters.

MLE for MA model

- The exact-likelihood estimates are preferred over the conditional ones, especially when the MA model is close to being noninvertible.
- The exact method, however, requires more intensive computation.
- Reference:
  `http://econ.nsysu.edu.tw/ezfiles/124/1124/img/Chapter17_MaximumLikelihoodEstimation.pdf`

Optimal 1-step-ahead forecast of MA(1)

- Assume that the forecast origin is h and Ω_h denote the information available at time h.
- We are interested in forecasting r_{h+l}, where l is the forecast horizon, $l \geq 1$.
- Let $\hat{r}_h(l)$ be the forecast of r_{h+l} using the minimum squared error loss function. In other words, the forecast $\hat{r}_h(l)$ is chosen such that

$$E\{[r_{h+l} - \hat{r}_h(l)]^2 | \Omega_h\} \leq \min_g E[(r_{h+l} - g)^2 | \Omega_h],$$

where $\hat{r}_h(l)$ is the l-step ahead forecast of r_t at the forecast origin h.

Optimal 1-step-ahead forecast of MA(1)

- From the MA(1) process, we have

$$r_{h+1} = \mu + e_{h+1} - \theta e_h$$

- Under the minimum squared error loss function, the point forecast of r_{h+1} given Ω_h is the conditional expectation,

$$\hat{r}_h(1) = E(r_{h+1}|\Omega_h) = \mu - \theta e_h,$$

- The associated forecast error is

$$e_h(1) = r_{h+1} - \hat{r}_h(1) = e_{h+1}$$

Optimal 1-step-ahead forecast of MA(1)

For the 1-step-ahead forecast of an MA(1) process,

$$\hat{r}_h(1) = E(r_{h+1}|\Omega_h) = \mu - \theta e_h,$$
$$e_h(1) = r_{h+1} - \hat{r}_h(1) = e_{h+1}$$

- The optimal forecast is $\mu - \theta e_h$. The optimal forecast error is e_{h+1}.
- The variance of the 1-step-ahead forecast error is $\mathrm{Var}(e_{h+1}) = \sigma^2$.
- However, the error e_h is not directly observable.

How to calculate e_h?

For the 1-step-ahead forecast of an MA(1) process,

$$\hat{r}_h(1) = E(r_{h+1}|\Omega_h) = \mu - \theta e_h,$$

- One approach is to use the autoregressive representation

$$e_h = r_h + \theta r_{h-1} + \theta^2 r_{h-2} + \theta^3 r_{h-3} + \ldots$$

- But this is cumbersome.

How to calculate e_h?

- Another approach is to use the equation

$$e_t = r_t - \mu + \theta e_{t-1}$$

  and realize that this gives a recursive formula to numerically compute the error.

- Given θ, and given the initial condition $e_0 = 0$

$$\begin{aligned}
e_1 &= r_1 - \mu + \theta e_0 \\
e_2 &= r_2 - \mu + \theta e_1 \\
&\dots \\
e_h &= r_h - \mu + \theta e_{h-1}
\end{aligned}$$

- This gives a recursive formula to compute all the errors.

Optimal 2-step-ahead forecast of MA(1)

For the 2-step-ahead forecast of an MA(1) process,

$$r_{h+2} = \mu + e_{h+2} - \theta e_{h+1}$$

We have

$$\begin{aligned} \hat{r}_h(2) &= E(r_{h+2}|\Omega_h) = \mu, \\ e_h(2) &= r_{h+2} - \hat{r}_h(2) = e_{h+2} - \theta e_{h+1} \end{aligned}$$

The variance of the forecast error is $\text{Var}(e_h(2)) = (1 + \theta^2)\sigma^2$, which is greater than or equal to that of the 1-step-ahead forecast error.

Optimal Forecast of MA(2) model

Similarly, the l-step-ahead forecast of an MA(2) model is,

$$r_{h+l} = \mu + e_{h+l} - \theta_1 e_{h+l-1} - \theta_2 e_{h+l-2}$$

from which we obtain

$$\begin{aligned}
\hat{r}_h(1) &= E(r_{h+1}|\Omega_h) = \mu - \theta_1 e_h - \theta_2 e_{h-1}, \\
\hat{r}_h(2) &= E(r_{h+2}|\Omega_h) = \mu - \theta_2 e_h, \\
\hat{r}_h(l) &= E(r_{h+l}|\Omega_h) = \mu, \text{ for } l > 2
\end{aligned}$$

The multistep-ahead forecasts of an MA(2) model go to the mean of the series after two steps.

$MA(\infty)$ Process

Recall the Wold's Theorem, a general linear process can be represented as $MA(\infty)$:

$$\begin{aligned} r_t &= \mu + e_t + \theta_1 e_{t-1} + \ldots \\ &= \mu + \sum_{i=0}^{\infty} \theta_i e_{t-i} \end{aligned}$$

with $\sum_{i=0}^{\infty} |\theta_i| < \infty$.

$MA(\infty)$ Process

The $MA(\infty)$ is given by,

$$r_t = \mu + \sum_{i=0}^{\infty} \theta_i e_{t-i}$$

We have

$$E(r_t) = \mu$$

$$Var(r_t) = \sigma^2 \sum_{i=0}^{\infty} \theta_i^2$$

$$Cov(r_t, r_{t-j}) \leq Var(r_t)$$

with $\theta_0 = 1$.

$MA(\infty)$ Process

For $MA(\infty)$ to be stationary, we need the absolute summability:

$$\sum_{j=0}^{\infty} |\theta_j| < \infty \Longleftrightarrow Var(r_t) < \infty$$

Intuitively, this requires that the effects of the past shocks represented by θ_j eventually die away.

AR(1) model

- A first-order autoregressive, or AR(1) process is,

$$r_t = \phi_0 + \phi_1 r_{t-1} + e_t, \quad e_t \sim WN(0, \sigma^2)$$

- Write in lag operator form

$$\begin{aligned} r_t - \phi_1 r_{t-1} &= \phi_0 + e_t \\ (1 - \phi_1 L) r_t &= \phi_0 + e_t \end{aligned}$$

- If $\phi_1 > 0$, r_t and r_{t-1} are positively correlated;
- If $\phi_1 < 0$, r_t and r_{t-1} are negatively correlated;

Condition mean and variance of AR(1)

The AR(1) model is given by,

$$r_t = \phi_0 + \phi_1 r_{t-1} + e_t, \quad e_t \sim WN(0, \sigma^2)$$

- Condition mean,

$$E(r_t|r_{t-1}) = \phi_0 + \phi_1 r_{t-1}.$$

- Condition variance

$$\begin{aligned} Var(r_t|r_{t-1}) &= E((r_t - E(r_t|r_{t-1}))^2|\Omega_{t-1}) \\ &= E(e_t^2|\Omega_{t-1}) = \sigma^2 \end{aligned}$$

- Given the past return r_{t-1}, the current return is centered around $\phi_0 + \phi_1 r_{t-1}$ with standard deviation σ.

Uncondition mean and variance of AR(1)

Given an AR(1) model,

$$r_t = \phi_0 + \phi_1 r_{t-1} + e_t.$$

The unconditional mean and variance are given by,

$$E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1},$$
$$Var(r_t) = \gamma_0 = \frac{\sigma^2}{1 - \phi_1^2}$$

Proof (Approach 1): $E(r_t) = \mu$

Let $\mu = \frac{\phi_0}{1-\phi_1}$. The AR(1) model can be rewritten as,

$$
\begin{aligned}
r_t &= \phi_0 + \phi_1 r_{t-1} + e_t, \\
r_t - \mu &= \phi_1(r_{t-1} - \mu) + e_t
\end{aligned}
$$

- By repeated substitutions, the AR(1) model can be rewritten as,

$$
\begin{aligned}
r_t - \mu &= e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + .., \\
&= \sum_{j=0}^{\infty} \phi_1^j e_{t-j}
\end{aligned}
$$

Three results

Based on the following result,

$$\begin{aligned} r_t - \mu &= e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + .., \\ &= \sum_{j=0}^{\infty} \phi_1^j e_{t-j} \end{aligned}$$

we have,

$$\begin{aligned} E(r_t e_t) &= E\left[\left(\sum_{j=0}^{\infty} \phi_1^j e_{t-j}\right) e_t\right] \\ &= E[e_t^2] = \sigma^2 \end{aligned}$$

Three results

Based on the following result,

$$
\begin{aligned}
r_t - \mu &= e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + .., \\
&= \sum_{j=0}^{\infty} \phi_1^j e_{t-j}
\end{aligned}
$$

we have,

$$
\begin{aligned}
E(r_{t-1} e_t) &= E\left[\left(\sum_{j=0}^{\infty} \phi_1^j e_{t-1-j}\right) e_t\right] \\
&= 0
\end{aligned}
$$

Three results

Based on the following result,

$$\begin{aligned} r_t - \mu &= e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + .., \\ &= \sum_{j=0}^{\infty} \phi_1^j e_{t-j} \end{aligned}$$

we have,

$$\begin{aligned} E(r_{t+1} e_t) &= E\left[\left(\sum_{j=0}^{\infty} \phi_1^j e_{t+1-j}\right) e_t\right] \\ &= E[\phi_1 e_t^2] = \phi_1 \sigma^2 \end{aligned}$$

Proof (Approach 1): $E(r_t) = \mu$

The AR(1) model can be rewritten as,

$$\begin{aligned} r_t - \mu &= e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + .., \\ &= \sum_{j=0}^{\infty} \phi_1^j e_{t-j} \end{aligned}$$

- If $|\phi_1| < 1$, this is a general linear process with geometrically declining coefficients;
- The impact of a shock becomes smaller and smaller as time passes.
- We can use this result to prove the unconditional mean and variance of r_t.

Proof (Approach 1): $E(r_t) = \mu$

Given an AR(1) model,

$$r_t - \mu = \sum_{j=0}^{\infty} \phi_1^j e_{t-j}$$

The unconditional mean is:

$$E(r_t) = \mu + E\left(\sum_{j=0}^{\infty} \phi_1^j e_{t-j}\right) = \frac{\phi_0}{1 - \phi_1}.$$

Proof (Approach 1): $Var(r_t) = \gamma_0$

Given an AR(1) model,

$$r_t - \mu = \sum_{j=0}^{\infty} \phi_1^j e_{t-j}$$

The unconditional variance is

$$\begin{aligned} Var(r_t) &= Var\left(\mu + \sum_{j=0}^{\infty} \phi_1^j e_{t-j}\right) \\ &= \sum_{j=0}^{\infty} \phi_1^{2j} Var(e_{t-j}) = \sigma^2 \sum_{j=0}^{\infty} \phi_1^{2j} \end{aligned}$$

If $|\phi_1| < 1$, $Var(r_t) = \frac{\sigma^2}{1-\phi_1^2}$.

Proof (Approach 2): $E(r_t) = \mu$

In the second approach, we need to assume that r_t is a weakly stationary process.

Given an AR(1) model,

$$r_t = \phi_0 + \phi_1 r_{t-1} + e_t.$$

Taking the expectation of the prior equation and because $E(e_t) = 0$, we obtain

$$E(r_t) = \phi_0 + \phi_1 E(r_{t-1}).$$

Under the stationarity condition, $E(r_t) = E(r_{t-1}) = \mu$,

$$E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1}$$

Proof (Approach 2): $Var(r_t) = \gamma_0$

Given an AR(1) model,

$$r_t = \phi_0 + \phi_1 r_{t-1} + e_t.$$

Using $\phi_0 = (1 - \phi_1)\mu$, The AR(1) model can be written as,

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + e_t$$

Now square both sides of the prior equation and take expectations:

$$\begin{aligned} E(r_t - \mu)^2 = \phi_1^2 E(r_{t-1} - \mu)^2 \\ + 2\phi_1 E[(r_{t-1} - \mu)e_t] + E(e_t^2) \end{aligned}$$

Proof (Approach 2): $Var(r_t) = \gamma_0$

Recall that $r_{t-1} - \mu$ is a linear function of $e_{t-1}, e_{t-2}, \ldots$,

$$r_{t-1} - \mu = e_{t-1} + \phi_1 e_{t-2} + \phi_1^2 e_{t-3} + \ldots.$$

But e_t is uncorrelated with $e_{t-1}, e_{t-2}, \ldots$, so e_t must be uncorrelated with $r_{t-1} - \mu$. Therefore,

$$E[(r_{t-1} - \mu)e_t] = 0$$

Again, assuming covariance-stationarity, we have

$$E(r_t - \mu)^2 = E(r_{t-1} - \mu)^2 = \gamma_0$$

Proof (Approach 2): $Var(r_t) = \gamma_0$

Recall

$$E(r_t - \mu)^2 = \phi_1^2 E(r_{t-1} - \mu)^2 + 2\phi_1 E[(r_{t-1} - \mu)e_t] + E(e_t^2).$$

Because

$$E(r_t - \mu)^2 = E(r_{t-1} - \mu)^2 = \gamma_0,$$

we have

$$\gamma_0 = \phi_1^2 \gamma_0 + 0 + \sigma^2$$

or

$$\gamma_0 = \sigma^2 / (1 - \phi_1^2)$$

Stationarity of AR(1) model

- AR(1) model

$$r_t = \phi_0 + \phi_1 r_{t-1} + e_t$$

The variance of r_t is given by,

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

The necessary and sufficient condition for the AR(1) model to be weakly stationary is $|\phi_1| < 1$.

Random Walk

The AR(1) model can be rewritten as,

$$r_t - \mu = \sum_{j=0}^{\infty} \phi_1^j e_{t-j}$$

If $\phi_1 = 1$, then the sum does not converge:

$$r_t - \mu = e_t + e_{t-1} + e_{t-2} + \dots$$

Or we could write,

$$r_t = r_{t-1} + e_t$$

Autocovariances of AR(1) models

Given an AR(1) model,

$$r_t = \phi_0 + \phi_1 r_{t-1} + e_t.$$

The autocovariances of AR(1) are,

$$Cov(r_t, r_{t-j}) = \gamma_j = \begin{cases} \phi_1 \gamma_1 + \sigma^2 & \text{if } j = 0, \\ \phi_1 \gamma_{j-1} & \text{if } j > 0. \end{cases}$$

Proof: $\gamma_j = \phi_1 \gamma_{j-1}, j > 0$

Recall

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + e_t$$

We could multiply the prior equation by $r_{t-j} - \mu$ and take expectations:

$$\begin{aligned} E[(r_t - \mu)(r_{t-j} - \mu)] &= \phi_1 E[(r_{t-1} - \mu)(r_{t-j} - \mu)] \\ &\quad + E[e_t(r_{t-j} - \mu)] \end{aligned}$$

Note that

$$\begin{aligned} E[(r_{t-1} - \mu)(r_{t-j} - \mu)] &= E[(r_{t-1} - \mu)(r_{(t-1)-(j-1)} - \mu)] \\ &= \gamma_{j-1}. \end{aligned}$$

Autocovariances of AR(1)

$$E[(r_t - \mu)(r_{t-j} - \mu)] = \phi_1 E[(r_{t-1} - \mu)(r_{t-j} - \mu)] \\ + E[e_t(r_{t-j} - \mu)]$$

Note that

$$E[(r_{t-1} - \mu)(r_{t-j} - \mu)] = \gamma_{j-1}, \\ E[e_t(r_{t-j} - \mu)] = 0$$

Therefore,

$$\gamma_j = \phi_1 \gamma_{j-1}$$

We can show that $\gamma_j = \phi_1^j \gamma_0$.

Autocovariances of AR(1)

Since

$$\gamma_j = \phi_1^j \gamma_0,$$

$$\gamma_0 = \frac{\sigma^2}{1 - \phi_1^2}$$

Therefore,

$$\gamma_1 = \phi_1 \frac{\sigma^2}{1 - \phi_1^2}, \ \gamma_2 = \phi_1^2 \frac{\sigma^2}{1 - \phi_1^2}, \cdots, \gamma_j = \phi_1^j \frac{\sigma^2}{1 - \phi_1^2}$$

Autocorrelations of AR(1)

- Since $\gamma_j = \phi_1 \gamma_{j-1}$, we have,

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \frac{\phi_1 \gamma_{j-1}}{\gamma_0} = \phi_1 \rho_{j-1}$$

- That is,

$$\rho_0 = 1,\ \ \rho_1 = \phi_1,\ \ \rho_2 = \phi_1^2, \cdots, \rho_j = \phi_1^j$$

Autocorrelations of AR(1)

$$\rho_0 = 1,\ \rho_1 = \phi_1,\ \rho_2 = \phi_1^2, \cdots, \rho_j = \phi_1^j$$

- The autocorrelation of a weakly stationary AR(1) series decays exponentially with rate ϕ_1 and starting value $\rho_0 = 1$;
- If ϕ_1 is small, the autocorrelations decay rapidly to zero with k.
- If ϕ_1 is large (close to 1), the autocorrelations decay moderately.
- The AR(1) parameter ϕ_1 describes the persistency in the time series.

Autocorrelations of AR(1)

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \frac{\phi_1 \gamma_{j-1}}{\gamma_0} = \phi_1 \rho_{j-1}$$

- For a positive ϕ_1, the plot of ACF of an AR(1) model shows a nice exponential decay;
- For a negative ϕ_1, the plot consists of two alternating exponential decays with rate ϕ_1^2;

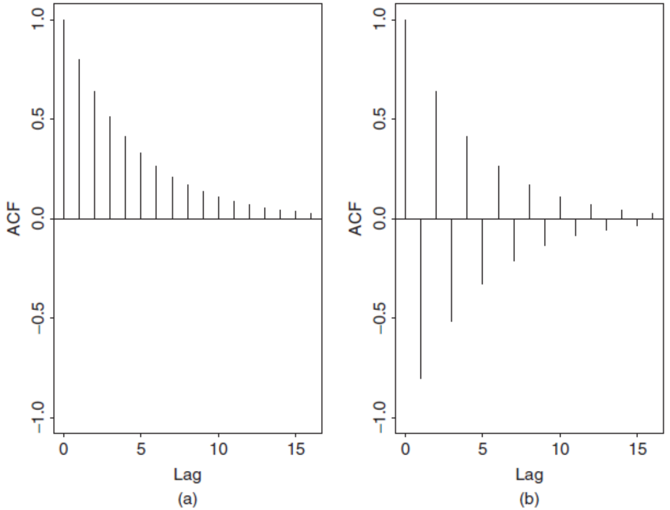


Figure 2.3 Autocorrelation function of an AR(1) model: (a) for $\phi_1 = 0.8$ and (b) for $\phi_1 = -0.8$.

AR(2) Process

- Consider an AR(2) model with intercept

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + e_t, \text{ where } e_t \sim WN(0, \sigma^2)$$

- Using the same techniques as that of the AR(1) case, we obtain

$$E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}$$

provided that $\phi_1 + \phi_2 \neq 1$.

AR(2) Process

- Consider an AR(2) model with intercept

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + e_t, \text{ where } e_t \sim WN(0, \sigma^2)$$

- Using $\phi_0 = (1 - \phi_1 - \phi_2)\mu$, we can rewrite the AR(2) model as,

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + e_t$$

Multiplying the prior equation by $r_{t-j} - \mu$, we have,

$$\begin{aligned}(r_{t-j} - \mu)(r_t - \mu) &= \phi_1(r_{t-j} - \mu)(r_{t-1} - \mu) \\ &= +\phi_2(r_{t-j} - \mu)(r_{t-2} - \mu) + (r_{t-j} - \mu)e_t\end{aligned}$$

AR(2) Process

- Taking expectation and using $E[(r_{t-j} - \mu)e_t] = 0$ for $j > 0$, we obtain

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2}, \text{ for } j > 0$$

- Dividing the above equation by γ_0, we have the property,

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

  for the ACF of r_t.
- The lag-1 ACF satisfies,

$$\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_{-1} = \phi_1 + \phi_2 \rho_1,$$

AR(2) Process

- The ACF of r_t satisfies,

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

- Therefore, for a stationary AR(2) series r_t, we have

$$\begin{aligned} \rho_0 &= 1, \\ \rho_1 &= \frac{\phi_1}{1 - \phi_2}, \\ \rho_j &= \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2}, \ \ j \geq 2 \end{aligned}$$

Stationarity of AR(2) model

- AR(2) model

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + e_t$$

- The AR(2) process is stationary if we can invert the lag polynomial to write it as a general linear process, if

$$\begin{aligned}(1 - \phi_1 L - \phi_2 L^2) r_t &= e_t \\ r_t &= (1 - \phi_1 L - \phi_2 L^2)^{-1} e_t\end{aligned}$$

- When is this valid?

Stationarity of AR(2) model

- We can look at the roots of the polynomial equation (by replacing the lag operator L by a variable, z, and set the resulting polynomial equal to zero:),

$$\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 = 0$$

- Solutions of this equation are

$$z = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}$$

Formula to compute the roots of a quadratic function

The quadratic function is given by,

$$ax^2 + bx + c = 0$$

The roots of x is given by,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Stationarity of AR(2) model

- r_t is stationary if all the roots of $1 - \phi_1 z - \phi_2 z^2 = 0$ "lie outside the unit circle".
- If the roots all are real numbers (that is, none of the roots are complex numbers), then we can say that r_t is stationary if the absolute values of all of these real roots are greater than one.
- If a root equals one or minus one, it is called a unit root.
- If there is at least one unit root, or if any root lies between plus and minus one, then the series is not stationary.

AR(p) model

- The AR(p) is given by,

$$r_t = \phi_0 + \phi_1 r_{t-1} + ... + \phi_p r_{t-p} + e_t, \quad e_t \sim WN(0, \sigma^2)$$

- The mean of a stationary series is,

$$E(r_t) = \frac{\phi_0}{1 - \phi_1 - ... - \phi_p}$$

Autocorrelations of AR(p)

- Autocorrelations of AR(p): Monotonic decreasing or oscillating, the behavior depends on coefficients.
- Both AR(1) and AR(p) can be transformed into $MA(\infty)$ under the stationarity condition.

Stationarity of AR(p) model

Theorem 8

If $r_t \sim AR(p)$, r_t is a stationary process if and only if the modulus of all the roots of the polynomial

$$\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p$$

are greater than one.

The modulus of a real number x is its absolute value $|x|$.

The modulus of a complex number $a + bi$ is $\sqrt{a^2 + b^2}$.

Identifying AR models in practice

Two general approaches are available for determining the value of p.

- Partial autocorrelation function
- Information criterion

Partial autocorrelation function

- Determining the order of an autoregressive process from its autocorrelation function is difficult. To resolve this problem the partial autocorrelation function is introduced.
- If we compare an AR(I) with an AR(2) we see that although in both processes each observation is related to the previous ones, the type of relationship between observations separated by more than one lag is different in both processes:

Partial autocorrelation function

- In the AR(1) the effect of r_{t-2} on r_t is always through r_{t-1}, and given r_{t-1}, the value of r_{t-2} is irrelevant for predicting r_t.

$$r_t = \phi_0 + \phi_1 r_{t-1} + e_t$$

- Nevertheless, in an AR(2) in addition to the effect of r_{t-2} which is transmitted to r_t through r_{t-1}, there exists a direct effect on r_{t-2} on r_t.

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + e_t$$

Partial autocorrelation function

- In general, an AR(p) has direct effects on observations separated by $1, 2, ..., p$ lags and the direct effects of the observations separated by more than p lags are null.

$$r_t = \phi_0 + \phi_1 r_{t-1} + ... + \phi_p r_{t-p} + e_t.$$

Partial autocorrelation function

- The partial autocorrelation coefficient of order j, denoted by $\phi_{j,j}$, is defined as the correlation coefficient between r_t and r_{t-j}, after their mutual linear dependency on the intervening variables $r_{t-1}, r_{t-2}, ..., r_{t-j+1}$ has been removed.
- The conditional correlation

$$Corr(r_t, r_{t-j}|r_{t-1}, r_{t-2}, ..., r_{t-j+1}) = \phi_{jj}$$

is usually referred as the partial autocorrelation in time series. e.g. $\phi_{11} = Corr(r_t, r_{t-1}) = \rho_1$, $\phi_{22} = Corr(r_t, r_{t-2}|r_{t-1})$.

Partial autocorrelation function

- Technically,
  - Regress r_t on $r_{t-1}, r_{t-2}, ..., r_{t-j+1}$ to find the fitted value $\hat{r}_t$ so that we eliminate from r_t, the effect of $r_{t-1}, r_{t-2}, ..., r_{t-j+1}$:

$$r_t = \beta_1 r_{t-1} + ... + \beta_{j-1} r_{t-j+1} + u_t,$$

  - Regress r_{t-j} on $r_{t-1}, r_{t-2}, ..., r_{t-j+1}$ to find the fitted value $\hat{r}_{t-j}$ so that we eliminate the effect of $r_{t-1}, r_{t-2}, ..., r_{t-j+1}$ from r_{t-j}:

$$r_{t-j} = \gamma_1 r_{t-1} + ... + \gamma_{j-1} r_{t-j+1} + v_t,$$

    where, again, v_t contains the part of r_{t-j} not common to the intermediate observations.

Partial autocorrelation function

- Technically,
  - We calculate the conditional correlation

$$
\begin{aligned}
&Corr(r_t, r_{t-j}|r_{t-1}, ..., r_{t-j+1}) \\
&\quad = \frac{Cov[(r_t - \hat{r}_t)(r_{t-j} - \hat{r}_{t-j})]}{\sqrt{Var(r_t - \hat{r}_t)Var(r_{t-j} - \hat{r}_{t-j})}},
\end{aligned}
$$

  which is called the PACF of r_t and r_{t-j}, denoted by $\phi_{j,j}$, where

$$
\begin{aligned}
\hat{r}_t &= E(r_t|r_{t-1}, r_{t-2}, ..., r_{t-j+1}), \\
\hat{r}_{t-j} &= E(r_{t-j}|r_{t-1}, r_{t-2}, ..., r_{t-j+1}).
\end{aligned}
$$

  - $Corr(r_t, r_{t-j}|r_{t-1}, r_{t-2}, ..., r_{t-j+1})$ is also the simple correlation coefficient between u_t and v_t.

Partial autocorrelation function

- This definition is analogous to that of the partial correlation coefficient in regression. It can be proved that the three above steps are equivalent to fitting the multiple regression:

$$r_t = \phi_{0,j} + \phi_{1,j} r_{t-1} + \ldots + \phi_{j,j} r_{t-j} + e_{jt},$$

  and thus $\phi_{j,j}$ is the PACF of r_t and r_{t-j}.

Partial autocorrelation function

- Consider the following AR models in consecutive orders:

$$
\begin{aligned}
r_t &= \phi_{0,1} + \phi_{1,1} r_{t-1} + e_{1t} \\
r_t &= \phi_{0,2} + \phi_{1,2} r_{t-1} + \phi_{2,2} r_{t-2} + e_{2t} \\
r_t &= \phi_{0,3} + \phi_{1,3} r_{t-1} + \phi_{2,3} r_{t-2} + \phi_{3,3} r_{t-3} + e_{3t} \\
r_t &= \phi_{0,4} + \phi_{1,4} r_{t-1} + \phi_{2,4} r_{t-2} + \phi_{3,4} r_{t-3} + \phi_{4,4} r_{t-4} + e_{4t}
\end{aligned}
$$

$\cdots,$

  where $\phi_{0,j}$, $\phi_{i,j}$ and $\{e_{jt}\}$ are respectively, the constant term, the coefficient of r_{t-i}, and the error term of an AR(j) model.

- The estimate $\hat{\phi}_{1,1}$ of the first equation is called the lag-1 sample PACF of r_t.

Partial autocorrelation function

- Consider the following AR models in consecutive orders:

$$
\begin{aligned}
r_t &= \phi_{0,1} + \phi_{1,1} r_{t-1} + e_{1t} \\
r_t &= \phi_{0,2} + \phi_{1,2} r_{t-1} + \phi_{2,2} r_{t-2} + e_{2t} \\
r_t &= \phi_{0,3} + \phi_{1,3} r_{t-1} + \phi_{2,3} r_{t-2} + \phi_{3,3} r_{t-3} + e_{3t} \\
r_t &= \phi_{0,4} + \phi_{1,4} r_{t-1} + \phi_{2,4} r_{t-2} + \phi_{3,4} r_{t-3} + \phi_{4,4} r_{t-4} + e_{4t}
\end{aligned}
$$

$\cdots,$

where $\phi_{0,j}$, $\phi_{i,j}$ and $\{e_{jt}\}$ are respectively, the constant term, the coefficient of r_{t-i}, and the error term of an AR(j) model.

- The estimate $\hat{\phi}_{2,2}$ of the first equation is called the lag-2 sample PACF of r_t.
    - $\hat{\phi}_{2,2}$ shows the added contribution of r_{t-2} to r_t over the AR(1) model $r_t = \phi_0 + \phi_1 r_{t-1} + e_{1t}$.

Partial autocorrelation function

- Consider the following AR models in consecutive orders:

$$
\begin{aligned}
r_t &= \phi_{0,1} + \phi_{1,1} r_{t-1} + e_{1t} \\
r_t &= \phi_{0,2} + \phi_{1,2} r_{t-1} + \phi_{2,2} r_{t-2} + e_{2t} \\
r_t &= \phi_{0,3} + \phi_{1,3} r_{t-1} + \phi_{2,3} r_{t-2} + \phi_{3,3} r_{t-3} + e_{3t} \\
r_t &= \phi_{0,4} + \phi_{1,4} r_{t-1} + \phi_{2,4} r_{t-2} + \phi_{3,4} r_{t-3} + \phi_{4,4} r_{t-4} + e_{4t}
\end{aligned}
$$

$\cdots,$

- The estimate $\hat{\phi}_{3,3}$ of the first equation is called the lag-3 sample PACF of r_t.
    - $\hat{\phi}_{3,3}$ shows the added contribution of r_{t-3} to r_t over the AR(3) model.

Partial autocorrelation function

- This definition is equivalent to fitting the multiple regression:

$$r_t = \phi_{0,j} + \phi_{1,j} r_{t-1} + \ldots + \phi_{j,j} r_{t-j} + e_{jt},$$

and thus $\phi_{j,j}$ is the PACF of r_t and r_{t-j}.

- Consider a zero mean stationary process. Multiply both sides by r_{t-k},

$$\begin{aligned} r_t r_{t-k} = \phi_{1,j} r_{t-1} r_{t-k} + \ldots \\ + \phi_{j,j} r_{t-j} r_{t-k} + e_{jt} r_{t-k}, \end{aligned}$$

Partial autocorrelation function

- Consider a zero mean stationary process. Multiply both sides by r_{t-k},

$$r_t r_{t-k} = \phi_{1,j} r_{t-1} r_{t-k} + \ldots \\ + \phi_{j,j} r_{t-j} r_{t-k} + e_{jt} r_{t-k},$$

- Take the expectations to both sides of the prior equation,

$$\gamma_k = \phi_{1,j}\gamma_{k-1} + \ldots + \phi_{j,j}\gamma_{k-j},$$

- Divide both sides of the prior equation by γ_0,

$$\rho_k = \phi_{1,j}\rho_{k-1} + \ldots + \phi_{j,j}\rho_{k-j},$$

Partial autocorrelation function

- Divide both sides of the prior equation by γ_0,

$$\rho_k = \phi_{1,j}\rho_{k-1} + \ldots + \phi_{j,j}\rho_{k-j},$$

- For $k = 1, 2, \ldots, j$, we have the following system of equations,

$$\begin{aligned}
\rho_1 &= \phi_{1,j} + \phi_{2,j}\rho_1 + \ldots + \phi_{j,j}\rho_{1-j} \\
\rho_2 &= \phi_{1,j}\rho_1 + \phi_{2,j} + \ldots + \phi_{j,j}\rho_{2-j} \\
&\cdots \\
\rho_j &= \phi_{1,j}\rho_{j-1} + \phi_{2,j}\rho_{j-2} + \ldots + \phi_{j,j}
\end{aligned}$$

Partial autocorrelation function

- Using Cramer's rule successively for $j = 1, 2, \ldots$

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}$$

Formula to calculate PACF

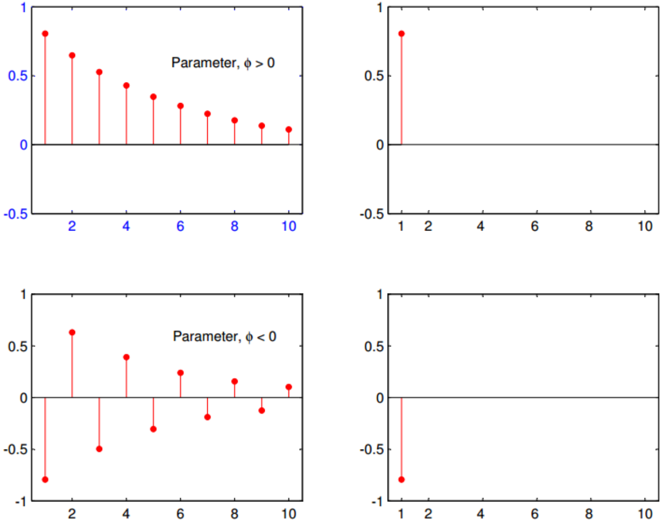
In general,

$$\phi_{j,j} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{j-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{j-3} & \rho_2 \\ \cdots & & & & & \\ \cdots & & & & & \\ \rho_{j-1} & \rho_{j-2} & \rho_{j-3} & \cdots & \rho_1 & \rho_j \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{j-2} & \rho_{j-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{j-3} & \rho_{j-2} \\ \cdots & & & & & \\ \cdots & & & & & \\ \rho_{j-1} & \rho_{j-2} & \rho_{j-3} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

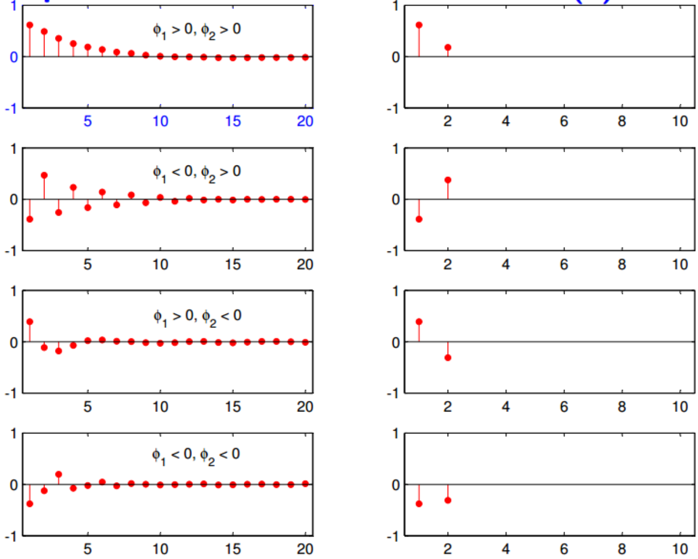
Partial autocorrelation function

- From this definition it is clear that an AR(p) process will have the first p nonzero partial autocorrelation coefficients and, therefore, in the partial autocorrelation function (PACF) the number of nonzero coefficients indicates the order of the AR process.
- This property will be a key element in identifying the order of an autoregressive process.

The partial autocorrelation function - AR(1) models



The partial autocorrelation function - AR(2) models



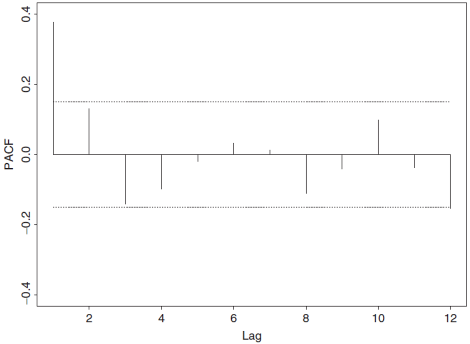


Figure 2.6 Sample partial autocorrelation function of U.S. quarterly real GNP growth rate from 1947.II to 1991.I. Dotted lines give approximate pointwise 95% confidence interval.

This figure shows the PACF of the quarterly growth rate of U.S. real gross national product (GNP), seasonally adjusted, from the second quarter of 1947 to the first quarter of 1991.

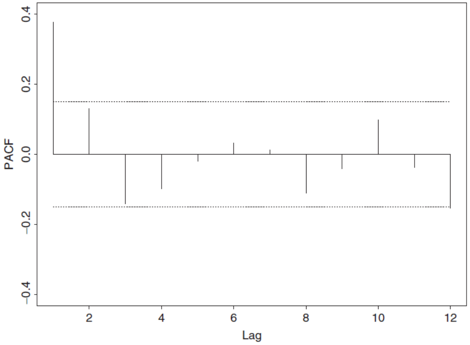


Figure 2.6 Sample partial autocorrelation function of U.S. quarterly real GNP growth rate from 1947.II to 1991.I. Dotted lines give approximate pointwise 95% confidence interval.

- The two dotted lines of the plot denote the approximate two standard error limits.
- The plot suggests an AR(3) model for the data because the first three lags of sample PACF appear to be large.

Information criteria

- Akaike Information Criterion (AIC)

$$AIC = -\frac{2}{T}\ln \hat{L} + \frac{2k}{T},$$

where k is the number of parameters, T is the sample size and $\hat{L}$ is the maximized value of the likelihood function of the model.

- Schwarz-Bayesian Information Criterion(BIC)

$$BIC = -\frac{2}{T}\ln \hat{L} + \frac{\ln(T)k}{T},$$

- The model with the lowest AIC and BIC is preferred.
- The penalty term is larger in BIC than in AIC. Therefore, BIC tends to select a lower AR model when the sample size is moderate or large.

Likelihood function for Gaussian AR(1)

- A stationary Gaussian AR(1) process takes the form

$$r_t = \phi_0 + \phi_1 r_{t-1} + e_t$$

  with $e_t \sim i.i.d.N(0, \sigma^2)$ and $|\phi_1| < 1$. In this case, $\theta = (\phi_0, \phi_1, \sigma^2)'$.
- Consider the pdf of r_1, the first observations in the sample. This is a random variable with mean and variance

$$E(r_1) = \mu = \frac{\phi_0}{1 - \phi_1}$$

$$Var(r_1) = \frac{\sigma^2}{1 - \phi_1^2}$$

Likelihood function for Gaussian AR(1)

- Since r_t is a linear function of $\{e_t\}'$s and $\{e_t\}_{t=-\infty}^{\infty}$ is Normal, r_1 is also Normal, $r_1 \sim N(\phi_0/(1-\phi_1), \sigma^2/(1-\phi_1^2))$
- The density of the first observation takes the form,

$$f_{r_1}(r_1;\theta) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2/(1-\phi_1^2)}} \exp\left[-\frac{[r_1-\phi_0/(1-\phi_1)]^2}{2\sigma^2/(1-\phi_1^2)}\right].$$

Likelihood function for Gaussian AR(1)

- Now let's consider the distribution of the second observation r_2 conditional on the observation r_1.

$$
\begin{aligned}
r_2 &= \phi_0 + \phi_1 r_1 + e_2 \\
f_{r_2|r_1} &\sim N((\phi_0 + \phi_1 r_1), \sigma^2) \\
f_{r_2|r_1} &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(r_2 - \phi_0 - \phi_1 r_1)^2}{2\sigma^2}\right]
\end{aligned}
$$

Similarly,

$$
f_{r_3|r_2,r_1} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(r_3 - \phi_0 - \phi_1 r_2)^2}{2\sigma^2}\right]
$$

Likelihood function for Gaussian AR(1)

In general, the values of $r_1, r_2, ..., r_{t-1}$ matter for r_t only through the value of r_{t-1}. function becomes

$$
\begin{aligned}
f_{r_t|r_{t-1},\cdots,r_1} &= f_{r_t|r_{t-1}} \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(r_t - \phi_0 - \phi_1 r_{t-1})^2}{2\sigma^2}\right].
\end{aligned}
$$

Likelihood function for Gaussian AR(1)

Then the log likelihood can be written as,

$$
\begin{aligned}
L(\theta) &= \log f_{r_1} + \sum_{t=2}^{T} \log f_{r_t|r_{t-1}} \\
&= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2/(1-\phi_1^2)) \\
&\quad - \frac{[y_1 - \phi_0/(1-\phi_1)]^2}{2\sigma^2/(1-\phi_1^2)} - \frac{T-1}{2}\log(2\pi) \\
&\quad - \frac{T-1}{2}\log(\sigma^2) - \sum_{t=2}^{T}\frac{(r_t - \phi_0 - \phi_1 r_{t-1})^2}{2\sigma^2}
\end{aligned}
$$

One-step-ahead forecast of AR(1)

- From the AR(1) model,

$$r_{h+1} = \phi_0 + \phi_1 r_h + e_{h+1},$$

  where h is the forecast origin.

- Under the minimum squared error loss function, the point forecast of r_{h+1} given Ω_h is the conditional expectation,

$$\hat{r}_h(1) = E(r_{h+1}|\Omega_h) = \phi_0 + \phi_1 r_h,$$

  where Ω_h denotes the information available at time h.

One period ahead forecast of AR(1)

- Under the minimum squared error loss function, the point forecast of r_{h+1} given Ω_h is the conditional expectation,

$$\hat{r}_h(1) = E(r_{h+1}|\Omega_h) = \phi_0 + \phi_1 r_h,$$

- The optimal one-step-ahead forecast is a linear function of the final observed value.
- In practice, we compute $\hat{r}_h(1) = \hat{\phi}_0 + \hat{\phi}_1 r_h$ using the estimates.

One period ahead forecast of AR(1)

- 1-step ahead forecast error of AR(1):

$$e_h(1) \equiv r_{h+1} - \hat{r}_h(1) = e_{h+1}$$

  Thus, e_{h+1} is the un-predictable part of r_{h+1}. It is the shock at time $h+1$.

- Variance of 1-step ahead forecast error:

$$Var(e_h(1)) = \sigma^2.$$

Two periods ahead forecast of AR(1)

- By back-substitution

$$
\begin{aligned}
r_{h+2} &= \phi_0 + \phi_1 r_{h+1} + e_{h+2} \\
&= \phi_0 + \phi_1(\phi_0 + \phi_1 r_h + e_{h+1}) + e_{h+2} \\
&= \phi_0 + \phi_0\phi_1 + \phi_1^2 r_h + \phi_1 e_{h+1} + e_{h+2}
\end{aligned}
$$

- Thus

$$
\begin{aligned}
\hat{r}_h(2) &= E(r_{h+2}|\Omega_h) \\
&= E(\phi_0 + \phi_0\phi_1 + \phi_1^2 r_h + \phi_1 e_{h+1} + e_{h+2}|\Omega_h) \\
&= \phi_0 + \phi_0\phi_1 + \phi_1^2 r_h
\end{aligned}
$$

Two periods ahead forecast of AR(1)

- The two periods ahead optimal forecast of AR(1) is

$$\hat{r}_h(2) = E(r_{h+2}|\Omega_h) = \phi_0 + \phi_0\phi_1 + \phi_1^2 r_h.$$

- That is, the 2 periods ahead forecast is also a linear function of the final observed value, but with the coefficients ϕ_1^2.
- In practice, we compute $\hat{r}_h(2) = \hat{\phi}_0 + \hat{\phi}_0\hat{\phi}_1 + \hat{\phi}_1^2 r_h$ using the estimates.

Two periods ahead forecast of AR(1)

- In two-period ahead forecast,

$$\hat{r}_h(2) = \hat{\phi}_0 + \hat{\phi}_0\hat{\phi}_1 + \hat{\phi}_1^2 r_h$$

- 2-step ahead forecast error:

$$e_h(2) = r_{h+2} - \hat{r}_h(2) = \phi_1 e_{h+1} + e_{h+2}.$$

- Variance of 2-step ahead forecast error:

$$Var[e_h(2)] = (\phi_1^2 + 1)\sigma^2.$$

which is greater than or equal to $Var[e_h(1)]$, implying that uncertainty in forecasts increases as the number of steps increases.

k period ahead forecast of AR(1)

- By further back-substitution

$$r_{h+k} = \phi_0 \sum_{j=0}^{k-1} \phi_1^j + \phi_1^k r_h + e_{h+k} + .. + \phi_1^{k-1} e_{h+1}$$

- Thus

$$\hat{r}_h(k) = E(r_{h+k} | \Omega_h) = \phi_0 \sum_{j=0}^{k-1} \phi_1^j + \phi_1^k r_h$$

K periods ahead forecast of AR(1)

- The k periods ahead optimal forecast is

$$\hat{r}_h(k) = E(r_{h+k}|\Omega_h) = \phi_0 \sum_{j=0}^{k-1} \phi_1^j + \phi_1^k r_h.$$

- That is, the *k* periods ahead forecast is also a linear function of the final observed value, but with the coefficients ϕ_1^k.
- In practice, we compute $\hat{r}_h(k) = \hat{\phi}_0 \sum_{j=0}^{k-1} \hat{\phi}_1^j + \hat{\phi}_1^k r_h$ using the estimates.

k period ahead forecast of AR(1)

- k-step ahead forecasting error

$$e_h(k) = r_{h+k} - \hat{r}_h(k) = e_{h+k} + .. + \phi_1^{k-1} e_{h+1}$$

- k-step ahead forecasting variance

$$\mathrm{Var}[e_h(k)] = \sigma^2 \sum_{j=0}^{k-1} \phi_1^{2j}$$

k period ahead forecast of AR(1)

Forecast interval (limit) (FI): General formula for a $100(1-\alpha)\%$ forecast interval for r_{h+k} is given by,

$$\left[\hat{r}_h(k) - N_{\frac{\alpha}{2}}\sqrt{\mathrm{Var}[e_h(k)]},\ \hat{r}_h(k) + N_{\frac{\alpha}{2}}\sqrt{\mathrm{Var}[e_h(k)]}\right],$$

where $N_{\alpha/2}$ denotes the $\alpha/2$-quantile of the standard normal distribution.

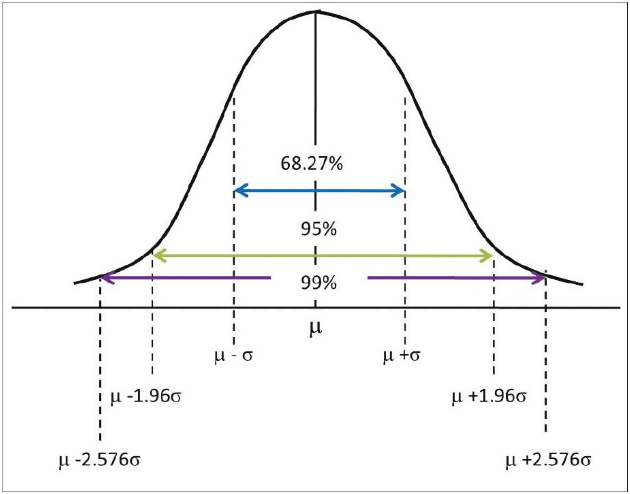
K period ahead forecast of AR(1)

Forecast interval (limit) (FI): The $100(1-\alpha)\%$ forecast interval for k period ahead forecast of AR(1) is given by,

$$\left[\hat{r}_h(k) - N_{\frac{\alpha}{2}}\hat{\sigma}\sqrt{\sum_{j=0}^{k-1}\hat{\phi}_1^{2j}},\ \hat{r}_h(k) + N_{\frac{\alpha}{2}}\hat{\sigma}\sqrt{\sum_{j=0}^{k-1}\hat{\phi}_1^{2j}}\right],$$

where $N_{\alpha/2}$ is the $\alpha/2$-quantile of the standard normal distribution for the selected confidence level α.

Confidence intervals for standard normal



k period ahead forecast of AR(1)

- For a 95% interval ($\alpha = 0.05$):

$$\left[\hat{r}_h(k) - 1.96\hat{\sigma}\sqrt{\sum_{j=0}^{k-1}\hat{\phi}_1^{2j}},\ \hat{r}_h(k) + 1.96\hat{\sigma}\sqrt{\sum_{j=0}^{k-1}\hat{\phi}_1^{2j}}\right]$$

- For a 90% interval ($\alpha = 0.10$):

$$\left[\hat{r}_h(k) - 1.645\hat{\sigma}\sqrt{\sum_{j=0}^{k-1}\hat{\phi}_1^{2j}},\ \hat{r}_h(k) + 1.645\hat{\sigma}\sqrt{\sum_{j=0}^{k-1}\hat{\phi}_1^{2j}}\right]$$

One-step-ahead forecast of AR(p)

- From the AR(p) model, we have,

$$r_{h+1} = \phi_0 + \phi_1 r_h + \ldots + \phi_p r_{h+1-p} + e_{h+1}.$$

- Under the minimum squared error loss function, the point forecast of r_{h+1} given Ω_h is the conditional expectation:

$$\hat{r}_h(1) = E(r_{h+1}|\Omega_h) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i}$$

- The optimal one-step-ahead forecast is a linear function of $\{r_h, r_{h-1}, \ldots, r_{h+1-p}\}$.
- In practice, we compute $\hat{r}_h(1) = \hat{\phi}_0 + \sum_{i=1}^{p} \hat{\phi}_i r_{h+1-i}$ using the estimates.

One period ahead forecast of AR(1)

- The associated forecast error:

$$e_h(1) = r_{h+1} - \hat{r}_h(1) = e_{h+1}$$

  Thus, e_{h+1} is the un-predictable part of r_{h+1}. It is the shock at time $h+1$.

- Variance of 1-step ahead forecast error:

$$Var(e_h(1)) = \sigma^2.$$

Two periods ahead forecast of AR(p)

- From the AR(p) model, we have

$$r_{h+2} = \phi_0 + \phi_1 r_{h+1} + \dots + \phi_p r_{h+2-p} + e_{h+2}.$$

- Taking conditional expectation, we have

$$\begin{aligned}\hat{r}_h(2) &= E(r_{h+2}|\Omega_h) \\ &= \phi_0 + \phi_1 \hat{r}_h(1) + \phi_2 r_h + \dots + \phi_p r_{h+2-p}\end{aligned}$$

- and the associated forecast error

$$\begin{aligned}e_h(2) = r_{h+2} - \hat{r}_h(2) &= \phi_1(r_{h+1} - \hat{r}_h(1)) + e_{h+2} \\ &= e_{h+2} + \phi_1 e_{h+1}\end{aligned}$$

Two periods ahead forecast of AR(p)

- The associated 2-step ahead forecast error

$$e_h(2) = e_{h+2} + \phi_1 e_{h+1}$$

- The variance of the 2-step ahead forecast error is,

$$Var(e_h(2)) = (1 + \phi_1^2)\sigma^2,$$

  which is greater than or equal to $Var[e_h(1)]$, implying that uncertainty in forecasts increases as the number of steps increases.
- This is in agreement with common sense that we are more uncertain about r_{h+2} than r_{h+1} at the time index h for a linear time series.

Multistep-ahead forecast of AR(p)

- In general, we have

$$r_{h+l} = \phi_0 + \phi_1 r_{h+l-1} + \dots + \phi_p r_{h+l-p} + e_{h+l}$$

- The l-step ahead forecast based on the minimum squared error loss function is the conditional expectation of r_{h+l} given Ω_h, which is obtained by

$$\hat{r}_h(l) = \phi_0 + \sum_{i=1}^{p} \hat{r}_h(l-i)$$

Example

Quarterly growth rate of U.S. real gross national product (GNP), seasonally adjusted, from the second quarter of 1947 to the first quarter of 1991.

The model is specified as an AR(3) model,

$$r_t = 0.005 + 0.35 r_{t-1} + 0.18 r_{t-2} - 0.14 r_{t-3} + e_t, \hat{\sigma} = 0.01.$$

where $\{e_t\}$ denotes a white noise with variance σ^2.

Given r_h, r_{h-1} and r_{h-2}, we can predict r_{h+1} as

$$\hat{r}_{h+1} = 0.005 + 0.35 r_h + 0.18 r_{h-1} - 0.14 r_{h-2}.$$

Example

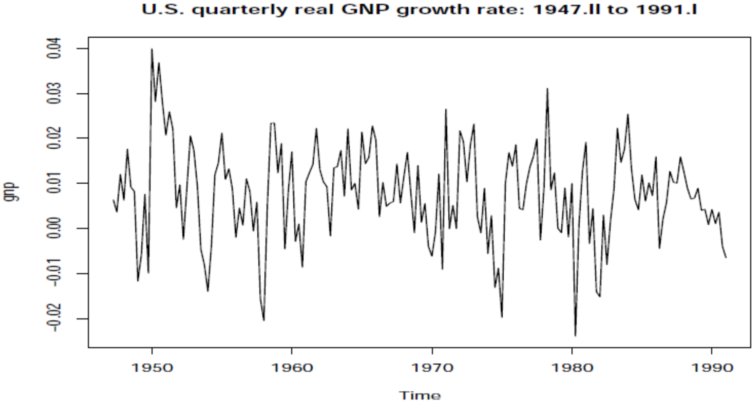


Figure 1: U.S. quarterly growth rate of real GNP: 1947-1991

Simple ARMA models

- An ARMA model combines the ideas of AR and MA models into a compact form so that the number of parameters used is kept small, achieving parsimony in parameterization.
- A time series r_t follows an ARMA(1,1) model if it satisfies

$$r_t - \phi_1 r_{t-1} = \phi_0 + e_t - \theta_1 e_{t-1},$$

where $\{e_t\}$ is a white noise series. The left-hand side of the prior equation is the AR component of the model and the right-hand side gives the MA component.

Mean of ARMA(1,1) model

The ARMA(1,1) model is given by,

$$r_t - \phi_1 r_{t-1} = \phi_0 + e_t - \theta_1 e_{t-1},$$

Taking expectation of the prior equation, we have,

$$E(r_t) - \phi_1 E(r_{t-1}) = \phi_0 + E(e_t) - \theta_1 E(e_{t-1})$$
$$E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1}$$

This result is exactly the same as that of the AR(1) model.

Variance of ARMA(1,1)

Assuming for simplicity that $\phi_0 = 0$.

$$r_t - \phi_1 r_{t-1} = e_t - \theta_1 e_{t-1},$$

Multiplying the model by e_t and taking expectation, we have,

$$E(r_t e_t) = E(e_t^2) - \theta_1 E(e_t e_{t-1}) = E(e_t^2) = \sigma^2$$

Recall that $E[r_t e_{t+1}] = 0$ since r_t is a linear function of e_{t-i} for $i \geq 0$ and e_t is white noise.

Variance of ARMA(1,1)

Rewriting the model as,

$$r_t = \phi_1 r_{t-1} + e_t - \theta_1 e_{t-1},$$

and taking the variance of the prior equation, we have,

$$Var(r_t) = \phi_1^2 Var(r_{t-1}) + \sigma^2 + \theta_1^2 \sigma^2 - 2\phi_1\theta_1 E(r_{t-1}e_{t-1}),$$

where we make use of the fact that r_{t-1} and e_t are uncorrelated. Recall that $E(r_t e_t) = \sigma^2$, we obtain that

$$Var(r_t) - \phi_1^2 Var(r_{t-1}) = (1 + \theta_1^2 - 2\phi_1\theta_1)\sigma^2,$$

Variance of ARMA(1,1)

$$Var(r_t) - \phi_1^2 Var(r_{t-1}) = (1 + \theta_1^2 - 2\phi_1\theta_1)\sigma^2,$$

Therefore, if the series r_t is weakly stationary, then $Var(r_t) = Var(r_{t-1})$ and we have,

$$Var(r_t) = \frac{(1 + \theta_1^2 - 2\phi_1\theta_1)\sigma^2}{1 - \phi_1^2},$$

The stationarity condition is $\phi_1^2 < 1$, (i.e. $|\phi_1| < 1$), which is the same as that of the AR(1) model.

Autocovariance of ARMA(1,1)

Assuming for simplicity that $\phi_0 = 0$. Therefore, $\mu = 0$.

$$r_t - \phi_1 r_{t-1} = e_t - \theta_1 e_{t-1},$$

Multiplying the prior model by r_{t-j} to obtain

$$r_t r_{t-j} - \phi_1 r_{t-1} r_{t-j} = e_t r_{t-j} - \theta_1 e_{t-1} r_{t-j},$$

Autocovariance of ARMA(1,1)

$$r_t r_{t-j} - \phi_1 r_{t-1} r_{t-j} = e_t r_{t-j} - \theta_1 e_{t-1} r_{t-j},$$

- $j = 1$
  Taking expectation and using $E(r_t e_t) = \sigma^2$ for $t-1$, we have

  $$\gamma_1 - \phi_1 \gamma_0 = -\theta_1 \sigma^2,$$

  where $\gamma_j = Cov(r_t, r_{t-j})$. This result is different from that of the AR(1) case for which $\gamma_1 - \phi_1 \gamma_0 = 0$.

Autocovariance of ARMA(1,1)

$$r_t r_{t-j} - \phi_1 r_{t-1} r_{t-j} = e_t r_{t-j} - \theta_1 e_{t-1} r_{t-j},$$

- $j = 2$
  Taking expectation, we have
  $$\gamma_2 - \phi_1 \gamma_1 = 0,$$
  where $\gamma_j = Cov(r_t, r_{t-j})$, which is identical to that of the AR(1) case.
- $j > 1$
  The same technique yields
  $$\gamma_j - \phi_1 \gamma_{j-1} = 0.$$

ACF of ARMA(1,1)

$$\rho_1 = \phi_1 - \frac{\theta_1 \sigma^2}{\gamma_0},$$
$$\rho_j = \phi_1 \rho_{j-1}, \quad \text{for } j > 1$$

- The ACF of an ARMA(1,1) models behaves very much like that of an AR(1) model except for the exponential decay starts with lag 2.
- The ACF of an ARMA(1,1) model does not cut off at any finite lag.

PACF of ARMA(1,1)

- The PACF of an ARMA(1,1) does not cut off at any finite lag either.
- It behaves very much like that of an MA(1) model except that the exponential decay starts with lag 2 instead of lag 1.

General ARMA model

A general ARMA(p,q) model is in the form

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + e_t - \sum_{i=1}^{q} \theta_i e_{t-i},$$

where $\{e_t\}$ is a white noise series and p and q are nonnegative integers. The AR and MA models are special cases of ARMA(p,q) model.

General ARMA model

A general ARMA(p,q) model can be written as

$$\Phi(L)r_t = \phi_0 + \Theta(L)e_t,$$

where $\Phi(L)$ and $\Theta(L)$ are the AR polynomial and MA polynomial of the model, in which

$$\Phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p,$$
$$\Theta(L) = 1 - \theta_1 L - \ldots - \theta_q L^q$$

Stationarity conditions regard the autoregressive part of ARMA(p,q) model. Invertibility conditions regard the moving average part of ARMA(p,q) model.

Stationarity conditions

Theorem 9

Consider the stochastic process defined by

$$\Phi(L)r_t = \phi_0 + \Theta(L)e_t,$$

with $\Phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ *and* $\Theta(L) = 1 - \theta_1 L - ... - \theta_q L^q$. *This process is called stationary if the modulus of all the roots,* z_i, $i = 1, ..., p$, *of the polynomial equation:*

$$\Phi(z) = 0 \Rightarrow 1 - \phi_1 z - ... - \phi_p z^p = 0$$

are greater than one, $|z_i| > 1$ *for all* $i = 1, ..., p$.

Invertibility conditions

Theorem 10

Consider the stochastic process defined by

$$\Phi(L)r_t = \phi_0 + \Theta(L)e_t,$$

with $\Phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ *and* $\Theta(L) = 1 - \theta_1 L - ... - \theta_q L^q$. *This process is called invertible if the modulus of all the roots,* z_i, $i = 1, ..., q$, *of the polynomial equation:*

$$\Theta(z) = 0 \Rightarrow 1 - \theta_1 z - ... - \theta_p z^q = 0$$

are greater than one, $|z_i| > 1$ *for all* $i = 1, ..., q$.

Verify if the following models are stationary, invertible, or both.

$$(1 - L)r_t = (1 - 1.5L)e_t$$

For stationary, the roots of the polynomial equation $1 - z = 0$ should lie outside the unit circle.

$1 - z = 0$ has a unit root and therefore the process is not stationary. The root of $1 - 1.5z = 0$ is $2/3$ which is smaller than one in absolute value, and therefore this process is not invertible either.

Verify if the following models are stationary, invertible, or both.

$$(1 - 0.6L)r_t = (1 - 1.2L + 0.2L^2)e_t$$

For stationary, the roots of the polynomial equation $1 - 0.6z = 0$ should lie outside the unit circle.

The polynomial $1 - 0.6z = 0$ has root $1/0.6 = 5/3$ which is bigger than one in absolute value, and therefore this process is stationary.

Verify if the following models are stationary, invertible, or both.

$$(1 - 0.6L)r_t = (1 - 1.2L + 0.2L^2)e_t$$

For invertibility, the roots of the characteristic equation $1 - 1.2z + 0.2z^2 = 0$ should lie outside the unit circle.

The polynomial factors are

$$1 - 1.2z + 0.2z^2 = \frac{1}{5}(z - 1)(z - 5)$$

and because one of the roots is 1, the process is not invertible

One period ahead forecast of ARMA(p,q)

A general ARMA(p,q) model is in the form,

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + e_t - \sum_{i=1}^{q} \theta_i e_{t-i},$$

Denote the forecast origin by h and the available information by Ω_h. The 1-step-ahead forecast of r_{h+1} can be easily obtained from the model as

$$\hat{r}_h(1) = E(r_{h+1}|\Omega_h) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i} - \sum_{i=1}^{q} \theta_i e_{h+1-i}$$

One period ahead forecast of ARMA(p,q)

The 1-step-ahead forecast of r_{h+1} can be easily obtained from the model as

$$\hat{r}_h(1) = E(r_{h+1}|\Omega_h) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i} - \sum_{i=1}^{q} \theta_i e_{h+1-i}$$

and the associated forecast error is
$e_h(1) = r_{h+1} - \hat{r}_h(1) = e_{h+1}$.

The variance of 1-step-ahead forecast error is
$Var[e_h(1)] = \sigma^2$.

ℓ-step-ahead forecast of ARMA(p,q)

The ℓ-step-ahead forecast of $r_{h+\ell}$ can be easily obtained from the model as

$$\hat{r}_h(\ell) = E(r_{h+\ell}|\Omega_h) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_h(\ell - i) - \sum_{i=1}^{q} \theta_i e_h(\ell - i),$$

where $\hat{r}_h(\ell - i) = r_{h+l-i}$ if $\ell - i \leq 0$ and $e_h(\ell - i) = 0$ if $\ell - i > 0$ and $e_h(\ell - i) = e_{h+\ell-i}$ if $\ell - i \leq 0$.

The ℓ-step-ahead forecast error is $e_h(\ell) = r_{h+\ell} - \hat{r}_h(\ell)$.

Suppose that r_t follows the model

$$r_t = r_{t-1} + e_t - 0.9e_{t-1}$$

and we have $r_{1001} = 1.2$ and $\hat{r}_{1000}(1) = 1.0$, where $r_t(1)$ denotes the 1-step ahead prediction of r_{t+1} at the forecast origin t. Compute $\hat{r}_{1001}(1)$.

Suppose that r_t follows the model

$$r_t = r_{t-1} + e_t - 0.9e_{t-1}$$

Note that $\hat{r}_{1000}(1) = E(r_{1001}|\Omega_{1000}) = r_{1000} - 0.9e_{1000}$.
We have,

$$\begin{aligned} e_{1001} &= r_{1001} - \hat{r}_{1000}(1) = 1.2 - 1 = 0.2 \\ \hat{r}_{1001}(1) &= r_{1001} - 0.9e_{1001} = 1.2 - 0.9 \times 0.2 = 1.02. \end{aligned}$$

Three model representation for an ARMA model

- ARMA(p,q) form:

$$\Phi(L) r_t = \phi_0 + \Theta(L) e_t,$$

  with $\Phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ and $\Theta(L) = 1 - \theta_1 L - ... - \theta_q L^q$.
  - This representation is compact and useful in parameter estimation.
  - It is also useful in computing recursively multistep-ahead forecasts of r_t.

AR representation

- By long division,

$$\frac{\Phi(L)}{\Theta(L)} = 1 - \pi_1 L - \pi_2 L^2 - ... \equiv \pi(L)$$

- For instance, if $\Phi(L) = 1 - \phi_1 L$ and $\Theta(L) = 1 - \theta_1 L$, we have

$$\pi(L) = \frac{1 - \phi_1 L}{1 - \theta_1 L} = 1 - (\phi_1 - \theta_1)L - \theta_1(\phi_1 - \theta_1)L^2 - \theta_1^2(\phi_1 - \theta_1)L^3 - \cdots$$

AR representation

- Dividing $\Theta(L)$ on both sides of the following formula,
$$\Phi(L)r_t = \phi_0 + \Theta(L)e_t,$$
with $\Phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$ and $\Theta(L) = 1 - \sum_{i=1}^{p} \theta_i L^i$.
- The ARMA(p,q) model can be written as an AR representation,
$$r_t = \frac{\phi_0}{1 - \theta_1 - ... - \theta_q} + e_t + \pi_1 r_{t-1} + \pi_2 r_{t-2} + ...,$$
where we use the fact that $Lc = c$, in which c is a constant.

AR representation

- The ARMA(p,q) model can be written as an AR representation,

$$r_t = \frac{\phi_0}{1 - \theta_1 - ... - \theta_q} + e_t + \pi_1 r_{t-1} + \pi_2 r_{t-2} + ...,$$

- The AR representation tells how r_t depends on its past values.

MA representation

- By long division,

$$\frac{\Theta(L)}{\Phi(L)} = 1 + \psi_1 L + \psi_2 L^2 + \ldots \equiv \psi(L)$$

- For instance, if $\Phi(L) = 1 - \phi_1 L$ and $\Theta(L) = 1 - \theta_1 L$, we have

$$\psi(L) = \frac{1 - \theta_1 L}{1 - \phi_1 L} = 1 + (\phi_1 - \theta_1)L + \phi_1(\phi_1 - \theta_1)L^2 + \phi_1^2(\phi_1 - \theta_1)L^3 + \cdots$$

MA representation

- Dividing $\Phi(L)$ on both sides of the following formula,
$$\Phi(L) r_t = \phi_0 + \Theta(L) e_t,$$
with $\Phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$ and $\Theta(L) = 1 - \sum_{i=1}^{p} \theta_i L^i$.
- The ARMA(p,q) model can be written as an MA representation,
$$r_t = \mu + e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \ldots,$$
where $\mu = \phi_0 / (1 - \phi_1 - \ldots - \phi_p)$ and we use the fact that $Lc = c$, in which c is a constant.

MA representation

- The ARMA(p,q) model can be written as an MA representation,

$$r_t = \mu + e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \ldots,$$

- The MA representation shows the impact of the past shock e_{t-i} $(i > 0)$ on the current return r_t.
- For a stationary series, ψ_i converges to zero as $i \to \infty$. Thus, the effect of any shock is transitory.

MA representation

- The MA representation is particularly useful in computing variances of forecast errors.
- At the forecast origin h, the ℓ-step-ahead forecast is,

$$\hat{r}_h(\ell) = \mu + \psi_\ell e_h + \psi_{\ell+1} e_{h-1} + \ldots,$$

- For a ℓ-step ahead forecast, the forecast error is

$$e_h(\ell) = e_{h+\ell} + \psi_1 e_{h+\ell-1} + \ldots + \psi_{\ell-1} e_{h+1},$$

- The variance of forecast error is

$$Var[e_h(\ell)] = (1 + \psi_1^2 + \ldots + \psi_{\ell-1}^2)\sigma^2.$$

Consider the monthly crude oil prices from January 1986 to February 2016. The original data are from FRED and also available in m-COILWTICO.txt.

(a) Obtain the time plot of the oil prices and its first differenced series.

(b) Based on the plots, is the first differenced series weakly stationary? Why?

(c) Let r_t be the first differenced price series. Test $H_0 : \rho_1 = ... = \rho_{12} = 0$ versus $H_a : \rho_i \neq 0$ for some $1 \leq i \leq 12$. Draw your conclusion.

Consider the monthly crude oil prices from January 1986 to February 2016. The original data are from FRED and also available in m-COILWTICO.txt.

(d) Build an AR model for r_t, including model checking. Refine the model by excluding all estimates with t-ratio less than 1.645. Write down the fitted model.

(e) Build an ARMA model for r_t, including model checking. Write down the fitted model.

(f) Use the fitted AR model to compute 1-step to 4-step ahead forecasts of r_t at the forecast origin February, 2016. Also, compute the corresponding 95% interval forecasts.

R Code: CodesToIllustrativeExample.r

(a) Obtain the time plot of the oil prices and its first differenced series.

See **price.pdf** and **diffprice.pdf**.

(b) Based on the plots, is the first differenced series weakly stationary? Why?

No, the variability is much higher in the later part of the series.

(c) Let r_t be the first differenced price series. Test $H_0 : \rho_1 = ... = \rho_{12} = 0$ versus $H_a : \rho_i \neq 0$ for some $1 \leq i \leq 12$. Draw your conclusion.

The Ljung-Box statistic shows Q(12) = 110.11 with p-value close to zero. There are serial correlations in the data.

(d) Build an AR model for r_t, including model checking. Refine the model by excluding all estimates with t-ratio less than 1.645. Write down the fitted model.

The refined model is

$$(1 - 0.38L + 0.18L^6)r_t = a_t, \qquad \sigma^2 = 15.11$$

- AR(6) model is preferred by the AIC criterion
- fit an AR(6) model

The refined model is

$$(1 - 0.38L + 0.18L^6)r_t = a_t, \qquad \sigma^2 = 15.11$$

- t-ratio = Estimate/Std Error. Excluding all estimates with t-ratio less than 1.645 (in absolute value). Note that t-ratios greater than 1.645 or 1.96(in absolute value) suggest that the coefficient is statistically significantly different from 0 at the 90% or 95% confidence level.
- Re-estimate the refined model.
- The Ljung-Box statistics of the residuals can be used to check the adequacy of a fitted model.

(e) Build an ARMA model for r_t, including model checking. Write down the fitted model.

Starting with an AR(1) model, we refined it to the following ARMA model

$$(1 - 0.64L)r_t = (1 - 0.30L - 0.13L^3 - 0.19L^6)e_t, \quad \sigma^2 = 14.95,$$

The model is adequate.

Outline

Random Walk without a drift

- A time series $\{p_t\}$ is a random walk if it satisfies

$$p_t = p_{t-1} + e_t,$$

  where p_0 is a real number denoting the starting value of the process and $\{e_t\}$ is the white noise series.
- The random walk model is a special AR(1) model with $\phi_0 = 0$ and $\phi_1 = 1$. Therefore, it is called a unit-root time series.

Efficient market hypothesis

Efficient market hypothesis states that stock prices in efficient markets follow a random walk process without a drift such that there is no scope for profitable speculation in the stock market: the change in the stock price from one period to the next is essentially random and unpredictable.

Random Walk without a drift

- A random walk $\{p_t\}$ satisfies

$$p_t = p_{t-1} + e_t,$$

- We can write

$$\begin{aligned} p_t &= p_{t-1} + e_t \\ &= p_{t-2} + e_{t-1} + e_t \\ &\cdots \\ &= p_0 + \sum_{i=1}^{t} e_i \end{aligned}$$

Random Walk without a drift

- A random walk $\{p_t\}$ satisfies

$$p_t = p_0 + \sum_{i=1}^{t} e_i$$

- The past never disappears. Shocks have permanent effects.
- A random walk process tends to wander without mean-reversion.

Random Walk without a drift

$$p_t = p_0 + \sum_{i=1}^{t} e_i$$

We have

$$\begin{aligned} E(p_t) &= p_0 \\ Var(p_t) &= Var(\sum_{i=1}^{t} e_i) = t\sigma^2 \\ Cov(p_t, p_s) &= \min\{s, t\}\sigma^2 \end{aligned}$$

Proof: $Cov(p_t, p_s) = \min\{s, t\}\sigma^2$

$$p_t = p_0 + \sum_{i=1}^{t} e_i$$

Without loss of generality, assume $s < t$.

$$\begin{aligned} Cov(p_t, p_s) &= Cov(p_s + e_{s+1} + \dots + e_t, p_s) \\ &= Var(p_s) = \sigma^2 s \end{aligned}$$

It follows that

$$\begin{aligned} Corr(p_t, p_s) &= \frac{s\sigma^2}{\sigma^2\sqrt{s}\sqrt{t}} \\ &= \sqrt{s/t} \end{aligned}$$

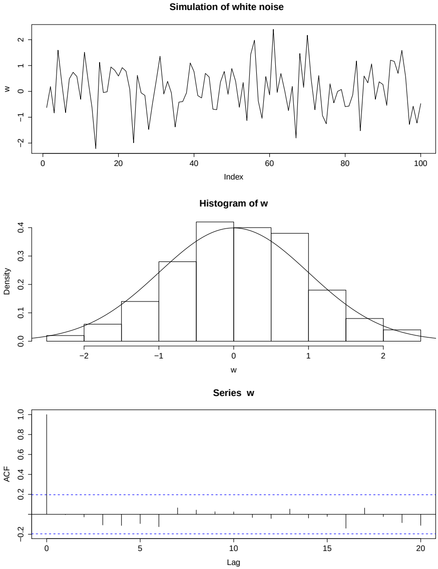
Random Walk without a drift

- Random walk without a drift is a nonstationary process: Although its mean is constant over time, its variance diverges to infinity as t increases.
- A random walk has strong memory! Its sample ACF approaches 1 for any finite lag.
- Repeated substitution shows

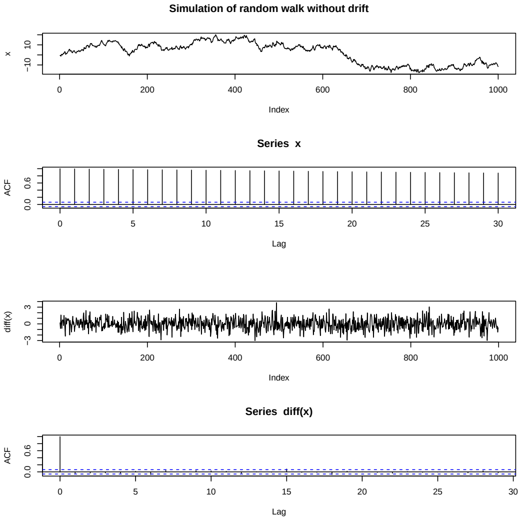
$$p_t = p_0 + \sum_{i=1}^{t} \phi_i e_i,$$

where $\phi_i = 1$ for all i. Thus, ϕ_i does not converge to zero. The effect of any shock is permanent.

Simulation of white noise: `simulate_WN.R`



Simulation of random walk without drift: `simulate_RW.R`



Random Walk with a drift

- In stock price applications it is natural to assume that the log-prices are governed by

$$p_t = \mu + p_{t-1} + e_t, \quad \mu \neq 0$$

where $\mu = E(p_t - p_{t-1})$ and $\{e_t\}$ is the white noise series.

Random Walk with drift

- The random walk with drift model is given by,

$$p_t = \mu + p_{t-1} + e_t,$$

- Note that

$$\begin{aligned} p_t &= \mu + p_{t-1} + e_t \\ &= \mu + \mu + p_{t-2} + e_{t-1} + e_t \\ &\dots \\ &= \mu t + p_0 + \sum_{i=1}^{t} e_i \end{aligned}$$

- μ represents the time trend of the log price p_t and is often referred to as the drift of the model.

Random Walk with drift

$$p_t = \mu t + p_0 + \sum_{i=1}^{t} e_i,$$

We have,

$$E(p_t) = \mu t + p_0$$
$$Var(p_t) = Var(\sum_{i=1}^{t} e_i) = t\sigma^2$$
$$Cov(p_t, p_s) = \min\{s, t\}\sigma^2$$

Interpretation of the constant term

- MA(q) model

$$r_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q},$$

  The constant term μ is the mean of the series $\{r_t\}$.
- AR(p) model

$$r_t = \phi_0 + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + e_t,$$

  The constant term ϕ_0 is related to the mean via $\mu = \phi_0/(1 - \phi_1 - \dots - \phi_p)$.

Interpretation of the constant term

- Random walk with drift

$$p_t = \mu + p_{t-1} + e_t,$$

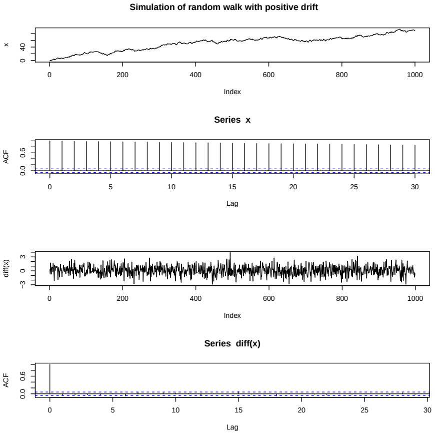
Note that

$$p_t = \mu t + p_0 + \sum_{i=1}^{t} e_i.$$

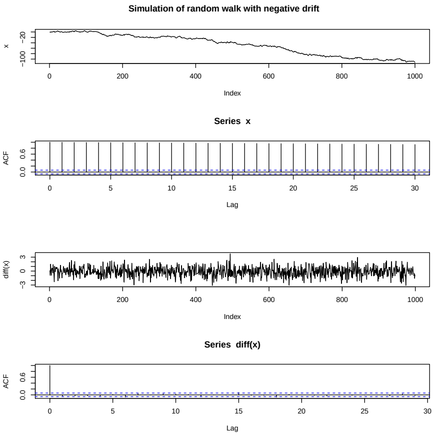
If we graph p_t against the time index t, μ is the time slope .

  - A positive slope μ implies that the log price eventually goes to infinity.
  - A negative slope μ implies that the log price would converge to $-\infty$ as t increases.

R code for random walk with drift: `simulate_RW_wDrift.R`



R code for random walk with drift: `simulate_RW_wDrift.R`



Difference between dynamic and regression models

- AR(1) model

$$r_t = \phi_0 + \phi_1 r_{t-1} + e_t,$$

- Regression model

$$y_t = \beta_0 + \beta_1 x_t + e_t.$$

For the AR(1) model to be meaningful, the coefficient ϕ_1 must satisfy $|\phi_1| \leq 1$.
For the regression models, the coefficient β_1 can assume any fixed real number.

Trend-stationary time series

- A trend-stationary time series model is defined by,

$$p_t = \beta_0 + \beta_1 t + r_t,$$

  where r_t is a stationary time series, for example, a stationary AR(p) series.
- p_t grows linearly in time with rate β_1 and hence can exhibit behavior similar to that of a random-walk model with drift.
- The series with a deterministic trend is non-stationary!

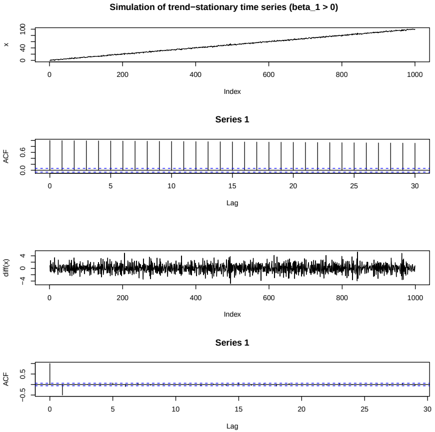
Trend-stationary time series

$$p_t = \beta_0 + \beta_1 t + r_t,$$

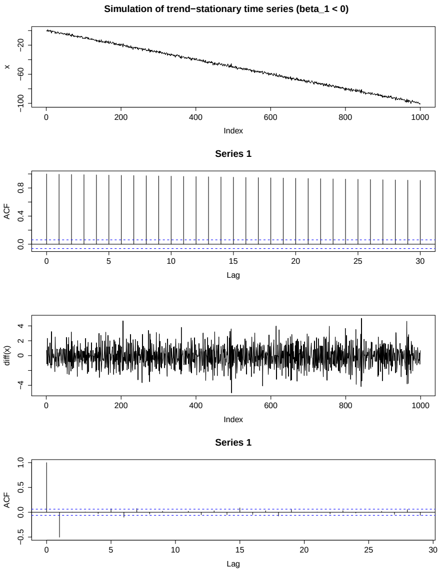
We have,

$$\begin{aligned} E(p_t) &= \beta_0 + \beta_1 t \\ Var(p_t) &= Var(\beta_0 + \beta_1 t + r_t) \\ &= Var(r_t) \end{aligned}$$

Since r_t is a stationary time series, its variance is finite and time invariant.



Simulation of trend−stationary time series (beta_1 > 0)
x
0
40
100
0
200
400
600
800
1000
Index
Series 1
ACF
0.0
0.6
0
5
10
15
20
25
30
Lag
diff(x)
−4
0
4
0
200
400
600
800
1000
Index
Series 1
ACF
−0.5
0.5
0
5
10
15
20
25
30
Lag


Simulation of trend−stationary time series (beta_1 < 0)
x
−100
−60
−20
0
200
400
600
800
1000
Index
Series 1
ACF
0.0
0.4
0.8
0
5
10
15
20
25
30
Lag
diff(x)
−4
0
2
4
0
200
400
600
800
1000
Index
Series 1
ACF
−0.5
0.0
0.5
1.0
0
5
10
15
20
25
30
Lag

Difference between Random walk with drift and Trend-stationary time series

- The random walk with drift assumes both the mean and variance are time dependent.
- The trend-stationary model assumes the mean depends on time and variance is finite and time invariant.

ARIMA model

- If one extends the model by allowing the AR polynomial to have 1 as a characteristic root, then the model becomes the well-known autoregressive integrated moving average (ARIMA) model.
- An ARIMA model is said to be unit-root nonstationary because its AR polynomial has a unit root.
- Like a random-walk model, an ARIMA model has strong memory because the ψ_i coefficients in its MA presentation do not decay over time to zero.

ARIMA model

- A time series p_t is said to be an ARIMA(p,1,q) process if the change series $r_t = p_t - p_{t-1} = (1 - L)p_t$ follows a stationary and invertible ARMA(p,q) model.
  - p is the number of autoregressive terms;
  - d is the number of nonseasonal differences;
  - q is the number of moving-average terms;
- In finance, price series are commonly believed to be nonstationary, but the log return series $r_t = \ln(P_t) - \ln(P_{t-1})$ is stationary. In this case, the log price series is unit-root nonstationary and hence can be treated as an ARIMA process.

ARIMA model

- Let $\{p_t\}$ be a General Stochastic Trend Model:

$$(1-L)^d p_t = r_t, \quad d \geq 1.$$

If r_t is a weakly stationary and invertible ARMA model,

$$\Phi(L) r_t = \Theta(L) e_t,$$

where $\Phi(L)$ and $\Theta(L)$ have no common roots. Then:

$$\Phi(L)(1-L)^d p_t = \Theta(L) e_t,$$

p_t is called ARIMA(p,d,q) model.

ARIMA model

- ARIMA(0,1,0) model:

  $(1-L)p_t = \mu + e_t$ or $p_t = p_{t-1} + \mu + e_t$ (random walk)

- The ARIMA(0,1,1) model

$$(1 - L)p_t = \mu + (1 - \theta L)e_t,$$

  where $|\theta| < 1$.

- ARIMA(1,1,0) model:

$$(1 - \phi L)(1 - L)p_t = \mu + e_t$$

Example 4

- Simulated 100 values from three models:
  - ARIMA(1,1,0) model:

$$(1 - 0.8L)(1 - L)p_t = a_t$$

  - The ARIMA(0,1,1) model

$$(1 - L)p_t = \mu + (1 - 0.75L)e_t,$$

  - ARIMA(1,1,1) model:

$$(1 - 0.9L)(1 - L)p_t = (1 - 0.5L)e_t$$

- Show the sample ACF and PACF.
- Let $r_t = (1 - L)p_t$. Show the sample ACF and PACF of r_t.

Unit-root test

Dickey, D. A.; Fuller, W. A. (1979). ''Distribution of the estimators for autoregressive time series with a unit root". *Journal of the American Statistical Association*. 74 (366a): 427-431.

Distribution of the estimators for autoregressive time series with a unit root **[PDF]** tandfonl

DA Dickey, WA Fuller - Journal of the American statistical …, 1979 - Taylor & Francis

Let n observations Y 1, Y 2,⋯, Y n be generated by the model Y t= pY t− 1+ et, where Y 0 is a fixed constant and {et} t-1 n is a sequence of independent normal random variables with mean 0 and variance σ2. Properties of the regression estimator of p are obtained under the assumption that p=±1. Representations for the limit distributions of the estimator of p and of the regression t test are derived. The estimator of p and the regression t test furnish methods of testing the hypothesis that p= 1.

☆ Save 99 Cite Cited by 35149 Related articles All 11 versions

Unit-root test

- Let p_t be the log price of an asset. To test that p_t is not predictable (i.e. has a unit root), two models are commonly employed:

$$p_t = \phi_1 p_{t-1} + e_t$$
$$p_t = \phi_0 + \phi_1 p_{t-1} + e_t$$

- The hypothesis of interest is $H_0 : \phi_1 = 1$ vs $H_a : \phi_1 < 1$.

Dickey-Fuller test

- Consider the model $p_t = \phi_1 p_{t-1} + e_t$.
- Dickey-Fuller test is the usual t-ratio of the OLS estimate of ϕ_1.
- The hypothesis of interest is $H_0 : \phi_1 = 1$ vs $H_a : \phi_1 < 1$.
- The OLS method gives

$$\hat{\phi}_1 = \frac{\sum_{t=1}^{T} p_{t-1} p_t}{\sum_{t=1}^{T} p_{t-1}^2}, \quad \hat{\sigma}^2 = \frac{\sum_{t=1}^{T} (p_t - \hat{\phi}_1 p_{t-1})^2}{T-1},$$

where $p_0 = 0$ and T is the sample size.

Dickey-Fuller test

- The t ratio is

$$\text{DF} \equiv \text{t ratio} = \frac{\hat{\phi}_1 - 1}{\text{std}(\hat{\phi}_1)} = \frac{\sum_{t=1}^{T} p_{t-1} e_t}{\hat{\sigma}\sqrt{\sum_{t=1}^{T} p_{t-1}^2}}$$

  where $p_0 = 0$ and T is the sample size.
- If $\{e_t\}$ is a white noise series with finite moments of order slightly greater than 2, then the DF statistic converges to a function of the standard Brownian motion as $T \to \infty$.

Dickey-Fuller test

AR(1) model:

- Given the model,

$$p_t = \phi_0 + \phi_1 p_{t-1} + e_t,$$

  we test whether $H_0 : \phi_1 = 1$ vs $H_a : \phi_1 < 1$.

- Equivalently, we consider the model,

$$\Delta p_t = \phi_0 + \delta p_{t-1} + e_t,$$

  we test whether $H_0 : \delta = 0$ vs $H_a : \delta < 0$ ($\delta = \phi_1 - 1$).

The augmented Dickey-Fuller test

Consider the series x_t,

- we can extend

$$x_t = \phi_0 + \phi_1 x_{t-1} + e_t,$$

to an AR(p) process,

$$x_t = c_t + \beta x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + e_t,$$

where c_t is a deterministic function of the time index t, which can be zero or a constant or $c_t = \omega_0 + \omega_1 t$. $\Delta x_t = x_t - x_{t-1}$.

The augmented Dickey-Fuller test

- The augmented DF unit-root test for an AR(p) model is based on

$$x_t = c_t + \beta x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + e_t,$$

- The hypothesis of interest is $H_0 : \beta = 1$ vs $H_a : \beta < 1$.

The augmented Dickey-Fuller test

- The t ratio of $\hat{\beta} - 1$

$$\text{ADF} = \frac{\hat{\beta} - 1}{\text{std}(\hat{\beta})},$$

is the ADF unit-root test statistic, where $\hat{\beta}$ denotes the OLS estimate of β. Again, the statistic has a non-standard limiting distribution.

The augmented Dickey-Fuller test

- Note that the model

$$x_t = c_t + \beta x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + e_t,$$

can be written as,

$$\Delta x_t = c_t + \beta_c x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + e_t,$$

where $\beta_c = \beta - 1$. One can then test the equivalent hypothesis $H_0 : \beta_c = 0$ vs $H_a : \beta_c < 0$.

Outline

Seasonal Time Series

- Before one can specify a model for a given data set, one must have an initial guess about the data generation process.
- The first step is always to plot the time series. In most cases such a plot gives first answers to questions like: "Is the time series under consideration stationary?" or "Do the time series show a seasonal pattern?"

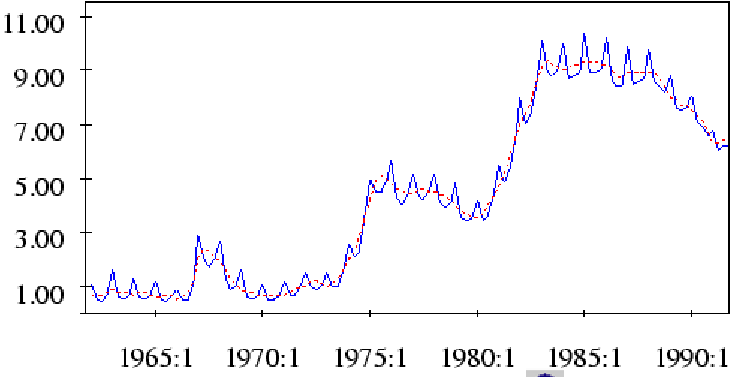


Figure 2: Quarterly unemployment rate for Germany (West) from 1962:1 to 1991:4.

Seasonal Time Series

- Figure 2 displays the quarterly unemployment rate for Germany (West) from the first quarter of 1962 to the fourth quarter of 1991. Denote the quarterly unemployment rate by X_t.
- The solid blue line represents the original series and the dashed line shows the seasonally adjusted series.
- It is easy to see, that this quarterly time series possesses a distinct seasonal pattern with spikes recurring always in the first quarter of the year.
- After the inspection of the plot, one can use the sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) to specify the order of the ARMA part.

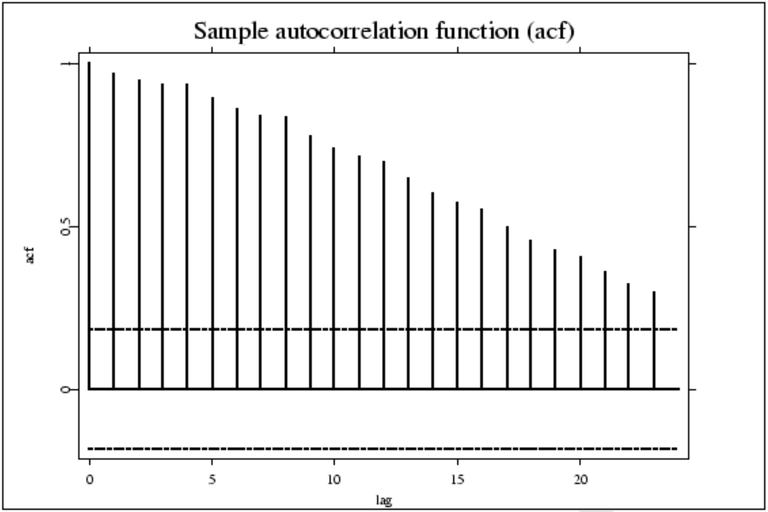


Figure 3: Sample ACF of the unemployment rate for Germany (West) (X_t) from 1962:1 to 1991:4.

Seasonal Differencing

- Figures 3 shows the sample ACF of the original data of the unemployment rate (X_t). The sample ACF of the unemployment rate declines very slowly, i.e. that this time series is clearly nonstationary.
- But it is difficult to isolate any seasonal pattern as all autocorrelations are dominated by the effect of the nonseasonal unit root.

Seasonal Differencing

- In general, for a seasonal time series X_t with periodicity s, seasonal differencing means

$$\Delta_s X_t = X_t - X_{t-s} = (1 - L^s) X_t,$$

  - $s = 4$ for quarterly data,
  - $s = 12$ for monthly data,
  - $s = 24$ for daily effects in hourly data,
  - $s = 168$ for weekly effects in hourly data,
- The conventional difference $\Delta X_t = X_t - X_{t-1} = (1 - L) X_t$, is referred to as the regular differencing.

Differencing

- 1st difference

$$r_t = p_t - p_{t-1},$$

If p_t is the log price, then the 1st difference is simply the log return. Typically, 1st difference means the "change" or "increment" of the original series.

- Seasonal difference

$$y_t = p_t - p_{t-s},$$

where s is the periodicity, e.g. $s = 4$ for quarterly series and $s = 12$ for monthly series. If p_t denotes quarterly earnings, then y_t is the change in earning from the same quarter one year before.

Seasonal Differencing

- We first consider taking the first differences of the unemployment rate ($\Delta X_t = X_t - X_{t-1}$).
- Figures 4 shows the sample ACF of the first differences of the unemployment rate (ΔX_t).

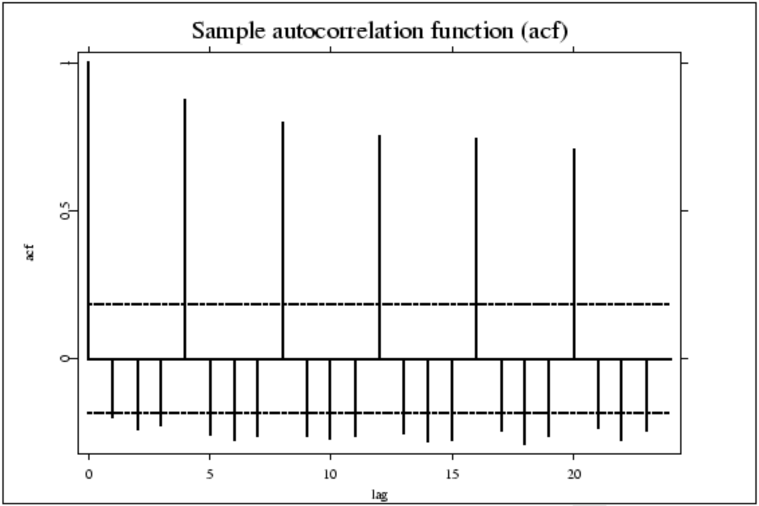


Figure 4: Sample ACF of the first differences of the unemployment rate for Germany (West) (ΔX_t) from 1962:1 to 1991:4.

Seasonal Differencing

- The ACF is strong when the lag is a multiple of periodicity 4. This is a well-documented behavior of sample ACF of a seasonal time series.
- We now consider taking another difference of the data, that is,

$$\begin{aligned}\Delta_4(\Delta X_t) = (1 - L^4)\Delta X_t &= \Delta X_t - \Delta X_{t-4} \\ &= X_t - X_{t-1} - X_{t-4} + X_{t-5},\end{aligned}$$

  where the operator $\Delta_4 = (1 - L^4)$ is called a seasonal differencing.

Seasonal Differencing

- Figures 5 shows the sample ACF of the seasonally differenced first differences of the unemployment rate $(\Delta_4(\Delta x_t))$.
- By means of this transformation we obtain a stationary time series that can be modeled by fitting an appropriate ARMA model.

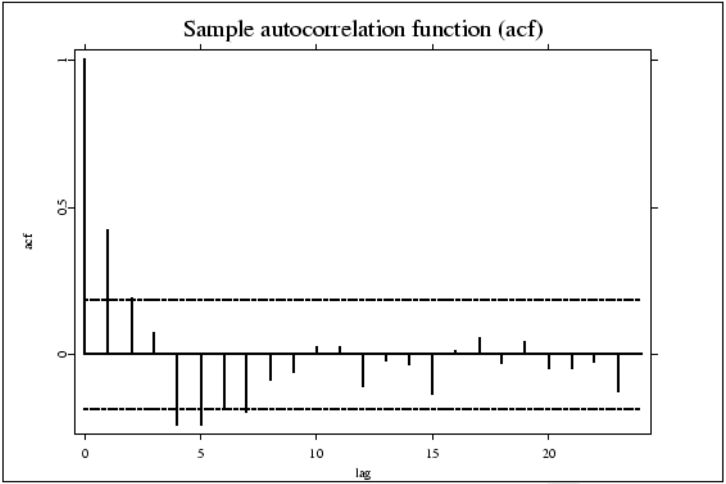


Figure 5: Sample ACF of the seasonally differenced first differences of the unemployment rate for Germany (West) ($\Delta_4(\Delta x_t)$) from 1962:1 to 1991:4.

Seasonal MA model

- Seasonal MA model

  $$r_t = e_t + \Theta e_{t-s},$$

  is a first-order seasonal MA model with $s = 12$, which is equivalent to an MA(12) model with $\theta_i = 0$ for $i = 1, ..., 11$, and $\theta_{12} \neq 0$.

This model is used to capture the dependence between the same months.

Seasonal MA(1) model

- SMA(1) model with periodicity 12:
  $r_t = e_t + \Theta e_{t-12}$.
- The autocovariance of SMA(1) model with periodicity 12,

$$
\begin{aligned}
\gamma(h) &= cov(e_t + \Theta e_{t-12}, e_{t-h} + \Theta e_{t-h-12}) \\
&= E[(e_t + \Theta e_{t-12})(e_{t-h} + \Theta e_{t-h-12})] \\
&= E(e_t e_{t-h}) + \Theta E(e_t e_{t-h-12}) \\
&\qquad + \Theta E(e_{t-12} e_{t-h}) + \Theta^2 E(e_{t-12} e_{t-h-12})
\end{aligned}
$$

Seasonal MA(1) model

- The autocovariance of SMA(1) model with periodicity 12 is

$$\gamma(h) = \begin{cases} (1+\Theta^2)\sigma^2 & \textit{if } h = 0, \\ \Theta\sigma^2 & \textit{if } h = \pm 12, \\ 0 & \textit{otherwise}. \end{cases}$$

- The autocorrelation of SMA(1) model with periodicity 12 is

$$\rho(h) = \begin{cases} 1 & \textit{if } h = 0, \\ \frac{\Theta}{1+\Theta^2} & \textit{if } h = \pm 12, \\ 0 & \textit{otherwise}. \end{cases}$$

Standard AR model

- Standard AR(1) model

$$r_t = \phi r_{t-1} + e_t$$

Based on this model, for all time points the dependence on the past is explained by the previous observation at time $t - 1$.

Seasonal AR model

- Seasonal AR model

$$r_t = \phi r_{t-12} + e_t,$$

  is a first-order seasonal AR model with $s = 12$, which is equivalent to an AR(12) model with $\phi_i = 0$ for $i = 1, ..., 11$, and $\phi_{12} \neq 0$.

This model is used to capture the dependence between the same months.

Seasonal AR(1) model

- SAR(1) model: $r_t = \Phi r_{t-s} + e_t$
- SAR(1) model has the following properties,
  $E[r_t] = 0$
  $\gamma(0) = Var[r_t] = \frac{\sigma^2}{1-\Phi^2}$

$$\rho(h) = \begin{cases} 1 & if\, h = 0 \\ \Phi^{|h|} & if\, h = \pm s \cdot k, k = 1, 2, ... \\ 0 & otherwise. \end{cases}$$

Differ from a standard AR(1) model only with respect to the dependence lag s.

Pure seasonal ARMA

- A pure seasonal ARMA model $ARMA(P,Q)_s$ is defined as

$$\begin{aligned} r_t &= \Phi_1 r_{t-s} + \Phi_2 r_{t-2s} + \ldots + \Phi_P r_{t-Ps} \\ &\quad + e_t + \Theta_1 e_{t-s} + \Theta_2 e_{t-2s} + \ldots + \Theta_Q e_{t-Qs} \\ \Phi_P(L^s) r_t &= \Theta_Q(L^s) e_t, \end{aligned}$$

  where $\Phi_P(L^s) = 1 - \sum_{i=1}^{P} \Phi_i L^{is}$ and $\Theta_Q(L^s) = 1 + \sum_{i=1}^{Q} \Theta_i L^{is}$.
- $\Phi_P(L^s)$ and $\Theta_Q(L^s)$ are seasonal autoregressive and moving average operators.
- This model has no components of standard ARMA models.

Multiplicative seasonal ARMA

- Box-Jenkins multiplicative seasonal ARMA model, denoted by $ARMA_{(p,q)} \times (P, Q)_s$:

$$\Phi_P(L^s)\phi_p(L)r_t \quad = \quad \theta_q(L)\Theta_Q(L^s)e_t,$$

  where $\phi_p(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$ and $\theta_q(L) = 1 + \sum_{i=1}^{q} \theta_i L^i$.
- The non-seasonal parts ϕ_p and θ_q control short-term correlations (up to half a season, lag $\approx s/2$), while the seasonal parts Φ_P and Θ_Q control the decay of the correlations over multiple seasons.
- This model is widely applicable to many seasonal time series.

Multiplicative seasonal Models

- The multiplicative seasonal model is defined as,

$$(1 - L^s)(1 - L)r_t = (1 - \theta L)(1 - \Theta L^s)e_t,$$

  where s is periodicity of the series, $s > 1$, e_t is a white noise series, $|\theta| < 1$, and $|\Theta| < 1$.
- The multiplicative seasonal model is referred to as the airline model.

Multiplicative seasonal model

Focusing on the MA part,

$$\begin{aligned} w_t &= (1-\theta L)(1-\Theta L^s)e_t, \\ &= e_t - \theta e_{t-1} - \Theta e_{t-s} + \theta\Theta e_{t-s-1}, \end{aligned}$$

where

$$\begin{aligned} w_t &= (1-L^s)(1-L)r_t \\ &= (1-L^s)(r_t - r_{t-1}) \\ &= r_t - r_{t-1} - r_{t-s} + r_{t-s-1}. \end{aligned}$$

It is called regular and seasonal differenced series.

Multiplicative seasonal model

It's easy to obtain that $E(w_t) = 0$ and

$$\begin{aligned}
\gamma_0 &= Var(w_t) = (1+\theta^2)(1+\Theta^2)\sigma^2 \\
\gamma_1 &= Cov(w_t, w_{t-1}) = -\theta(1+\Theta^2)\sigma^2 \\
\gamma_{s-1} &= Cov(w_t, w_{t-s+1}) = \theta\Theta\sigma^2 \\
\gamma_s &= Cov(w_t, w_{t-s}) = -\Theta(1+\theta^2)\sigma^2 \\
\gamma_{s+1} &= Cov(w_t, w_{t-s-1}) = \theta\Theta\sigma^2 \\
\gamma_\ell &= Cov(w_t, w_{t-\ell}) = 0, \text{ for } \ell \neq 0, 1, s-1, s, s+1.
\end{aligned}$$

Multiplicative seasonal model

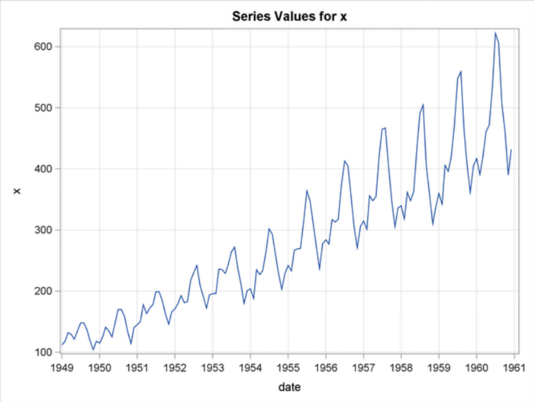
- The ACF of the w_t series is given by,

$$\begin{aligned}
\rho_1 &= \frac{-\theta}{1+\theta^2} \\
\rho_{s-1} &= \frac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)} = \rho_{s+1} \\
\rho_s &= \frac{-\Theta}{1+\Theta^2} \\
\rho_\ell &= 0, \text{ otherwise.}
\end{aligned}$$

- For example, if w_t is a quarterly time series, then $s = 4$ and for $\ell > 0$, the ACF ρ_ℓ is nonzero at lags 1, 3, 4, and 5 only.

International Airline Passengers data in Box and Jenkins (1976).

X_t= the number of Passengers in the t-th month.



- **Step 1.** Make a transformation: $p_t = log(X_t)$.
- **Step 2.** Remove nonstationary or seasonal components (try different models):

$$
\begin{aligned}
w_t &= (1 - B)p_t, \\
w_t &= (1 - B^{12})p_t, \\
w_t &= (1 - B)(1 - B^{12})p_t.
\end{aligned}
$$

- **Step 3.** Note that $w_t = (1 - B)(1 - B^{12})p_t$ is stationary. So we use the seasonal ARMA model to fit the data:

$$\Phi_P(L^{12})\phi_p(L)w_t = \theta_q(L)\Theta_Q(L^{12})e_t,$$

  Now, the problem is how to find p, q, P and Q !!!

- **Step 4.** Look at the ACF and PACF of w_t or try some different p, q, P and Q. For example, we try the model:

$$(1 - \phi L)(1 - \Phi L^{12})w_t = e_t.$$

- **Step 5.** Estimate the parameters in the model. Use MLE method. The results are: $\phi = -0.38$ and $\Phi = -0.52$.
- **Step 6.** Diagnostic checking. Calculate the residuals:

$$e_t = (1 + 0.38L)(1 + 0.52L^{12})w_t.$$

As for ARMA model, if $\{e_t\}$ are white noises, the model is correct.

- **Step 7.** If the model is wrong, we should try other models. Even if it is correct, we still need to try some possible models.
  For example, we try another model:

  $$w_t = (1 - \theta L)(1 - \Theta L^{12})e_t.$$

  Through **Step 5**, we obtain: $\theta = 0.4$ and $\Theta = 0.61$. Through **Step 6**, we know this model is correct, too.

- **Step 8.** Model selection: AIC, BIC, or SBC. The final model is:

$$w_t = (1 - 0.40L)(1 - 0.61L^{12})e_t.$$

or

$$(1-L)(1-L^{12})\log(X_t) = (1-0.40L)(1-0.61L^{12})e_t.$$

- **Step 9.** Forecasting.

Model comparison and averaging

- In applications, there is no true model for a given time series.
- All statistical models are approximations used to describe the dynamic dependence of the data.
- It is common to see that several models fit a given data set well, and the question of model comparison arises.
- Two approaches for the model comparison: in-sample comparison and out-of-sample comparison.

In-sample Comparison

- If the objective of data analysis is to gain insight into the dynamic structure of a time series, then one can use in-sample measurement to compare different models.
- By in-sample, we mean that all data are used in model estimation and comparison.
- In this case, information criteria, such as AIC and BIC, and the estimate of residual variance can be used for model comparison.
- For a selected criterion, the model with a smaller value is preferred.

Out-of-sample Comparison

- When the objective of time series modeling is forecasting, one should use the forecasting performance of the models in model comparison.
- A commonly used measure to quantify the forecasting performance of statistical models is the mean square of forecast errors (MSFE) in an out-of-sample exercise.
- This model comparison method is known as backtesting in the finance literature.

Backtesting procedure

We use 1-step ahead forecasts to introduce the method.

- Step 1: Divide the data set into estimation and forecasting subsamples. There is no specific rule to guide the division, but each subsample should contain sufficient data points so that the estimation and MSFE can be as accurate as possible.

Backtesting procedure

- Step 2: Perform model estimation using data in the estimation subsample and use the fitted model to obtain 1 -step ahead forecast and its forecast error. Specifically, suppose the estimation subsample is $\{r_t \mid t = 1, \ldots, h\}$. We estimate the model using the first h data points to compute the 1 -step ahead prediction $\hat{r}_h(1)$ and its forecast error $e_h(1) = r_{h+1} - \hat{r}_h(1)$. The data point r_{h+1} is not used in model estimation.

Backtesting procedure

- Step 3: Advance the estimation subsample by one data point, that is, $\{r_t \mid t = 1, \ldots, h+1\}$ (recursive window) and $\{r_t \mid t = 2, \ldots, h+1\}$ (rolling window). Reestimate the model using $h+1$ data points and compute the 1 -step ahead forecast and its forecast error. That is, compute $e_{h+1}(1) = r_{h+2} - \hat{r}_{h+1}(1)$ where $\hat{r}_{h+1}(1)$ is the 1 -step ahead prediction of the newly fitted model at the forecast origin $h+1$.
- Step 4: Repeat step 3 until we have the 1 -step ahead forecast error $e_{T-1}(1) = r_T - \hat{r}_{T-1}(1)$, where T is the sample size.

Backtesting procedure

- MSFE of the model is then given by

$$\text{MSFE}(m) = \frac{\sum_{j=h}^{T-1} [e_j(1)]^2}{T - h}$$

where m denotes the model used.

- One selects the model with the smallest MSFE as the best model for the data.
- In practice, one often uses the square root of MSFE instead of MSFE itself.

Backtesting procedure

- Other measurements of forecasting performance include the mean absolute forecast errors and bias, that is,

$$\mathrm{MAFE}(m) = \frac{\sum_{j=h}^{T-1} |e_j(1)|}{T-h}, \quad \mathrm{Bias}(m) = \frac{\sum_{j=h}^{T-1} e_j(1)}{T-h}$$

Backtesting procedure

- Other measurements of forecasting performance include the mean absolute forecast errors and bias, that is,

$$\mathrm{MAFE}(m) = \frac{\sum_{j=h}^{T-1} |e_j(1)|}{T-h}, \quad \mathrm{Bias}(m) = \frac{\sum_{j=h}^{T-1} e_j(1)}{T-h}$$

Backtesting procedure

- Other measurements of forecasting performance include the mean absolute forecast errors and bias, that is,

$$\text{MAFE}(m) = \frac{\sum_{j=h}^{T-1} |e_j(1)|}{T-h}, \quad \text{Bias}(m) = \frac{\sum_{j=h}^{T-1} e_j(1)}{T-h}$$

Model Averaging

- When several models fit a given time series well, instead of selecting a single model, one can use all the models to produce a combined forecast.
- Suppose that there are m models available and they all produce unbiased forecasts for a time series. By unbiased forecast, we mean that the expectation of the associated forecast error is 0.

Model Averaging

- Let $\hat{r}_{i,h+1}$ be the 1 -step ahead forecast of model i at the forecast origin h. Then, a combined forecast is

$$\hat{r}_{h+1} = \sum_{i=1}^{m} w_i \hat{r}_{i,h+1}$$

  where w_i is a nonnegative real number denoting the weight for model i and satisfies $\sum_{i=1}^{m} w_i = 1$.
- The weights w_i can be determined in various ways. For example, the posterior probability of model i in Bayesian inference or the simple average, namely, $w_i = \frac{1}{m}$.