Linear Time Series Analysis
Lecture 2: Stationarity and Properties of ARMA Models

Daniel Buncic

Institute of Mathematics & Statistics
University of St. Gallen
Switzerland

February 5, 2016
Version: [ltsa2]

Homepage
www.danielbuncic.com
University of St. Gallen

Outline/Table of Contents

Outline

Classical Decomposition

Recall that we defined the (additive) Classical Decomposition (Model) as:

$$X_t = m_t + s_t + Y_t,\ t = 1, \ldots, n \quad (1)$$

where

1) m_t was the trend component
2) $s_t = s_{t+d}$ (with $d =$ period) was the seasonal component and
3) Y_t was the cyclical component also sometimes referred to as the "*noise component*" (with $E(Y_t)$ normalised to 0)

Focus will be on **stationary models**: we will assume that the data has been detrended and de-seasonalised.

Will thus focus on modelling Y_t in (1) above.

Transforming Data

Notice from (1) that the relation is **additive**! This does not need to be the case, so model could also be multiplicative as

$$\tilde{X}_t = M_t \times S_t \times \tilde{Y}_t,\ t = 1, \ldots, n \quad (2)$$

but will become additive after **log-transforming** the data.

It is common to work with log-transformed (natural logarithm to base e) data in economics and finance, because:

1) it has a variance stabilising property
2) log changes are percentage changes for small changes (only approximate for large changes if discrete changes are assumed)

3) exponential relationships become linear in a time trend, multiplicative ones become additive

Example

Consider the following general and simple differential equation (also known as the exponential growth model) for the evolution of output over time

$$\frac{d\mathcal{Y}}{dt} = \mathcal{Y}\delta \tag{3}$$

where δ is some rate of growth if $\delta > 0$ (and a rate of decay if $\delta < 0$) and $\mathcal{Y}$ is the level of output at the time. From standard results, we can solve (3) by noting that

$$\frac{d\mathcal{Y}}{\mathcal{Y}} = \delta dt$$

$$\int \frac{d\mathcal{Y}}{\mathcal{Y}} = \int \delta dt$$
$$\ln|\mathcal{Y}| + c_1 = \delta t + c_2$$
$$\mathcal{Y} = e^{\delta t + C}$$
$$\mathcal{Y} = Ae^{\delta t} \tag{4}$$

where $C = c_2 - c_1$ and $A = e^C$. The relation in (4) is the general solution of the simple differential equation in (3). Taking logs of (4) we get

$$\ln(\mathcal{Y}) = \ln(A) + \delta t \tag{5}$$

which is linear in the time index t.

So log transform dampens exponential growth patterns and makes the whole relation linear.

Box-Cox transform

Another more general way to transform data is a power transform known as the Box-Cox transformtion, defined as:

$$y_t = \begin{cases} (x_t^{\lambda} - 1)/\lambda & \text{if } \lambda \neq 0 \\ \log(x_t) & \text{if } \lambda = 0 \end{cases} \quad (6)$$

where λ is a parameter that determines the shape of the transform.

Box-Cox transformtion is frequently employed to stabilise *"ill behaved"* data.

When dealing with time series data a number of things can influence the data adversely, ie.,

- outliers,

- jumps,
- breaks,
- funky seasonal patterns, etc.,

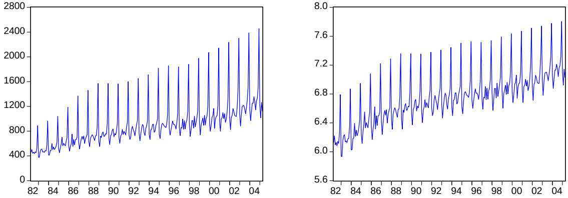
can make it difficult to work with the data.

λ parameter controls how the series is stabilised, so how to choose λ is therefore crucial.

In general, one sets up an optimisation routine for a given loss function and then does a grid search over it.

Implementation details are given in Johnson and Wichern (2007), page 193.

Transformations of Data: Example



(a) Raw series (b) Log-transformed

Figure 1: Retail Sales UK 1982:04 – 2005:04.

Stationarity and convergence of sample moments to population moments

Recall that we defined a stochastic process $\{X_t,\ t \in \mathbb{Z}\}$ to be weakly (or second order) stationary if

1) $\mu_X(t) = E[X_t] = \mu$ is independent of t;
2) $\gamma_X(t+h,t) = \text{Cov}(X_{t+h}, X_t) = E[(X_{t+h} - \mu_X(t+h))(X_t - \mu_X(t))] = \gamma(h)$ is independent of t for each integer h.

Also, recall that we defined the first two sample moments corresponding to the population moments $\mu_X(t)$ and $\gamma_X(t+h,t)$ as

1) Sample mean:

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t \tag{7}$$

2) Sample autocovariance function at lag h:

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x})(x_t - \bar{x}),\ -n < h < n. \tag{8}$$

How do we know that the sample quantitative converge to their population quantities?

That is,

$$\begin{aligned}\operatorname*{plim}_{n\to\infty}\ \bar{x} &= E(X_t)\\ &= \mu_X(t)\\ \operatorname*{plim}_{n\to\infty}\ \hat{\gamma}(h) &= \mathsf{Cov}(X_{t+h}, X_t)\\ &= \gamma_X(t+h, t)\end{aligned}$$

With "*normal*" cross sectional data, we had laws of large numbers (LLNs) that would guarantee the converge of sample moments to population moments, as $n \to \infty$.

- for independently and identically distributed data.
- time series data is, by assumption, not independent over time,
  $\Rightarrow$ **standard LLNs do not hold.**

Ensemble average

Assume that there exists some underlying DGP that creates data.

Use a computer to simulate one (1) artificial data from such a stochastic process:

$$\{x_t^{(1)}\}_{t\in\mathbb{Z}} = \{\ldots, x_{-2}^{(1)}, x_{-1}^{(1)}, x_0^{(1)}, x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}, x_{n+1}^{(1)}, \ldots\}. \tag{9}$$

$\{x_t^{(1)}\}_{t\in\mathbb{Z}}$ in (9) is just one artificial sequence that can be generate from our DGP.

Suppose we generate many (K) of these sequences. Then, $E(x_t)$ is probability limit of "*cross sectional*" series:

$$\mu_t = E(X_t) = \operatorname*{plim}_{K\to\infty} \frac{1}{K}\sum_{i=1}^{K} x_t^{(i)}. \tag{10}$$

(10) is known as the **ensemble average** of the stochastic process X_t at time period t.

Note: other moments of interest (Variance and Autocovariance) are conceptually analogously, ie., ensemble averages are probability limits

$$\begin{aligned}\gamma_t(j) &= E[(X_t - \mu_t)(X_{t-j} - \mu_{t-j})] \\ &= \operatorname*{plim}_{K\to\infty} \frac{1}{K}\sum_{i=1}^{K}(x_t^{(i)} - \mu_t)(x_{t-j}^{(i)} - \mu_{t-j}).\end{aligned} \tag{11}$$

We never have more than one observed series available with empirical data.

Problem

We are computing the expected value of the random variable at time t from a sample of n **time series** observations **and not** from **cross section** or **ensemble average**.

See Table (1) below for illustration.

Stationarity and Ergodicity

Ensemble Averages

Table 1: Example of time series and ensemble averages

$t \diagdown i$	1	2	3	4	$\cdots$	K
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
-1	$x_{-1}^{(1)}$	$x_{-1}^{(2)}$	$x_{-1}^{(3)}$	$x_{-1}^{(4)}$	$\cdots$	$x_{-1}^{(K)}$
0	$x_{0}^{(1)}$	$x_{0}^{(2)}$	$x_{0}^{(3)}$	$x_{0}^{(4)}$	$\cdots$	$x_{0}^{(K)}$
1	$x_{1}^{(1)}$	$x_{1}^{(2)}$	$x_{1}^{(3)}$	$x_{1}^{(4)}$	$\cdots$	$x_{1}^{(K)}$
2	$x_{2}^{(1)}$	$x_{2}^{(2)}$	$x_{2}^{(3)}$	$x_{2}^{(4)}$	$\cdots$	$x_{2}^{(K)}$
3	$x_{3}^{(1)}$	$x_{3}^{(2)}$	$x_{3}^{(3)}$	$x_{3}^{(4)}$	$\cdots$	$x_{3}^{(K)}$
4	$x_{4}^{(1)}$	$x_{4}^{(2)}$	$x_{4}^{(3)}$	$x_{4}^{(4)}$	$\cdots$	$x_{4}^{(K)}$
5	$x_{5}^{(1)}$	$x_{5}^{(2)}$	$x_{5}^{(3)}$	$x_{5}^{(4)}$	$\cdots$	$x_{5}^{(K)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
n	$x_{n}^{(1)}$	$x_{n}^{(2)}$	$x_{n}^{(3)}$	$x_{n}^{(4)}$	$\cdots$	$x_{n}^{(K)}$
$n+1$	$x_{n+1}^{(1)}$	$x_{n+1}^{(2)}$	$x_{n+1}^{(3)}$	$x_{n+1}^{(4)}$	$\cdots$	$x_{n+1}^{(K)}$
$n+2$	$x_{n+2}^{(1)}$	$x_{n+2}^{(2)}$	$x_{n+2}^{(3)}$	$x_{n+2}^{(4)}$	$\cdots$	$x_{n+2}^{(K)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$

Stationarity and Ergodicity

Ensemble Averages

Information content is mismatched!

- we are interested in an ensemble average
- but we only have one time series of data available to compute the average at each point in time.

For the **time series** average $\bar{x}$ to converge to the **ensemble concept** of $E(X_t) = \mu_t$ we need to put some restrictions on the memory of the stochastic process.

For a **time series** average to converge to the **ensemble concept** of $E(X_t)$, the stochastic process needs to be **ergodic**.

For the formal definition of **ergodicity**, we need the stochastic process to be weakly **stationary**, which we have already defined.

Definition (Ergodicity)

Let $f(\cdot)$ and $g(\cdot)$ be two bounded and real valued functions. A stationary process $\{X_t\}_{t\in\mathbb{Z}}$ is said to be ergodic if, for any $k, \ell \in \mathbb{Z}$ and $j \in \mathbb{N}_+$

$$\lim_{j\to\infty} |E\left[f(X_t,\ldots,X_{t+k}) \times g(X_{t+k+j},\ldots X_{t+k+j+\ell}\right]|$$
$$= |E\left[f(X_t,\ldots,X_{t+k})\right] \times E[g(X_{t+k+j},\ldots X_{t+j+k+\ell}]| \,.$$

The definition of ergodicity above states the following:

- if we take two stochastic process at two different time periods positioned "**far apart**" from one another, then these two processes are **almost** independently distributed from one another.

Notice here the difference between stationarity and ergodicity.

- Ergodicity focuses on the asymptotic independence of the process.
- Stationarity, is concerned with the time invariance of the moments of the process.

To be ergodic, the memory of a stochastic process should fade the further the two blocks are taken away from each other

- the dependence between increasingly distant observations disappears "**sufficiently rapidly**".

Theorem (LLN for covariance stationary processes)

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a covariance stationary process with finite mean ($\mu < \infty$) given by:

$$E(X_t) = \mu$$
$$E(X_t - \mu)(X_{t-j} - \mu) = \gamma(j) \tag{12}$$

$\forall\ t \in \mathbb{Z}$. If

$$\sum_{j=0}^{\infty} |\gamma(j)| < \infty \tag{13}$$

then,

$$\bar{x} = n^{-1} \sum_{t=1}^{n} x_t \xrightarrow{p} E(X_t) = \mu$$

and x_t is said to be ergodic for the mean.

The above LLN is important as it provides us with the conditions under which we can validly use time series averages to obtain information about the ensemble concept of the population moment

Transformations of stationary processes are stationary and ergodic

Theorem (Transformations of stationary processes)

If $\{X_t\}_{t\in\mathbb{Z}}$ is a covariance stationary and ergodic process and $g(\cdot)$ a (real-valued measurable) function then $\{Y_t\}_{t\in\mathbb{Z}}$ formed from

$$Y_t = g(\ldots, X_{t-2}, X_{t-1}, X_t, X_{t+1}, X_{t+2}, \ldots) \tag{14}$$

is also an **ergodic and covariance stationary process**.

Theorem (Ergodic theorem for stationary processes)

If $\{X_t\}_{t\in\mathbb{Z}}$ is a covariance stationary and ergodic process, $g(\cdot)$ a (real-valued measurable) function and $E[g(X_t)] < \infty$ (moment exists), then:

$$\bar{g}_n = n^{-1}\sum_{t=1}^{n} g(x_t) \xrightarrow{p} E[g(X_t)].$$

The **Transformations of stationary processes** result establishes that we can transform any covariance stationary and ergodic series by a (well-behaved) function and get again an ergodic and stationary series back in return.

The **Ergodic theorem for stationary processes** result says that the sample mean of any (well-behaved) functional transformation $g(\cdot)$ of a covariance stationary and

ergodic series converges to the corresponding population expectation, as long as the population expectation exists.

Summary of results on Stationarity and Ergodicity

Results above tell us that we can validly estimate **any** population moments from the **time series** average (and not the cross-section or **ensemble average**)

1) as long as the process $\{X_t\}_{t\in\mathbb{Z}}$ is covariance stationary and ergodic and
2) the population moments exist.

Background

Autoreg**R**essive (AR) and **M**oving**A**verage (MA) models, when put together, ARMA models

- the fundamental building block of time series analysis and
- have a long history in this literature.

Origins are from Yule (1927)

- first to consider the idea that a time series for which successive values are highly correlated can be generated from a series of uncorrelated and exogenous **White Noise** *"shocks"* Z_t
- White Noise process is transformed into an observed process X_t by a transformation which is called a **linear filter**

- the linear filter simply takes a weighted sum of the infinite history of White Noise *"shocks"* Z_t. The series is transformed as follows

$$\begin{aligned} X_t &= \mu + Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \psi_3 Z_{t-3} + \cdots \\ &= \mu + (1 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \cdots) Z_t \\ &= \mu + \psi(L) Z_t \end{aligned} \tag{15}$$

where μ is the unconditional mean or level of the series and $\psi(L)$ is the **lag polynomial**

- linear filter $\psi(L)$ is called a **transfer function**

- whether the generated series is stationary or not depends on the **stability** of the transfer function.
  - the filter is said to be stable if the ψ_j weights are absolutely summable, that is, if

$$\sum_{j=0}^{\infty} |\psi_j| < \infty \quad (16)$$

holds, with $\psi_0 = 1$. Under this condition, the x_t process will also be stationary.

First formalisation of Yule's idea of a **linear filter** was given by Wold (1938) and is known Wold's Representation Theorem

- also as the Wold Decomposition
- or just Wold's Theorem

Wold Decomposition Theorem

Theorem (Wold Decomposition)

If $\{X_t\}$ is a stationary time series, then it can be represented as:

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t, \quad (17)$$

where

- $\psi_0 = 1$ *and* $\sum_{j=0}^{\infty} \psi_j^2 < \infty$*;*
- $\{Z_t\} \sim \text{WN}(0, \sigma^2)$*;*
- $\text{Cov}(Z_s, V_t) = 0$ *for all* t*;*
- $\{V_t\}$ *is deterministic (perfectly forecastable)*

Note from above that we only **required square summability**, ie.,

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty \tag{18}$$

for Wold's theorem to hold, which is a **weaker condition** than the one of **absolute summability** assumed in Yule's formulation of a linear filter, that is,

$$\sum_{j=0}^{\infty} |\psi_j| < \infty. \tag{19}$$

Box-Jenkins Approach

The Wold representation is not feasible for modelling a realised process x_t

- only have finite number of observations
- the key problem in ARMA modelling is how best to approximate the infinite order term $\psi(L)$ in

$$X_t = \mu + \psi(L)Z_t,\ \{Z_t\} \sim \text{WN}(0, \sigma^2),\ \psi_0 = 1,\ \text{and} \sum \psi_i^2 < \infty. \tag{20}$$

The key idea is to approximate the infinite order polynomial $\psi(L)$ with a ratio of two finite polynomial

$$\psi(L) = \frac{\theta_q(L)}{\phi_p(L)}, \tag{21}$$

with p, q finite.

Using (21) we get

$$X_t = \mu + \frac{\theta(L)}{\phi(L)} Z_t$$

or equivalently

$$\begin{aligned} \phi(L)X_t &= \phi(L)\mu + \theta(L)Z_t \\ &= \phi(1)\mu + \theta(L)Z_t \\ (L)X_t &= c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} \\ &\quad + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \ldots + \theta_q Z_{t-q}. \end{aligned} \tag{22}$$

where $c = \phi(1)\mu$.

In **Box-Jenkins approach**, need to fined appropriate AR(p) and MA(q) components to model the dynamic behaviour of X_t in a **parsimonious** way.

3 basic steps in the Box-Jenkins modelling approach (need to assure that X_t is **stationary** and **free from seasonal variation**)

1) Model Selection (also called Identification)
2) Model Estimation (estimation of parameters)
3) Model Checking and Evaluation (forecasting)

- **Model selection part focuses** on finding the right orders p and q of the AR and MA components
  - to find orders, split the ARMA(p, q) model in its two component blocks
  - look at the properties of these two blocks individually, before we put the model back together into its ARMA form.

- look at sample autocorrelation functions (ACF) and partial autocorrelation functions (PACF).
- later on when modelling, will use **information criteria** to help find the most parsimonious model.

- Model **estimation** and **forecasting** with ARMA models will be outlined in the next Lecture Chapter.

ARMA Processes

Definition (ARMA Processes)

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

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = c + Z_t + \theta_1 Z_{t-1} + \ldots \theta_q Z_{t-q}, \tag{23}$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$ and the polynomials

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \tag{24}$$

and

$$\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q \tag{25}$$

have **no common roots (factors)**.

Some Results

The process $\{X_t\}$ is said to be an ARMA(p,q) process with mean μ if $\{X_t - \mu\}$ is an ARMA(p,q) process.

The time series $\{X_t\}$ is said to be an autoregressive process of order p (or AR(p)) if $\theta(z) \equiv 1$, and a moving-average process of order q (or MA(q)) if $\phi(z) \equiv 1$.

Existence and Uniqueness

A stationary solution $\{X_t\}$ of the equations for an ARMA(p,q) model exists (and is also the unique stationary solution) if and only if

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \neq 0 \quad \text{for all } |z| = 1. \tag{26}$$

Invertibility of Lag Polynomial

The lag polynomial $\theta(L)$ is invertible if

$$\theta(z) = 1 - \theta_1 z - \ldots \theta_p z^p \neq 0 \quad \text{for all } |z| \leq 1. \tag{27}$$

Rationale for ARMA models

Why is the seemingly atheoretical looking ARMA model so popular among practitioners and also as a tool for economic policy modelling?

- ARMA models are notoriously difficult to beat out-of-sample.
  - low number of parameters in ARMA models compared to bigger scale and more complicated models.
  - less parameters results in less forecast uncertainty due to estimation uncertainty.
- ARMA models as special versions of **reduced form economic models**
  - coming from a structural econom(etr)ic model.

To see this, consider the simplest macroeconomic model.

Take Consumption C_t, Output Y_t and Investment I_t in the following bivariate system

$$C_t = a_0 + a_1 Y_t + a_2 C_{t-1} + \varepsilon_t \quad (28)$$

$$Y_t \equiv C_t + I_t \quad (29)$$

where ε_t is some uncorrelated shock, with mean 0 and variance 1.

C_t and Y_t are endogenous, I_t and ε_t are exogenous. Substituting Y_t into C_t and re-arranging yields the **reduced form models**:

$$C_t - \delta_2 C_{t-1} = \delta_0 + \delta_1 I_t + \delta \varepsilon_t \quad (30)$$

$$Y_t - \delta_2 Y_{t-1} = \delta_0 - \delta_2 I_{t-1} + \sigma(\varepsilon_t + I_t) \quad (31)$$

where

$$\delta_2 = \frac{a_2}{1-a_1}, \delta_0 = \frac{a_0}{1-a_1} \quad (32)$$

and

$$\delta_1 = \frac{a_1}{1-a_1}, \delta = \frac{1}{1-a_1}. \quad (33)$$

If I_t White Noise and uncorrelated with ε_t with mean 0 and variance 1,

- then C_t is an AR(1) process.
- Y_t can be shown to have the properties of an ARMA(1,1) process

ACF and PACF

Definition (ACF)

The Autocorrelation function is defined as

$$\rho(j) = \frac{\mathrm{Cov}(X_t, X_{t-j})}{\mathrm{Var}(X_t)} = \frac{\gamma(j)}{\gamma(0)} \tag{34}$$

Definition (PACF)

The Partial Autocorrelation function is defined as

$$\phi_{jj} = \mathrm{Corr}(X_t, X_{t-j} | X_{t-1}, X_{t-2}, ..., X_{t-j+1}) \tag{35}$$

The PACF can be viewed as the last coefficient in a population linear regression of X_t on a constant and j lagged values of X_t.

Remark: The PACF measures the correlation between X_t and X_{t-j} by **controlling for** or **conditioning on** all lagged values of X_t up to $t-j+1$. So it gives the **net effect** (or influence) of X_{t-j} on X_t.

We can always work out the PACF after we have computed the ACFs

- PACFs are a function of the ACFs.
- there exist a few different ways on how to quickly work out the ACFs of a given AR, MA or mixed ARMA model.

For AR models, easiest to work out ACFs by writing model in de-meaned form $\tilde{X}_t = (X_t - \mu)$

- multiply AR process by $\tilde{X}_{t-j}$ on both sides and
- then take expectations.

For MA models, can follow a similar strategy.

- will need to evaluate $E(X_t Z_{t-j})$ though
- can use a recursive form for this (this will become clear later)

We will learn later how to **move between** AR, MA and ARMA models in general using the lag polynomial operator

Proposition (Mean)

Let X_t be a covariance stationary AR(p) process defined as

$$\phi(L) X_t = c + Z_t \tag{36}$$

where $Z_t \sim \text{WN}(0, \sigma^2)$.

Then its (unconditional) mean μ has the form

$$E(X_t) = \left(1 - \sum_{i=1}^{p} \phi_i\right)^{-1} c. \tag{37}$$

Proof.

Invert the AR(p) to its MA(∞) representation. This yields:

$$\begin{aligned} X_t &= \phi(L)^{-1} c + \phi(L)^{-1} Z_t \\ &= \phi(1)^{-1} c + \psi(L) Z_t \\ &= \left(1 - \textstyle\sum_{i=1}^{p} \phi_i\right)^{-1} c + \sum_{j=0}^{\infty} \psi_j Z_{t-j} \end{aligned} \quad (38)$$

Taking expectations of (38) yields

$$\begin{aligned} E(X_t) &= \left(1 - \textstyle\sum_{i=1}^{p} \phi_i\right)^{-1} c + \sum_{j=0}^{\infty} \psi_j \underbrace{E(Z_{t-j})}_{=0} \\ &= \left(1 - \textstyle\sum_{i=1}^{p} \phi_i\right)^{-1} c \end{aligned}$$

due to $\sum_{j=0}^{\infty} \psi_j E(Z_{t-j}) = 0$, $\forall j = 1, 2, \ldots$ □

Proposition (Autocovariance)

The AR(p) process defined in (36) has the following autocovariance recursions

$$\gamma(j) = \begin{cases} \phi_1 \gamma(1) + \phi_2 \gamma(2) + \ldots + \phi_p \gamma(p) + \sigma^2 & \text{for } j = 0 \quad (39a) \\ \phi_1 \gamma(j-1) + \phi_2 \gamma(j-2) + \ldots + \phi_p \gamma(j-p) & \text{for } j > 0 \quad (39b) \end{cases}$$

Proof.

Write (36) in demeaned $\left(\tilde{X}_t = X_t - \mu\right)$ form and expand the $\phi(L)$ term to get

$$\begin{aligned} \phi(L)\tilde{X}_t &= Z_t \\ \tilde{X}_t &= \phi_1\tilde{X}_{t-1} + \phi_2\tilde{X}_{t-2} + \ldots + \phi_p\tilde{X}_{t-p} + Z_t. \end{aligned} \tag{40}$$

Then multiply both sides of (40) by $\tilde{X}_{t-j}$ and take expectations, which gives

$$\begin{aligned} E\left[\tilde{X}_t\tilde{X}_{t-j}\right] &= \phi_1 E\left[\tilde{X}_{t-1}\tilde{X}_{t-j}\right] + \phi_2 E\left[\tilde{X}_{t-2}\tilde{X}_{t-j}\right] + \ldots \\ &+ \phi_p E\left[\tilde{X}_{t-p}\tilde{X}_{t-j}\right] + E\left[Z_t\tilde{X}_{t-j}\right] \end{aligned}$$

$$\begin{aligned} \gamma(j) &= \phi_1\gamma(j-1) + \phi_2\gamma(j-2) + \ldots \\ &+ \phi_p\gamma(j-p) + E\left[Z_t\tilde{X}_{t-j}\right] \end{aligned}$$

where $E\left[Z_t\tilde{X}_{t-j}\right] = \begin{cases} \sigma^2 & \text{for } j = 0 \\ 0 & \text{for } j > 0 \end{cases}$. □

The first $(p+1)$ equations in the recursive system in (39) can be put into the following matrix form

$$\gamma = \mathbf{A}\gamma + \mathbf{\Sigma} \tag{41}$$

$$(\mathbf{I} - \mathbf{A})\gamma = \mathbf{\Sigma}$$

$$\gamma = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{\Sigma} \tag{42}$$

where $\mathbf{I}$ is a $(p+1)$ identity matrix and

$$\begin{aligned} \gamma &= [\gamma(0)\ \ \gamma(1)\ \ \gamma(2)\ \ \cdots\ \gamma(p)]' \\ \mathbf{\Sigma} &= [\sigma^2\ 0\ 0\ \cdots\ 0]'. \end{aligned}$$

The $\mathbf{A}$ matrix for the different AR(p) models can be seen to follow the sequence

$$\mathbf{A}_{AR(1)} = \begin{bmatrix} 0 & \phi_1 \\ \phi_1 & 0 \end{bmatrix}, \ \mathbf{A}_{AR(2)} = \begin{bmatrix} 0 & \phi_1 & \phi_2 \\ \phi_1 & \phi_2 & 0 \\ \phi_2 & \phi_1 & 0 \end{bmatrix},$$

$$\mathbf{A}_{AR(3)} = \begin{bmatrix} 0 & \phi_1 & \phi_2 & \phi_3 \\ \phi_1 & \phi_2 & \phi_3 & 0 \\ \phi_2 & (\phi_1+\phi_3) & 0 & 0 \\ \phi_3 & \phi_2 & \phi_1 & 0 \end{bmatrix},$$

$$\mathbf{A}_{AR(4)} = \begin{bmatrix} 0 & \phi_1 & \phi_2 & \phi_3 & \phi_4 \\ \phi_1 & \phi_2 & \phi_3 & \phi_4 & 0 \\ \phi_2 & (\phi_1+\phi_3) & \phi_4 & 0 & 0 \\ \phi_3 & (\phi_2+\phi_4) & \phi_1 & 0 & 0 \\ \phi_4 & \phi_3 & \phi_2 & \phi_1 & 0 \end{bmatrix}, \text{ etc}$$

Once the $(p+1)$ entries in γ vector are found, find the autocovariance $\gamma(j)$ of any order $j > p$ from the difference equation of autocovariance given in (39b), ie. , from

$$\gamma(j) = \phi_1\gamma(j-1) + \phi_2\gamma(j-2) + \cdots + \phi_p\gamma(j-p)\,. \tag{43}$$

Deflating (43) by $\gamma(0) = \mathrm{Var}\left(\tilde{X}_t\right)$ gives us the **autocorrelation function** (ACF), which is also a difference equation of order p, taking the form

$$\rho(j) = \phi_1\rho(j-1) + \phi_2\rho(j-2) + \cdots + \phi_p\rho(j-p)\,. \tag{44}$$

Notice here also that we can re-write the relation for the case when $j = 0$ in (39a) by noting that

$$
\begin{aligned}
\gamma(0) &= \phi_1\gamma(1) + \cdots + \phi_p\gamma(p) + \sigma^2 \\
\gamma(0) - \phi_1\gamma(1) - \cdots - \phi_p\gamma(p) &= \sigma^2 \\
\gamma(0)\left[1 - \phi_1\rho(1) - \cdots - \phi_p\rho(p)\right] &= \sigma^2 \\
\gamma(0) &= \frac{\sigma^2}{\left[1 - \phi_1\rho(1) - \cdots - \phi_p\rho(p)\right]} \qquad (45) \\
\text{Var}(X_t) = \gamma(0) &= \frac{\sigma^2}{\left[1 - \sum_{i=1}^{p}\phi_i\rho(i)\right]} \qquad (46)
\end{aligned}
$$

with $\rho(j) = \frac{\gamma(j)}{\gamma(0)}$.

Example AR(1)

Let $\tilde{X}_t = (X_t - \mu)$ be generated by

$$\phi(L)\tilde{X}_t = Z_t \qquad (47)$$

where $\phi(L) = 1 - \phi_1 L$ and $Z_t \sim \text{WN}\left(0, \sigma^2\right)$.

$$\gamma(j) = \begin{cases} \phi_1\gamma(1) + \sigma^2 & \text{for } j = 0 \\ \phi_1\gamma(j-1) & \text{for } j > 0 \end{cases} \qquad (48)$$

so that we have $\gamma = [\gamma(0) \;\; \gamma(1)]'$, $\mathbf{\Sigma} = \left[\sigma^2 \;\; 0\right]'$ and $\mathbf{A} = \begin{bmatrix} 0 & \phi_1 \\ \phi_1 & 0 \end{bmatrix}$ and the system looks like

$$
\begin{aligned}
\begin{bmatrix} \gamma(0) \\ \gamma(1) \end{bmatrix} &= \begin{bmatrix} 0 & \phi_1 \\ \phi_1 & 0 \end{bmatrix} \begin{bmatrix} \gamma(0) \\ \gamma(1) \end{bmatrix} + \begin{bmatrix} \sigma^2 \\ 0 \end{bmatrix} \qquad (49) \\
\gamma &= \mathbf{A}\gamma + \mathbf{\Sigma}. \qquad (50)
\end{aligned}
$$

We then get

$$\begin{aligned} \gamma &= (\mathbf{I} - \mathbf{A})^{-1} \boldsymbol{\Sigma} && (51) \\ \begin{bmatrix} \gamma(0) \\ \gamma(1) \end{bmatrix} &= \begin{bmatrix} \frac{\sigma^2}{\left(1-\phi_1^2\right)} \\ \frac{\phi_1 \sigma^2}{\left(1-\phi_1^2\right)} \end{bmatrix} && (52) \end{aligned}$$

and corresponding ACFs

$$\begin{bmatrix} \frac{\gamma(0)}{\gamma(0)} \\ \frac{\gamma(1)}{\gamma(0)} \end{bmatrix} = \begin{bmatrix} \rho(0) \\ \rho(1) \end{bmatrix} = \begin{bmatrix} 1 \\ \phi_1 \end{bmatrix}. \quad (53)$$

All ACFs for $j > 1$ are generated from the recursion $\rho(j) = \phi_1 \rho(j-1)$, so for

$$\begin{aligned} j = 2: \quad \rho(2) = {} & \phi_1 \rho(1) && = \phi_1^2 \\ j = 3: \quad \rho(3) = {} & \phi_1 \rho(2) && = \phi_1^3 \\ j = m: \quad \rho(m) = {} & \phi_1 \rho(m-1) && = \phi_1^m. \\ & \vdots \end{aligned} \quad (54)$$

Example AR(2)

Consider the AR(2) $\left(1-\phi_1 L-\phi_2 L^2\right)\tilde{X}_t = Z_t$, giving

$$\begin{aligned}\gamma(0) &= \phi_1\gamma(1)+\phi_2\gamma(2)+\sigma^2\\ \gamma(1) &= \phi_1\gamma(0)+\phi_2\gamma(1)\\ \gamma(2) &= \phi_1\gamma(1)+\phi_2\gamma(0)\end{aligned}$$

Here we have the relation

$$\begin{bmatrix}\gamma(0)\\ \gamma(1)\\ \gamma(2)\end{bmatrix} = \begin{bmatrix}0 & \phi_1 & \phi_2\\ \phi_1 & \phi_2 & 0\\ \phi_2 & \phi_1 & 0\end{bmatrix}\begin{bmatrix}\gamma(0)\\ \gamma(1)\\ \gamma(2)\end{bmatrix} + \begin{bmatrix}\sigma^2\\ 0\\ 0\end{bmatrix} \tag{55}$$

which then gives the solution

$$\gamma = D^{-1}\sigma^2\begin{bmatrix}(1-\phi_2)\\ \phi_1\\ \phi_1^2+\phi_2(1-\phi_2)\end{bmatrix} \tag{56}$$

where $D=(1+\phi_2)\left[(1-\phi_2)^2-\phi_1^2\right]$ is the determinant of $(\mathbf{I}-\mathbf{A})$.

The ACF vector is given by

$$\begin{bmatrix}\rho(0)\\ \rho(1)\\ \rho(2)\end{bmatrix} = \begin{bmatrix}1\\ \frac{\phi_1}{(1-\phi_2)}\\ \frac{\phi_1^2}{(1-\phi_2)}+\phi_2\end{bmatrix} \tag{57}$$

and any higher order can again be compute from the ACF recursion

$$\rho(j) = \phi_1 \rho(j-1) + \phi_2 \rho(j-2), \forall j > 2. \quad (58)$$

Hence for

$$\begin{array}{llll} j=3: & \rho(3) = & \phi_1\rho(2) + \phi_2\rho(1) & = \frac{\phi_1\phi_2(2-\phi_2)+\phi_1^3}{(1-\phi_2)} \\ j=4: & \rho(4) = & \phi_1\rho(3) + \phi_2\rho(2) & = \frac{\phi_1^2\phi_2(2-\phi_2)+\phi_1^4}{(1-\phi_2)} + \frac{\phi_2\phi_1^2}{(1-\phi_2)} + \phi_2^2 \\ & \vdots & & \end{array} \quad (59)$$

Yule Walker equations and PACFs

There exists a relationship between the ϕ_i parameters ($\forall i = 1, \dots, p$) in the AR(p) model and the ACFs.

Take $j = 1, \dots, p$ in (44) and form the following system of equations for the ACFs

$$\begin{array}{lll} \rho(1) & = & \phi_1 + \phi_2\rho(1) + \cdots + \phi_p\rho(p-1) \\ \rho(2) & = & \phi_1\rho(1) + \phi_2 + \cdots + \phi_p\rho(p-2) \\ & & \vdots \\ \rho(p) & = & \phi_1\rho(p-1) + \phi_2\rho(p-2) + \cdots + \phi_p \end{array}$$

which in matrix form can be re-expressed as

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(p) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(p-1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(p-2) \\ \rho(2) & \rho(1) & 1 & \cdots & \rho(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \rho(p-3) & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_p \end{bmatrix} \tag{60}$$

$$\rho = \mathbf{R}\phi. \tag{61}$$

These equations above in (60) are known as **Yule Walker** equations.

Because PACF defined as $\phi_{jj} = \text{Corr}\left(X_t, X_{t-j} | X_{t-1}, X_{t-2}, \ldots, X_{t-j+1}\right)$ **controls for** impact of $\{X_{t-i}\}_{i=1}^{j-1}$ when gauging correlation between X_{t-j} and X_t, think of the PACF within **population regression** context:

$$X_t = \phi_{j1} X_{t-1} + \phi_{j2} X_{t-2} + \cdots + \phi_{jj} X_{t-j} + Z_t \tag{62}$$

where the regression coefficient on the j^{th} term, ie., ϕ_{jj}, gives the j^{th} PACF.

Another way to get PACFs is to use Yule Walker equations in (60) to formulate system of linear equations:

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(j-1) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(j-1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(j-2) \\ \rho(2) & \rho(1) & 1 & \cdots & \rho(j-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(j-1) & \rho(j-2) & \rho(j-3) & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{j1} \\ \phi_{j2} \\ \phi_{j3} \\ \vdots \\ \phi_{jj} \end{bmatrix} \tag{63}$$

$$\rho_j = \mathbf{R}_j \phi_j \tag{64}$$

and then extract PACFs recursively by solving for ϕ_{jj} term as needed.

For example, set $\phi_{11} = \rho(1)$ for the first PACF and then obtain the remaining PACFs, for $j = 2, 3, 4, \ldots$, from

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}}, \quad \phi_{33} = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{vmatrix}},$$

$$\phi_{44} = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(2) & \rho(1) \\ \rho(1) & 1 & \rho(2) & \rho(2) \\ \rho(2) & \rho(1) & 1 & \rho(3) \\ \rho(3) & \rho(2) & \rho(1) & \rho(4) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) & \rho(3) \\ \rho(1) & 1 & \rho(2) & \rho(2) \\ \rho(2) & \rho(1) & 1 & \rho(1) \\ \rho(3) & \rho(2) & \rho(1) & 1 \end{vmatrix}}, \cdots$$

where $|\cdot|$ denotes the determinant. Notice here that we are using **Cramer's Rule** to solve for the last entry of ϕ_{jj} of interest to us.

Remark: Cramer's Rule

To solve for the i^{th} element of a system of equations $Ax = b$ with Cramer's Rule, one needs to find

$$x_i = \frac{|A_i|}{|A|}$$

where A_i is the *augmented matrix* and is formed by replacing the i^{th} column of matrix A with vector b and $|\cdot|$ denotes the determinant.

The representation above in (63) is a good and intuitive way to understand how the PACFs are related to the ACFs

- but is inefficient way to compute them
- a fast and efficient way to compute the PACFs is to use the so called **Durbin-Levinson Algorithm**.

Definition (Durbin-Levinson Algorithm)

The PACFs can be obtained recursively from the following relations. Initialise $\phi_{11} = \rho(1)$ and let $\phi_{jk} = \phi_{j-1,k} - \phi_{jj}\phi_{j-1,j-k}$ for all $k = 1, 2, \ldots, j-1$ and $j > 1$. Then, ϕ_{jj} is computed as

$$\phi_{jj} = \frac{\rho(j) - \sum_{k=1}^{j-1} \phi_{j-1,k}\rho(j-k)}{1 - \sum_{k=1}^{j-1} \phi_{j-1,k}\rho(k)}, \ \forall j > 1.$$

Example: General PACF calculation

PACF recursively extracted from the two approaches

First, note that $\phi_{11} = \rho(1)$. Then,

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{vmatrix}} = \frac{\rho(3)\left[1 - \rho(1)^2\right] + \rho(1)^3 + \rho(2)\left[\rho(1)\left\{\rho(2) - 2\right\}\right]}{\left[\rho(2) - 1\right]\left[2\rho(1)^2 - \rho(2) - 1\right]}$$

Using the **Durbin-Levinson Algorithm**, we get for $j = 2$

$$\phi_{22} = \frac{\rho(2) - \sum_{k=1}^{1} \phi_{1k}\rho(2-k)}{1 - \sum_{k=1}^{1} \phi_{1k}\rho(k)}$$

$$= \frac{\rho(2) - \phi_{11}\rho(1)}{1 - \phi_{11}\rho(1)}$$

$$= \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}.$$

For $j = 3$, we get (with $\phi_{2k} = \phi_{1k} - \phi_{22}\phi_{1,2-k}$)

$$\phi_{33} = \frac{\rho(3) - \sum_{k=1}^{2} \phi_{2k}\rho(3-k)}{1 - \sum_{k=1}^{2} \phi_{2k}\rho(k)}$$

$$= \frac{\rho(3) - [(\phi_{11} - \phi_{22}\phi_{11})\rho(2) + \phi_{22}\rho(1)]}{1 - [(\phi_{11} - \phi_{22}\phi_{11})\rho(1) + \phi_{22}\rho(2)]}$$

$$= \frac{\rho(3) - [\rho(1)(1 - \phi_{22})\rho(2) + \phi_{22}\rho(1)]}{1 - [\rho(1)^2(1 - \phi_{22}) + \phi_{22}\rho(2)]}$$

$$= \frac{\rho(3)\left[1 - \rho(1)^2\right] + \rho(1)^3 + \rho(2)\left[\rho(1)\{\rho(2) - 2\}\right]}{[\rho(2) - 1][2\rho(1)^2 - \rho(2) - 1]} \quad \text{etc.}$$

Remark: PACFs

"Estimates of the PACFs ϕ_{jj} obtained using the Yule Walker equations become very sensitive to rounding errors and should not be used if the values of the parameter are close to the non-stationary boundaries." Box *et al.* (1994), page 68.

Example: PACF of AR(1)

Recall that AR(1) has $\rho(j) = \phi_1 \rho(j-1)$ for all $j = 0, 1, 2, \ldots$. Thus, we have

$$\begin{aligned} \phi_{11} &= \rho(1) = \phi_1 \\ \phi_{22} &= \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}. \end{aligned}$$

But

$$\begin{aligned} \rho(2) &= \phi_1 \rho(1) \\ &= \rho(1)^2 \end{aligned}$$

because of $\rho(1) = \phi_1$. Hence $\phi_{22} = 0$.

Also, for ϕ_{33} we have in the numerator

$$\rho(3)\left[1-\rho(1)^2\right]+\rho(1)^3+\rho(2)\left[\rho(1)\left\{\rho(2)-2\right\}\right].$$

With $\rho(j)=\rho(1)^j$ we get

$$\begin{aligned} & \rho(3)\left[1-\rho(1)^2\right]+\rho(1)^3+\rho(2)\left[\rho(1)\left\{\rho(2)-2\right\}\right] \\ = & \rho(1)^3\left[1-\rho(1)^2\right]+\rho(1)^3+\rho(1)^2\left[\rho(1)\left\{\rho(1)^2-2\right\}\right] \\ = & 2\rho(1)^3-\rho(1)^5+\rho(1)^5-2\rho(1)^3 \\ = & 0 \end{aligned}$$

so the PACF for $j>1=0$ for an AR(1) process.

Summary of ACF and PACF structure of AR(p) Models

Process	ACF	PACF
White Noise Z_t	0 for all j	0 for all j
AR(1)	$\rho(j)=\phi_1^j$ for all $j>0$	$\phi_{jj}=\rho(j)=\phi_1$ for $j=1$ and 0 for $j>1$
AR(p)	exp. decline to 0	non-zero for first p lags and 0 for $j>p$

Table 2: ACF and PACF properties of an AR(p) process.

Stationarity/Stability of AR(p) models

Stationarity of the AR(p) model is determined by the lag polynomial $\phi(L)$.

- common to use terms stationarity and stability of a time series process interchangeably

Definition (Stability of Lag Polynomial)

An AR(p) process $\phi(L)X_t = Z_t$ is said to be stable (stationary) if all the roots of $\phi(z) = 0$ lie **outside** the unit circle. Equivalently, we have the condition

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p \neq 0, \ \forall |z| \leq 1, \tag{65}$$

ie., the lag polynomial is **not** equal to zero for all $|z| \leq 1$, where $|\cdot|$ denotes the modulus.

Remark: $|x|$ is the modulus of x which is the absolute value for real x and $\sqrt{a^2 + b^2}$ for complex x, ie., when $x = a + bi$, where $i = \sqrt{-1}$.

Remark: The lag polynomial is expressed in terms of the **variable** z and not the **lag operator** L. This is a technical necessity because L is an operator and not a variable, so cannot be used like a variable to find the solutions of the polynomial.

An alternative definition is to state the properties of the roots of $\phi(z) = 0$ in terms of the **Factored Polynomial**, where we express $\phi(z)$ as

$$1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = (1 - \lambda_1 z)(1 - \lambda_2 z) \ldots (1 - \lambda_p z). \tag{66}$$

Definition (Stability of Factored Polynomial)

An AR(p) process $\phi(L) X_t = Z_t$ is said to be stable (stationary) if all the roots of $(1 - \lambda_1 z)(1 - \lambda_2 z) \ldots (1 - \lambda_p z) = 0$ lie **inside** the unit circle. Equivalently, we have the condition

$$(1 - \lambda_1 z)(1 - \lambda_2 z) \ldots (1 - \lambda_p z) \neq 0, \ \forall |\lambda| \geq 1, \tag{67}$$

ie., the **factored polynomial** is **not** equal to zero for all $|\lambda| \geq 1$, where $|\cdot|$ denotes the modulus.

Fact: The roots of the **lag polynomial** are equal to the **inverse** of the roots of the **factored polynomial**

Example AR(2) roots:

Let X_t be an AR(2) process, taking the form

$$\phi(L) X_t = Z_t$$

where

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2.$$

Then

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

and to check for the stability of the AR(2) we need to find

$$\phi(z) = 0.$$

For the AR(2) process to be stable we need $|z_i| > 1, \forall i = 1, 2$.

The **roots of the lag polynomial** are found as the solutions to

$$1 - \phi_1 z - \phi_2 z^2 = 0 \tag{68}$$

which are at

$$z_1 = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}, \; z_2 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2}. \tag{69}$$

These will be real as long as $\phi_1^2 + 4\phi_2 \geq 0$.

If you have forgotten how to find the roots of a quadratic, consider the general problem of finding the solutions of the second order polynomial (quadratic function)

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

which will be

$$x_1 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}, \; x_2 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}.$$

There will always be 2 roots, possibly complex and repeated (Fundamental theorem of Algebra).

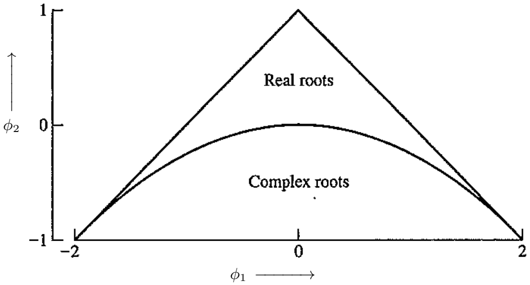


Figure 2: Stability region of an AR(2) model.

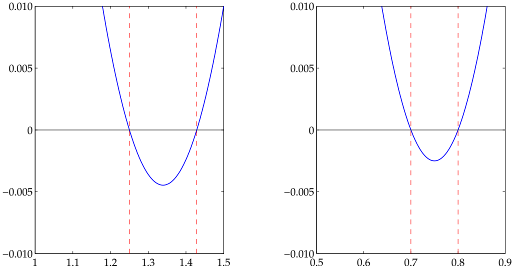
Numerical Example AR(2) with real roots

Let X_t follow an AR(2) process of the form $\left(1 - \phi_1 L - \phi_2 L^2\right) X_t = Z_t$, with parameters $\phi_1 = 1.5$, $\phi_2 = -0.56$. The lag polynomial is then

$$\phi(z) = 1 - 1.5z + 0.56z^2.$$

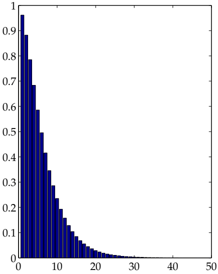
Plugging the values for ϕ into (69) yields the roots $z_1 = 1.4286$ and $z_2 = 1.25$. These are both greater than 1 in absolute value hence the AR(2) process is stable/stationary.

The factored roots $(1 - \lambda_1 z)(1 - \lambda_2 z)$ are $\lambda_1 = 0.7$ and $\lambda_2 = 0.8$ and we can easily see that $\lambda_i^{-1} = z_i$.

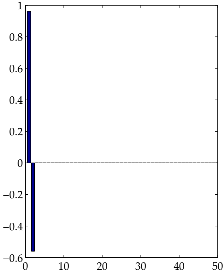


(a) Roots of lag polynomial (b) Factored roots

Figure 3: Roots of the AR(2) process.



(a) Theoretical ACF of the AR(2)



(b) Theoretical PACF of the AR(2)

Figure 4: Theoretical ACFs and PACFs of the AR(2) process.

Numerical Example Stable AR(2) with complex roots

Let the AR(2) parameters now be $\phi_1 = 1.4$, $\phi_2 = -0.85$. The lag polynomial is given by

$$\phi(z) = 1 - 1.4z + 0.85z^2$$

which has complex roots $z_{1,2} = 0.8235 \pm 0.7059i$ with modulus $\sqrt{0.8235^2 + 0.7059^2} = 1.0847 > 1$.

The factored roots $(1 - \lambda_1 z)(1 - \lambda_2 z)$ of $(\lambda^2 - \phi_1\lambda - \phi_2)$ are $\lambda_{1,2} = 0.7000 \pm 0.6000i$ with modulus $\sqrt{0.7000^2 + 0.6000^2} = 0.9220$.

We can again easily check that $\lambda^{-1} = z$ where the inverse of a complex number is computed as

$$z^{-1} = \frac{z_+}{|(z^-)^2|} \tag{70}$$

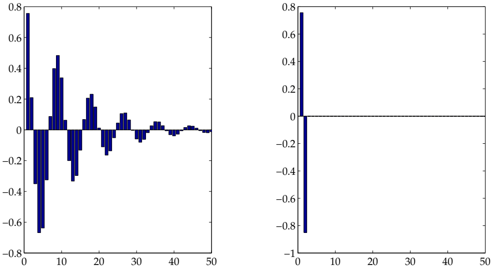
where $z^+ = a + bi$ ($z^- = a - bi$) and a and b are the coefficients of the complex number representation $x = a + bi$.

This yields $z^+ = 0.8235 + 0.7059i$ and $z^- = 0.8235 - 0.7059i$ so that

$$z^{-1} = \frac{0.8235 + 0.7059i}{|(0.8235 - 0.7059i)^2|}$$

$$= \frac{0.8235 + 0.7059i}{1.1764}$$

$$= 0.7000 + 0.6000i.$$



(a) Theoretical ACF of the AR(2) (b) Theoretical PACF of the AR(2)

Figure 5: Plots of the theoretical ACFs and PACFs from AR(2) with complex roots.

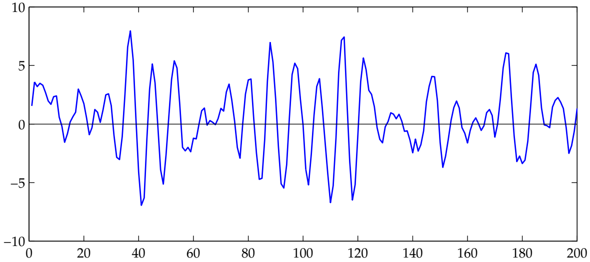


Figure 6: Time series plot of the AR(2) process with complex roots.

Mean of an MA(q) process

Proposition (Mean)

Let X_t be generated by

$$X_t = c + \theta(L) Z_t \tag{71}$$

where $Z_t \sim \text{WN}(0, \sigma^2)$ and $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$. Then

$$\mu = E(X_t) = c.$$

Proof.

Taking expectations of (71) we have

$$\begin{aligned} E(X_t) &= E(c) + E[\theta(L) Z_t] \\ E(X_t) &= c + \theta(L) E[Z_t] \\ E(X_t) &= c \end{aligned}$$

because $Z_t \sim \text{WN}(0, \sigma^2)$, hence uncorrelated at different time periods. □

Note: We did not make any statements about stationarity when defining the MA(q) process. $\Rightarrow$ an MA process is **always stationary**.

Autocovariance of an MA(q) process

Proposition (Autocovariance)

The process in (71) has autocovariances given by

$$\gamma(j) = \begin{cases} \sigma^2 \sum_{i=0}^{q-j} \theta_i \theta_{i+j} & \text{for } j \leq q \quad (72a) \\ 0 & \text{for } j > q \quad (72b) \end{cases}$$

for all $j = 0, 1, 2, \ldots,$ *where*

$$\sum_{i=0}^{q-j} \theta_i \theta_{i+j} = (\theta_j + \theta_1 \theta_{j+1} + \cdots + \theta_q \theta_{q-j}) \quad (73)$$

and $\theta_0 = 1$.

Proof.

Expand the $\mathrm{Cov}(X_t, X_{t-j}) = E\left[(X_t - \mu)(X_{t-j} - \mu)\right]$ terms to yield

$$\mathrm{Cov}(X_t, X_{t-j}) = E[(Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}) \times (Z_{t-j} + \theta_1 Z_{t-1-j} + \theta_2 Z_{t-2-j} + \cdots + \theta_q Z_{t-q-j})]$$

and then match all the time periods of Z_s for all $s = 1, \ldots, j$ and take expectations. Since the Z_t are uncorrelated across time, we get the desired result. $\square$

Example: MA(4)

Let $X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \theta_3 Z_{t-3} + \theta_4 Z_{t-4}$. Expanding (72) (ie. $\sum_{i=0}^{q-j} \theta_i \theta_{i+j}$) we have for $j = 0$ (the variance)

$$\gamma(0) = \left(1 + \theta_1^2 + \theta_2^2 + \theta_3^2 + \theta_4^2\right)\sigma^2,$$

and for $0 < j \leq q = 4$

$$\begin{aligned} \gamma(1) &= (\theta_1 + \theta_1\theta_2 + \theta_2\theta_3 + \theta_3\theta_4)\,\sigma^2 \\ \gamma(2) &= (\theta_2 + \theta_1\theta_3 + \theta_2\theta_4)\,\sigma^2 \\ \gamma(3) &= (\theta_3 + \theta_1\theta_4)\,\sigma^2 \\ \gamma(4) &= (\theta_4)\,\sigma^2. \end{aligned}$$

For any $j > q = 4$ we get $\gamma(j) = 0$.

Now it should be evident that an MA(q) process has:

$$\gamma(j) = \begin{cases} \theta_q \sigma^2 & \text{if } j = q \\ 0 & \text{for all } j > q. \end{cases} \tag{74}$$

This result is formalised by the following proposition on $q-$correlated series given in Brockwell and Davis (2002), page 50.

Proposition ($q-$correlated series)

If X_t is a stationary q-correlated time series with mean 0, then it can be represented as an MA(q) process.

⇒ only need to look at the correlation structure of the various White Noise components to work out the MA order of a process.

ACF and PACF

The ACF of an MA(q) process, is also found from the relation

$$\rho(j) = \frac{\gamma(j)}{\gamma(0)} \quad (75)$$

where $\rho(j)$ inherits the decay properties of $\gamma(j)$ so that for $j > q$, $\rho(j) = 0$.

The PACF of an MA(q) from the Yule Walker relations

- we can again use the Durbin-Levinson Algorithm or
- or Cramer's rule on Yule Walker

Example PACF of MA(2)

Consider the following MA(2) process, where

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$$

where

$$\rho(1) = (\theta_1 + \theta_1\theta_2)/(1 + \theta_1^2 + \theta_2^2)$$

and

$$\rho(2) = \theta_2/(1 + \theta_1^2 + \theta_2^2)$$

and we have $\rho(3) = \rho(4) =, \cdots, \rho(h) = 0$.

Then $\phi_{11} = \rho(1)$ and

$$\phi_{22} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}$$

$$\phi_{33} = \frac{\rho\left(1\right)^3 + \rho\left(2\right)\left[\rho\left(1\right)\left\{\rho\left(2\right)-2\right\}\right]}{\left[\rho\left(2\right)-1\right]\left[2\rho\left(1\right)^2-\rho\left(2\right)-1\right]}$$

$$\phi_{44} = \frac{\rho(1)^4 - 2\rho(1)^2\rho(2)^2 + 3\rho(1)^2\rho(2) + \rho(2)^4 - \rho(2)^2}{\rho(1)^4 - 2\rho(1)^2\rho(2)^2 + 4\rho(1)^2\rho(2) - 3\rho(1)^2 + \rho(2)^4 - 2\rho(2)^2 + 1}$$

$$\vdots$$

Higher order PACFs follow a similar pattern.

Since the PACF is a function of $\rho(1)$ and $\rho(2)$ and these are non-zero

$\Rightarrow$ PACF for an MA process will decay slowly towards zero.

Summary of ACF and PACF structure of MA(q) Models

Process	ACF	PACF
White Noise Z_t	0 for all j	0 for all j
MA(1)	$\rho(j) = \frac{\theta}{(1+\theta^2)}$ for $j = 1$ and 0 for $j > 1$	exponential decline to 0
MA(q)	non-zero for first q lags and 0 for $j > q$	exponential decline to 0

Table 3: ACF and PACF properties of an MA(q) process.

Invertibility of MA(q) models

- No assumptions about the $\theta(L)$ polynomial have been made regarding stationarity
- But need similar restrictions to have an **invertible** MA(q) model.

Definition (Invertibility)

The MA(q) process $X_t = c + \theta(L) Z_t$, $Z_t \sim \text{WN}(0, \sigma^2)$ is said to be invertible if the roots of $\theta(z)$ are greater than 1 in absolute value ($|z| > 1$), or equivalently, if

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q \neq 0, \ \forall |z| \leq 1.$$

Alternatively, in terms of the roots of the factored polynomial

$$(1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q) = (1 - \lambda_1 z)(1 - \lambda_2 z) \ldots (1 - \lambda_q z)$$

the MA(q) process is said to be invertible if $|\lambda_i| < 1, \ \forall i = 1. \ldots, q$.

Invertibility is needed:

a) to have a unique mapping between the ACF and the $\{\theta_i\}_{i=1}^q$ parameters (identification)

b) to have an AR(∞) representation of the MA(q) process.

c) to be able to estimate MA models by Maximum Likelihood.

- If the $\theta(L)$ polynomial is **not invertible**, can have, in general up to 2^q representations of the ACF of the MA(q) in terms of the $\{\theta_i\}_{i=1}^q$
- Will not be able to tell which set of $\{\theta_i\}_{i=1}^q$ parameters generated the observed series X_t.
  - referred to in the econometrics literature as an **identification** problem.

ARMA Algebra

Properties of MA(q) models

- Working with a non-invertible MA process is not a problem in general
  - model can be solved forward to get the X_t representation,
  - but will need all future values of X
  - working with a non-invertible MA process is not very practical.
- the value of Z_t associated with the invertible MA representation is frequently referred to as the **fundamental innovation** for X_t.

Having fundamental representation for process means we have same information regardless of whether we express an ARMA as an MA(∞) or as an AR(∞).

- we can therefore move between these definitions without loosing any *information* about the process.

ARMA Algebra

Properties of MA(q) models

- when an MA process is not invertible, we can always reformulate the model by backing out the θ parameters that correspond to the inverse of the non-invertible roots.
- inverses of non-invertible roots will be invertible,
  - corresponding θ parameters yield an invertible MA model with **the same first and second moments** as the non-invertible representation.
- Invertibility condition ensures that, as long as $\sum_{j=0}^{\infty} |\pi_j| < \infty$ (the coefficients of the AR(∞) representation), the AR(∞) representation of an MA(q) process exists.

That is, if $\theta\left(z\right) \neq 0,\ \forall \left|z\right| \leq 1$, then we can write

$$\begin{aligned} X_t &= c + \theta\left(L\right) Z_t \\ \theta\left(L\right)^{-1} X_t &= \theta\left(L\right)^{-1} c + Z_t \\ X_t &= \theta\left(1\right)^{-1} c + \sum_{j=1}^{\infty} \pi_j X_{t-j} + Z_t \end{aligned} \tag{76}$$

which is AR(∞), where $\left\{\pi_j\right\}_j^{\infty}$ are determined from $\left\{\theta_i\right\}_{i=1}^{q}$ of MA(q).

Example Invertibility/Identification

Let X_t be an MA(1) with the representation

$$\begin{aligned} X_t &= \theta\left(L\right) Z_t \\ X_t &= (1 + \theta_1 L) Z_t \\ &= Z_t + \theta_1 Z_{t-1}. \end{aligned}$$

The autocovariances of the process are

$$\begin{aligned} \gamma\left(0\right) &= (1 + \theta_1^2)\sigma^2 \\ \gamma\left(1\right) &= \theta_1 \sigma^2. \end{aligned}$$

The ACF is then

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\theta_1}{(1+\theta_1^2)}. \tag{77}$$

Note that for any θ_1, $|\rho(1)| \le \frac{1}{2}$. Re-arranging (77) to get the relation

$$\rho(1)\,\theta_1^2 - \theta_1 + \rho(1) = 0$$

we will get the following two solutions for θ:

$$\theta_1^{(1)} = \frac{1-\sqrt{1-4\rho(1)^2}}{2\rho(1)},\ \theta_1^{(2)} = \frac{1+\sqrt{1-4\rho(1)^2}}{2\rho(1)} \tag{78}$$

For example, for the two solutions in (78)

$$\theta_1^{(1)} = 0.5,\ \theta_1^{(2)} = 1/\theta_1^{(1)} = 2$$

we get the ACF value of $\rho(1) = 0.4$.

Need another restriction to **identify** the value of the θ_1 parameter from $\rho(1)$.

⇒ choose the invertible model, ie., the model with the **root of the factored polynomial** $\lambda = |\theta_1| < 1$, which here is the first root $\theta_1^{(1)} = 0.5$.

- with $\lambda = \theta_1^{(1)}$, we have that $z = 1/\theta_1^{(1)} = 2$, so the **root** of the **lag polynomial** is bigger than 1.

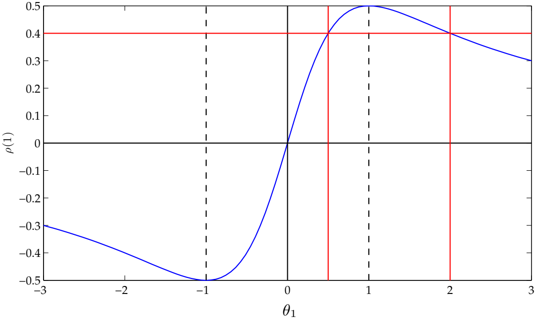


Figure 7: Plot of the mapping between $\rho(1)$ and θ_1 for an MA(1), where $\rho(1) = \theta_1/(1+\theta_1^2)$.

Example Non-invertible MA(2) to invertible MA(2)

Let X_t follow an MA(2) process, taking the form

$$X_t = \theta(L)Z_t \tag{79}$$

$$\begin{aligned} X_t &= (1 + \theta_1 L + \theta_2 L^2)Z_t \qquad (80) \\ &= Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} \end{aligned}$$

with $\theta_1 = -3.5$ and $\theta_2 = -2$ and $Z_t \sim \text{WN}(0,1)$. We then have

$$\theta(z) = 0$$

$$(1 - 3.5z - 2z^2) = 0$$

at $z_1 = -2$ and $z_2 = 0.25$.

This implies that the factored roots are λ_1

$$\theta(z) = (1 - \lambda_1 z)(1 - \lambda_2 z)$$

$$0 = (1 - \underbrace{[z_1^{-1}]}_{\lambda_1} z)(1 - \underbrace{[z_2^{-1}]}_{\lambda_2} z) \quad (81)$$

$$= (1 + 0.5z)(1 - 4z) \quad (82)$$

ie., $\lambda_1 = -0.5$ and $\lambda_2 = 4$.

$\Rightarrow$ modulus of z_1 (λ_1) is greater (smaller) than 1, modulus of z_2 (λ_2) is less (greater) than 1.

$\Rightarrow$ the MA(2) is not invertible.

The ACF can be simply found from (73) (and deflating by $\gamma(0)$) as

$$\rho(1) = \frac{(\theta_1 + \theta_1\theta_2)}{(1 + \theta_1^2 + \theta_{21}^2)} = \frac{(-3.5 + 7)}{(1 + 3.5^2 + 2^2)} = 0.20290$$

$$\rho(2) = \frac{\theta_2}{(1 + \theta_1^2 + \theta_{21}^2)} = \frac{-2}{(1 + 3.5^2 + 2^2)} = -0.11594$$

$$\rho(3) = 0$$

$$\vdots$$

From (82), the problematic root is z_2 (λ_2).

$\Rightarrow$ create invertible MA(2) process with same second moment structure as non-invertible one by inverting non-invertible root

Create

$$\theta(z)^{+} = (1 + 0.5z)(1 - \tfrac{1}{4}z)$$
$$\theta(z)^{+} = 1 + 0.25z - 0.125\,z^2 \tag{83}$$

where

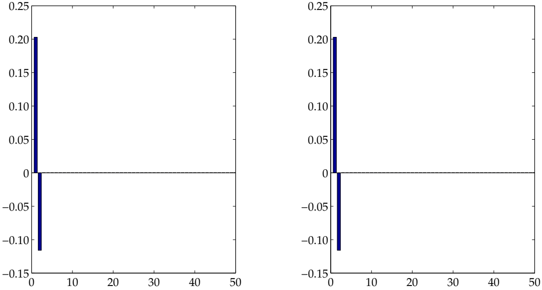
$$\theta_1^{+} = -(\lambda_1 + \lambda_2^{+})$$
$$= -(-0.5 + \frac{1}{4})$$

$$\theta_2^{+} = \lambda_1 \lambda_2^{+}$$
$$= -0.5 \times \frac{1}{4}$$

$\theta(z)^{+}$ denotes the invertible lag polynomial and λ_2^{+} is the invertible second factored root computed from λ_2^{-1}.

The autocovariances are formed again as

$$\rho(1) = \frac{(\theta_1 + \theta_1\theta_2)}{(1 + \theta_1^2 + \theta_{21}^2)} = \frac{(0.25 - 0.25 \times 0.125)}{(1 + 0.25^2 + 0.125^2)} = \quad 0.20290$$
$$\rho(2) = \frac{\theta_2}{(1 + \theta_1^2 + \theta_{21}^2)} = \frac{-0.125}{(1 + 0.25^2 + 0.125^2)} = -0.11594$$
$$\rho(3) = 0$$
$$\vdots$$



(a) $\theta(z) = (1 - 3.5z - 2z^2)$ (b) $\theta(z)^+ = (1 + 0.25z - 0.125\,z^2)$

Figure 8: Plots of the theoretical ACFs of the non-invertible and invertible MA(2) models.

Moving between AR, MA and ARMA representations

Recall that a general stationary and invertible ARMA(p,q) model is defined as

$$\phi(L)X_t = c + \theta(L)Z_t, \tag{84}$$

where $Z_t \sim \text{WN}(0, \sigma^2)$, and both $\phi(L)$ and $\theta(L)$ are invertible.

- will use the commonly followed notational convention to denote by $\psi(L)$ the weights of the MA(∞) representation of a general ARMA(p, q) model and
- will use $\pi(L)$ to denote the corresponding AR(∞) representation.

Definition (ARMA to MA(∞))

Let $\phi(L) X_t = c + \theta(L) Z_t$ be a stationary and invertible ARMA(p,q) process. Then the coefficients $\psi(L)$ of the MA(∞) representation, where $\psi(L) = \frac{\theta(L)}{\phi(L)}$ are given by the following recursion:

$$\psi_j = \theta_j + \sum_{k=1}^{p} \phi_k \psi_{j-k}, \ \forall j = 0, 1, 2, \ldots, \tag{85}$$

where $\psi_i = 0$ for $i < 0$, $\theta_0 = 1, \theta_j = 0$ for $j > q$. So the ARMA(p,q)

$$\begin{aligned} \phi(L) X_t &= c + \theta(L) Z_t \\ X_t &= \frac{c}{\phi(1)} + \frac{\theta(L)}{\phi(L)} Z_t \\ X_t &= \frac{c}{\phi(1)} + \psi(L) Z_t \end{aligned}$$

becomes an MA(∞), where the ψ_j coefficients are determined by (85).

Definition (ARMA to AR(∞))

Let $\phi(L) X_t = c + \theta(L) Z_t$ be a stationary and invertible ARMA(p,q) process. Then the coefficients $\pi(L)$ of the AR(∞) representation, where $\pi(L) = \frac{\phi(L)}{\theta(L)}$ are given by the following recursion:

$$\pi_j = -\phi_j - \sum_{k=1}^{q} \theta_k \pi_{j-k}, \ \forall j = 0, 1, 2, \ldots, \tag{86}$$

where $\pi_i = 0$ for $i < 0$, $\phi_0 = -1, \phi_j = 0$ for $j > p$. So the ARMA(p,q)

$$\begin{aligned} \phi(L) X_t &= c + \theta(L) Z_t \\ \frac{\phi(L)}{\theta(L)} X_t &= \frac{c}{\theta(1)} + Z_t \\ \pi(L) X_t &= \frac{c}{\theta(1)} + Z_t \end{aligned}$$

becomes an AR(∞), where the π_j coefficients are determined by (86).

Start with the relation

$$\psi(L) = \frac{\theta(L)}{\phi(L)}$$
$$\phi(L)\,\psi(L) = \theta(L) \tag{87}$$

where, after we expand the polynomial terms, we get:

$$\phi(L)\,\psi(L) = \theta(L)$$
$$(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 \ldots) \times$$
$$(\psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 \ldots) = (1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3 \ldots).$$

$\phi(L)$:				Left						Right
L^0 :	$\psi_0 L^0$	$+\psi_1 L^1$	$+\psi_2 L^2$	$+\psi_3 L^3$	$+\psi_4 L^4$	$+\psi_5 L^5$	$+\psi_6 L^6$	$\ldots$		$1L^0+$
$-\phi_1 L^1$:		$-\phi_1\psi_0 L^1$	$-\phi_1\psi_1 L^2$	$-\phi_1\psi_2 L^3$	$-\phi_1\psi_3 L^4$	$-\phi_1\psi_4 L^5$	$-\phi_1\psi_5 L^6$	$\ldots$		$\theta_1 L^1+$
$-\phi_2 L^2$:			$-\phi_2\psi_0 L^2$	$-\phi_2\psi_1 L^3$	$-\phi_2\psi_2 L^4$	$-\phi_2\psi_3 L^5$	$-\phi_2\psi_4 L^6$	$\ldots$		$\theta_2 L^2+$
$-\phi_3 L^3$:				$-\phi_3\psi_0 L^3$	$-\phi_3\psi_1 L^4$	$-\phi_3\psi_2 L^5$	$-\phi_3\psi_3 L^6$	$\ldots$	$=$	$\theta_3 L^3+$
$-\phi_4 L^4$:					$-\phi_4\psi_0 L^4$	$-\phi_4\psi_1 L^5$	$-\phi_4\psi_2 L^6$	$\ldots$		$\theta_4 L^4+$
$-\phi_5 L^5$:						$-\phi_5\psi_0 L^5$	$-\phi_5\psi_1 L^6$	$\ldots$		$\theta_5 L^5+$
$-\phi_6 L^6$:							$-\phi_6\psi_0 L^6$	$\ldots$		$\theta_6 L^6+$
$\vdots$					$\vdots$					$\vdots$

To recover the coefficients of $\Psi(L)$, we need to match the coefficients of the powers in L of the columns on the left to those on the right.

This gives us the recursions

$$\begin{aligned}
\psi_0 &= 1 \quad [\text{for } L^0] \\
\psi_1 - \phi_1\psi_0 &= \theta_1 \quad [\text{for } L^1] \\
\psi_2 - \phi_1\psi_1 - \phi_2\psi_0 &= \theta_2 \quad [\text{for } L^2] \\
\psi_3 - \phi_1\psi_2 - \phi_2\psi_1 - \phi_3\psi_0 &= \theta_3 \quad [\text{for } L^3] \\
&\vdots \\
\psi_j - \sum_{k=1}^{p} \phi_k\psi_{j-k} &= \theta_j.
\end{aligned} \tag{88}$$

last line in (88) yields the recursive formula shown in (85)

Properties of ARMA(p, q) models

Recall main objective of Box-Jenkins modelling is to approximate infinite lag polynomial $\Psi(L)$ by ratio of finite (and parsimonious) polynomials $\theta(L)$ and $\phi(L)$.

These finite polynomials make up the AR and MA parts of the joint ARMA model. Let us, therefore, define the properties of ARMA(p,q) processes.

Definition (Mean of ARMA(p, q))

Let $\phi(L) X_t = c + \theta(L) Z_t$ be stationary and invertible ARMA(p, q) model, then the mean of X_t can be found from the MA(∞) represention

$$E(X_t) = \frac{c}{\phi(1)} + \frac{\theta(L)}{\phi(L)} E(Z_t). \tag{89}$$

Hence, $E(X_t) = c/(1 - \phi_1 - \phi_2 - \ldots - \phi_p)$.

Definition (Autocovariance of ARMA(p,q))

Let $\tilde{X}_t = \left(X_t - \frac{c}{\phi(1)}\right) = \frac{\theta(L)}{\phi(L)} Z_t$. Then the Autocovariance of $\tilde{X}_t$ can be easily found from the MA(∞) representation, with $\psi(L) = \frac{\theta(L)}{\phi(L)}$ and $\psi_0 = 1$ as defined in (85) and $\gamma(j)$ as defined for an MA(∞) model, that is:

$$\gamma(j) = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+j}. \quad (90)$$

Note: ARMA processes inherits

- **stationarity (stability)** property from **AR(p)** part and
- **invertibility** property from **MA(q)** part.

Order identification in ARMA models

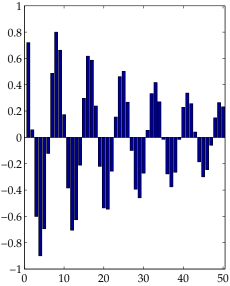
ACF and PACF cannot be used to identify the order of a stationary and invertible ARMA(p,q) model

- because the AR(p) and MA(q) coefficients get scrambled up in the ACFs and PACFs.
- no easy way to determine cut-off points in the ACFs or PACFs that would allow the lag order to be identified

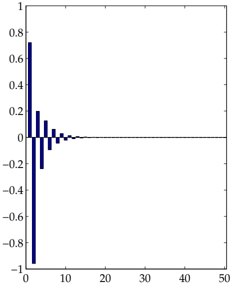
As a visual example, consider the ARMA(3,2) model given by

$$(1 - 1.3L + 0.8L^2 + 0.1L^3)X_t = (1 + 0.4L - 0.2L^2)Z_t \quad (91)$$

Corresponding plots of the ACF and PACF shown in Figure (9) below. It should be clear from this plot that there are no visible lag-order cut-off points.



(a) ACF of ARMA(3,2)



(b) PACF of ARMA(3,2)

Figure 9: Plots of the theoretical ACF and PACF of the ARMA(3,2) model given in (91).

A few different ways to see the scrambling up of the ACF/PACF algebraically.

1) Transform the ARMA(3,2) to an MA(∞) and then compute the autocovariance (ACV) as for the MA(q) model. That is, form

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= \phi_1\psi_0 + \theta_1 \\
\psi_2 &= \phi_1\psi_1 + \phi_2\psi_0 + \theta_2 \\
\psi_3 &= \phi_1\psi_2 + \phi_2\psi_1 + \phi_3\psi_0 \\
\psi_l &= \phi_1\psi_{l-1} + \phi_2\psi_{l-2} + \phi_3\psi_{l-3}, \text{ for all } l > 3
\end{aligned}$$

and then plug the obtained values into ACV formula for MA.

2) Multiply the ARMA process $\phi(L)X_t = \theta(L)Z_t$, $Z_t \sim \text{WN}(0, \sigma^2)$ by X_{t-j} and then take expectations. This yields

$$\gamma(j) = \underbrace{\phi_1\gamma(j-1) + \phi_2\gamma(j-2) + \ldots + \phi_p\gamma(j-p)}_{\text{from AR}(p)\text{ part}} + \underbrace{E(X_{t-j}Z_t)}_{=\sigma^2 \text{ if } j=0,\ 0 \text{ otherwise}} + \underbrace{\theta_1 E(X_{t-j}Z_{t-1}) + \theta_2 E(X_{t-j}Z_{t-2}) + \ldots + \theta_q E(X_{t-j}Z_{t-q})}_{\text{from MA}(q)\text{ part}}. \quad (92)$$

For the AR(p) earlier we had all the MA terms being zero, so need to evaluate $E(X_{t-j}Z_t)$ term for $j = 0, 1, 2, \ldots$

This term was equal to σ^2 for $j = 0$ and 0 for $j > 0$. Now we need to find expressions for $E(X_{t-j}Z_{t-1}), E(X_{t-j}Z_{t-2}), \ldots, E(X_{t-j}Z_{t-q})$ that also enter the relation in (92).

Let us define the following relations that will help us in determining the remaining expressions from the MA(q) part (92).

Let $E(X_K Z_L)$ be the correlation $s = (K - L)$ periods apart between X and Z. Then define

$$E(X_K Z_L) = \begin{cases} r(s) & \text{for all } s = 0, 1, 2, 3, \ldots \\ 0 & s < 0. \end{cases} \quad (93)$$

Note that the second relation in (93) follows from the fact that $E(X_t Z_{t+1})$, $E(X_t Z_{t+2}), \ldots = 0$ (why?).

What we need to find are the values for $r(s), s = 0, 1, 2, 3, \ldots$ which we can do by means of recursions.

To see this, take the ARMA(p, q) specification above and multiply the relation by Z_{t-j} and take expectations.

This yields

$$E(X_t Z_{t-j}) = \phi_1 E(X_{t-1} Z_{t-j}) + \phi_2 E(X_{t-2} Z_{t-j}) + \ldots + \phi_p E(X_{t-p} Z_{t-j}) + E\left(Z_{t-j} Z_t\right) + \theta_1 E\left(Z_{t-j} Z_{t-1}\right) + \theta_2 E\left(Z_{t-j} Z_{t-2}\right) + \ldots + \theta_q E\left(Z_{t-j} Z_{t-q}\right). \quad (94)$$

Now, evaluating this expression for $j = 0, 1, 2, 3 \ldots$ gives us:

$$\begin{aligned} r(0) &= \sigma^2 \\ r(1) &= \phi_1 r(0) + \theta_1 \sigma^2 \\ r(2) &= \phi_1 r(1) + \phi_2 r(0) + \theta_2 \sigma^2 \\ r(3) &= \phi_1 r(2) + \phi_2 r(1) + \phi_3 r(0) + \theta_3 \sigma^2 \\ r(4) &= \phi_1 r(3) + \phi_2 r(2) + \phi_3 r(1) + \phi_4 r(0) + \theta_4 \sigma^2 \end{aligned} \quad (95)$$

$$\begin{aligned} r(5) &= \underbrace{\phi_1 r(4) + \phi_2 r(3) + \phi_3 r(2) + \phi_4 r(1) + \phi_5 r(0)}_{\text{terms from AR}(p)\text{ part}} + \underbrace{\theta_5 \sigma^2}_{\text{term from MA}(q)\text{ part}} \\ &\vdots \\ r(j) &= \phi_1 r(j-1) + \phi_2 r(j-2) + \ldots + \underbrace{\phi_{j-2} r(2) + \phi_{j-1} r(1) + \phi_j r(0)}_{\text{some of these are zero if } p<j} + \underbrace{\theta_j \sigma^2}_{=0 \forall j>q.} . \end{aligned} \quad (96)$$

Since $E\left(X_{t-j} Z_{t-1}\right) = r(j-1)$, $E\left(X_{t-j} Z_{t-2}\right) = r(j-2)$, etc., we can see that the expectations in the MA(q) part in (92) can be replaced by the recursions in (95).

These will be functions of parameters of $\phi(L)$ and $\theta(L)$ lag polynomials and not of $\gamma(\cdot)$, so will be able to solve the $p + 1$ equations from $\gamma(0)$ to $\gamma(p)$ uniquely.

Then take difference equations structure of the AR(p) part to compute any higher order autocovariance, ie., for $j > p$.

Example ARMA(3,2)

Suppose we have the following ARMA($3, 2$) process:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} \tag{97}$$

In order to compute the ACV, need $r(s)$ from the recursive formulas above.

Multiply two sides of the ARMA equation in (97) by $Z_{t-j}, \forall j = 0, 1, 2, 3, \ldots$ and take expectations to yield:

$$\begin{aligned} r(0) &= \sigma^2 \\ r(1) &= \phi_1 r(0) + \theta_1 \sigma^2 &&= \sigma^2(\phi_1 + \theta_1) \\ r(2) &= \phi_1 r(1) + \phi_2 r(0) + \theta_2 \sigma^2 &&= \sigma^2(\phi_1^2 + \phi_1\theta_1 + \phi_2 + \theta_2) \\ &\vdots \\ r(j) &= \phi_1 r(j-1) + \phi_2 r(j-2) + \phi_3 r(j-3), \ j \geq 3. \end{aligned}$$

We can now multiply the two sides of the ARMA by $X_{t-j}, \forall j = 0, 1, 2, 3, \ldots$ and take expectations, which yields:

$$\begin{aligned} \gamma(0) &= \phi_1\gamma(1) + \phi_2\gamma(2) + \phi_3\gamma(3) + r(0) + \theta_1 r(1) + \theta_2 r(2) \\ \gamma(1) &= \phi_1\gamma(0) + \phi_2\gamma(1) + \phi_3\gamma(2) + \theta_1 r(0) + \theta_2 r(1) \\ \gamma(2) &= \phi_1\gamma(1) + \phi_2\gamma(0) + \phi_3\gamma(1) + \theta_2 r(0) \\ \gamma(3) &= \phi_1\gamma(2) + \phi_2\gamma(1) + \phi_3\gamma(0). \end{aligned} \tag{98}$$

We find a system of four equations with four variables in (98)

$\Rightarrow$ can be solved to obtain a unique solution for $\gamma(0)$ to $\gamma(3)$.

The other ACVs may be derived iteratively from the recursion:

$$\gamma(j) = \phi_1\gamma(j-1) + \phi_2\gamma(j-2) + \phi_3\gamma(j-3), \quad j > 3.$$

Sums of AR and MA processes

Definition (Sum of two MAs)

If the WN sequences of two MA models MA(q_1) and MA(q_2) are uncorrelated, then $MA(q_1) + MA(q_2) = MA(\max\{q_1, q_2\})$.

Definition (Sum of two ARs)

If the WN sequences of two AR models AR(p_1) and AR(p_2) are uncorrelated, then $AR(p_1) + AR(p_2) = ARMA(p_1 + p_2, \max\{p_1, p_2\})$.

References

Box, George E.P., Gwilym M. Jenkins and Gregory C. Reinsel (1994): *Time Series Analysis: Forecasting and Control*, *3rd Edition*, Prentice Hall.

Brockwell, Peter J. and Richard A. Davis (2002): *Introduction to Time Series and Forecasting*, *2nd Edition*, Springer.

Johnson, Richard A. and Dean W. Wichern (2007): *Applied Multivariate Statistical Analysis*, Pearson.

Wold, Herman O. A. (1938): *A Study in the Analysis of Stationary Time Series*, Stockholm: Almqvist & Wiksell.

Yule, George U. (1927): "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **226**, 267–298.

Exercises

1) Show that the process
$$X_t = A\cos(wt) + B\sin(wt),\ t \in \mathbb{Z}, \quad (99)$$
is stationary and find its mean and autocovariance function. Assume that A and B are uncorrelated random variables with mean 0 and variance 1 and that w is a fixed frequency in the interval $[0, \pi]$.

2) Find the ACVF of the time series
$$X_t = Z_t + 0.3Z_{t-1} - 0.4Z_{t-2}, \quad (100)$$
where $\{Z_t\} \sim \text{WN}(0, 1)$.

3) Show that the autoregressive equations
$$X_t = \phi_1 X_{t-1} + Z_t,\ t \in \mathbb{Z}, \quad (101)$$
where $\{Z_t\} \sim \text{WN}(0, 1)$ and $|\phi_1| = 1$, have no stationary solutions.

4) Let $\{Y_t\}$ be the AR(1) plus noise time series defined by
$$Y_t = X_t + W_t,$$
where $\{W_t\} \sim \text{WN}(0, \sigma_W^2)$, $\{X_t\}$ is the AR(1) process
$$X_t - \phi X_{t-1} = Z_t,\ \{Z_t\} \sim \text{WN}(0, \sigma_Z^2), \quad (102)$$
and $E[W_s Z_t] = 0$ for all s and t.

a) Show that $\{Y_t\}$ is stationary and find its autocovariance function.

b) Show that the time series $U_t = Y_t - \phi Y_{t-1}$ is 1-correlated and hence an MA(1) process.

5) Consider the ARMA$(2, 1)$ process defined by the equations

$$X_t - 0.75X_{t-1} + 0.125X_{t-2} = Z_t + 1.25Z_{t-1}, \quad \{Z_t\} \sim \mathsf{WN}(0, \sigma^2).$$

Is this process stationary and invertible?

6) Determine which of the following ARMA processes are stationary and which of them are invertible. (In each case $\{Z_t\}$ denotes white noise.)

a) $X_t + 0.2X_{t-1} - 0.48X_{t-2} = Z_t$.

b) $X_t + 1.9X_{t-1} + 0.88X_{t-2} = Z_t - 0.4Z_{t-1} + 0.04Z_{t-2}$.

c) $X_t + 0.6X_{t-1} = Z_t + 1.2Z_{t-1}$.

d) $X_t + 1.8X_{t-1} + 0.81X_{t-2} = Z_t$.

For those processes that are stationary, compute the first six coefficients $\psi_0, \psi_1, \ldots, \psi_5$ of the MA(∞) representation of $\{X_t\}$.

7) Show that the two MA(1) processes

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

and

$$Y_t = \tilde{Z}_t + \frac{1}{\theta}\tilde{Z}_{t-1}, \quad \{\tilde{Z}_t\} \sim \mathsf{WN}(0, \sigma^2\theta^2),$$

where $0 < |\theta| < 1$, have the same autocovariance function.

8) Consider the following MA(2) process

$$X_t = \mu + Z_t + \frac{7}{2}Z_{t-1} - 2Z_{t-2}, \quad \{Z_t\} \sim \mathsf{WN}(0, \sigma^2).$$

Show that this process is not invertible. Find an invertible MA(2) process with the same ACF as the process given above.

9) Find the autocovariance and autocorrelation functions of the MA(2) process

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}, \quad \{Z_t\} \sim \mathsf{IID}(0, \sigma^2).$$

Now, set $\theta_1 = -\frac{17}{10}$ and $\theta_2 = -2$. Compute the ACVF and ACF. Show that this MA(2) is not invertible.

Exercises

10) Let us consider the process

$$X_t = Z_t + bZ_{t-1} + b\rho Z_{t-2} + b\rho^2 Z_{t-2} + \dots, \quad \{Z_t\} \sim \mathsf{WN}(0, \sigma^2) \text{ and } |\rho| < 1.$$

a) Show that this MA(∞) process is stationary.

b) For which value of b is the process an AR(1)? Show that for any other finite value of b, the process is an ARMA(1,1) and identify its parameters.

11) Compute the ACF and PACF of the AR(2) process

$$X_t = 0.8X_{t-2} + Z_t, \quad \{Z_t\} \sim \mathsf{WN}(0, \sigma^2).$$

12) Show that the value at lag 2 of the partial ACF of the MA(1) process

$$X_t = Z_t + \theta Z_{t-1}, \; t \in \mathbb{Z}, \quad \{Z_t\} \sim \mathsf{WN}(0, \sigma^2),$$

is given by

$$\alpha(2) = -\frac{\theta^2}{1 + \theta^2 + \theta^4}.$$