Ch. 14 Stationary ARMA Process

A general linear stochastic model is described that suppose a time series to be generated by a linear aggregation of random shock. For practical representation it is desirable to employ models that use parameters parsimoniously. Parsimony may often be achieved by representation of the linear process in terms of a small number of autoregressive and moving average terms. This chapter introduces univariate ARMA process, which provide a very useful class of models for describing the dynamics of an individual time series. The ARMA model is based on a principle in philosophy called reductionism. The reductionism\(^1\) believe that anything can be understood once upon it is decomposed to its basic elements. Throughout this chapter we assume the time index \( T \) to be \( T = \{ \ldots, -2, -1, 0, 1, 2, \ldots \} \).

1 Moving Average Process

1.1 The First-Order Moving Average Process

A stochastic process \( \{ Y_t, \ t \in T \} \) is said to be a first order moving average process (MA(1)) if it can be expressed in the form

\[
Y_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1},
\]

where \( \mu \) and \( \theta \) are constants and \( \varepsilon_t \) is a white-noise process.

Remember that a white noise process \( \{ \varepsilon_t, \ t \in T \} \) is that

\[
E(\varepsilon_t) = 0
\]

and

\[
E(\varepsilon_t \varepsilon_s) = \begin{cases} 
\sigma^2 & \text{when } t = s \\
0 & \text{when } t \neq s 
\end{cases}
\]

1.1.1 Condition for Stationarity

The expectation of \( Y_t \) is given by

\[
E(Y_t) = E(\mu + \varepsilon_t + \theta \varepsilon_{t-1}) = \mu + E(\varepsilon_t) + \theta E(\varepsilon_{t-1}) = \mu, \quad \text{for all } t \in T.
\]

\(^1\)Thales (636-546 BC) was thought to be the first one to use reductionism in his writing.
The variance of $Y_t$ is
\[
\gamma_0 = E(Y_t - \mu)^2 = E(\varepsilon_t + \theta \varepsilon_{t-1})^2 \\
= E(\varepsilon_t^2 + 2\theta \varepsilon_t \varepsilon_{t-1} + \theta^2 \varepsilon_{t-1}^2) \\
= \sigma^2 + 0 + \theta^2 \sigma^2 \\
= (1 + \theta^2)\sigma^2.
\]

The first autocovariance is
\[
\gamma_1 = E(Y_t - \mu)(Y_{t-1} - \mu) = E(\varepsilon_t + \theta \varepsilon_{t-1})(\varepsilon_{t-1} + \theta \varepsilon_{t-2}) \\
= E(\varepsilon_t \varepsilon_{t-1} + \theta \varepsilon_t \varepsilon_{t-2} + \theta^2 \varepsilon_{t-1} \varepsilon_{t-2}) \\
= 0 + \theta \sigma^2 + 0 + 0 \\
= \theta \sigma^2.
\]

Higher autocovariances are all zero:
\[
\gamma_j = E(Y_t - \mu)(Y_{t-j} - \mu) = E(\varepsilon_t + \theta \varepsilon_{t-1})(\varepsilon_{t-j} + \theta \varepsilon_{t-j-1}) = 0 \quad \text{for } j > 1.
\]

Since the mean and the autocovariances are not functions of time, an $MA(1)$ process is weakly-stationary regardless of the value of $\theta$.

1.1.2 Conditions for Ergodicity

It is clear that the condition\(^2\)
\[
\sum_{j=0}^{\infty} |\gamma_j| = (1 + \theta^2) + |\theta \sigma^2| < \infty
\]
is satisfied. Thus the $MA(1)$ process is ergodic for any finite value of $\theta$.

1.1.3 The Dependence Structure

The $j$th autocorrelation of a weakly-stationary process is defined as its $j$th autocovariance divided by the variance
\[
r_j = \frac{\gamma_j}{\gamma_0}.
\]

By Cauchy-Schwarz inequality, we have $|r_j| \leq 1$ for all $j$.

\(^2\)See p.10 of Chapter 12.
From above results, the autocorrelation of an MA(1) process is

\[ r_j = \begin{cases} 
1 & \text{when } j = 0 \\
\frac{\theta \sigma^2}{(1 + \theta^2)\sigma^2} = \frac{\theta}{(1 + \theta^2)} & \text{when } j = 1 \\
0 & \text{when } j > 1
\end{cases} \]

The autocorrelation \( r_j \) can be plotted as a function of \( j \). This plot is usually called autocogram. See the plots of p.50 of Hamilton.

### 1.1.4 The Conditional First Two Moments of MA(1) process

Let \( F_{t-1} \) denote the information set available at time \( t-1 \). The conditional mean of \( u_t \) is

\[
E(Y_t | F_{t-1}) = E[\varepsilon_t + \theta \varepsilon_{t-1} | F_{t-1}] = \theta \varepsilon_{t-1} + E[\varepsilon_t | F_{t-1}] = \theta \varepsilon_{t-1}, \quad (\text{since } E(\varepsilon_t | F_{t-1}) = 0)
\]

and from this result, it implies that the conditional variance of \( Y_t \) is

\[
\sigma_t^2 = \text{Var}(Y_t | F_{t-1}) = E\{[Y_t - E(Y_t | F_{t-1})]^2 | F_{t-1}\} = E(\varepsilon_t^2 | F_{t-1}) = \sigma^2.
\]

While the conditional mean of \( Y_t \) depends upon the information at \( t-1 \), however, the conditional variance does not. Engle (1982) propose a class of models where the variance does depend upon the past and argue for their usefulness in economics. See Chapter 26.

### 1.2 The \( q \)-th Order Moving Average Process

A stochastic process \( \{Y_t, \ t \in T\} \) is said to be a moving average process of order \( q \) (MA(\( q \))) if it can be expressed in this form

\[
Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}
\]

where \( \mu, \theta_0, \theta_1, \theta_2, ..., \theta_q \) are constants with \( \theta_0 = 1 \) and \( \varepsilon_t \) is a white-noise process.
1.2.1 Conditions for Stationarity

The expectation of $Y_t$ is given by

$$E(Y_t) = E(\mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q})$$

$$= \mu + E(\varepsilon_t) + \theta_1 E(\varepsilon_{t-1}) + \theta_2 E(\varepsilon_{t-2}) + \ldots + \theta_q E(\varepsilon_{t-q}) = \mu, \quad \text{for all } t \in T.$$ 

The variance of $Y_t$ is

$$\gamma_0 = E(Y_t - \mu)^2 = E(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q})^2.$$ 

Since $\varepsilon_t$'s are uncorrelated, the variance of $Y_t$ is

$$\gamma_0 = \sigma^2 + \theta_1^2 \sigma^2 + \theta_2^2 \sigma^2 + \ldots + \theta_q^2 \sigma^2 = (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)\sigma^2.$$ 

For $j = 1, 2, \ldots, q$,

$$\gamma_j = E[(Y_t - \mu)(Y_{t-j} - \mu)]$$

$$= E[(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}) \times (\varepsilon_{t-j} + \theta_1 \varepsilon_{t-j-1} + \theta_2 \varepsilon_{t-j-2} + \ldots + \theta_q \varepsilon_{t-j-q})]$$

$$= E[\theta_j \varepsilon_{t-j}^2 + \theta_{j+1} \theta_1 \varepsilon_{t-j-1}^2 + \theta_{j+2} \theta_2 \varepsilon_{t-j-2}^2 + \ldots + \theta_q \theta_{q-j} \varepsilon_{t-j-q}^2].$$ 

Terms involving $\varepsilon$’s at different dates have been dropped because their product has expectation zero, and $\theta_0$ is defined to be unity. For $j > q$, there are no $\varepsilon$’s with common dates in the definition of $\gamma_j$, and so the expectation is zero. Thus,

$$\gamma_j = \left\{ \begin{array}{ll}
\theta_j + \theta_{j+1} \theta_1 + \theta_{j+2} \theta_2 + \ldots + \theta_q \theta_{q-j} \sigma^2 & \text{for } j = 1, 2, \ldots, q \\
0 & \text{for } j > q
\end{array} \right.$$ 

Since the mean and the autocovariances are not functions of time, an MA($q$) process is weakly-stationary regardless of the value of $\theta_i, i = 1, 2, \ldots, q$.

For example, for an MA(2) process,

$$\gamma_0 = [1 + \theta_1^2 + \theta_2^2] \sigma^2$$

$$\gamma_1 = [\theta_1 + \theta_2 \theta_1] \sigma^2$$

$$\gamma_2 = [\theta_2] \sigma^2$$

$$\gamma_3 = \gamma_4 = \ldots = 0$$
1.2.2 Conditions for Ergodicity

It is clear that the condition
\[ \sum_{j=0}^{\infty} |\gamma_j| < \infty \]
is satisfied. Thus the $MA(q)$ process is ergodic for any finite value of $\theta_i, i = 1, 2, ..., q$.

1.2.3 The Dependence Structure

The autocorrelation function is zero after $q$ lags. See the plots of p.50.

1.3 The Infinite-Order Moving Average Process

A stochastic process $\{Y_t, t \in T\}$ is said to be an infinite-order moving average process ($MA(\infty)$) if it can be expressed in this form
\[ Y_t = \mu + \sum_{j=0}^{\infty} \varphi_j \varepsilon_{t-j} = \mu + \varphi_0 \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + ..... \tag{1} \]
where $\mu, \varphi_0, \varphi_1, \varphi_2, ...$, are constants with $\varphi_0 = 1$ and $\varepsilon_t$ is a white-noise process.

1.3.1 Convergence of Infinite Series

Before we discuss the statistical properties of a $MA(\infty)$ process, we need an understanding of the theory of convergence of an infinite series.

Let $\{a_j\}$ be a sequence of numbers. Then the formal sum
\[ a_0 + a_1 + a_2 + \cdots + a_n + \cdots \]

\[ \varepsilon_{t-j} = \beta_0 u_{t-j}, \]
\[ \varphi_j = \beta_j / \beta_0, \quad j = 1, 2, \cdots, \]
and obtain the representation
\[ Y_t = \sum_{j=0}^{\infty} \varphi_j \varepsilon_{t-j}, \]
where $\varphi_0 = 1$ and $\varepsilon_t$ is uncorrelated $(0, \beta_0^2 \sigma_u^2)$ random variables.
or

\[ \sum_{j=0}^{\infty} a_j \]

is called an infinite series. The number \( a_0, a_1, \ldots, a_n, \ldots \) are its terms, and the numbers \( S_n \equiv \sum_{j=0}^{n} a_j \) its partial sums.

If \( \lim S_n \) exists, its value \( S \) is called the sum of the series. In this case, we say that the series converges and we write

\[ S \equiv \sum_{j=0}^{\infty} a_j < \infty. \]

If \( \lim S_n \) does not exists, we say that the series diverges.

**Theorem:**

Suppose that \( \sum_{j=0}^{\infty} a_j \) converges. Then \( \lim a_j = 0 \).

**Proof:**

Note first that, if \( \lim S_n = S \), then \( \lim S_{n-1} = S \). Now \( a_j = S_j - S_{j-1} \), so

\[ \lim a_j = \lim (S_j - S_{j-1}) = \lim S_j - \lim S_{j-1} = S - S = 0. \]

This does not say that, if \( \lim a_j = 0 \), then \( \sum a_j \) converges. Indeed this is not correct. It says that convergence of \( \sum a_j \) implies \( a_j \to 0 \). Hence if \( a_j \) does not tend to zero, the series cannot converge. Thus, in a given series \( \sum a_j \), we can examine \( \lim a_j \). If \( \lim a_j = 0 \), we have no information about convergence of divergence; but if \( \lim a_j \neq 0 \), either because it fail to exist or because it exists and has another value, then \( \sum a_j \) diverges.

**Theorem** (Cauchy Criterion).\(^4\)

A necessary and sufficient condition that a series \( \sum a_j \) converges is that, for each \( \zeta > 0 \), there exist an \( N(\zeta) \) for which\(^5\)

\[ |a_{j+1} + a_{j+2} + \cdots + a_m| < \zeta \quad \text{if } m > j > N. \]

\(^4\)The Cauchy Criterion is an assertion about the behavior of the terms of a sequence. It says that far out in the sequence all of them are close to each other.

\(^5\)This implies that \( \lim(a_{j+1} + a_{j+2} + \cdots + a_m) = 0 = \lim(a_{j+1}) + \lim(a_{j+2}) + \cdots + \lim(a_m) = 0 \), which is satisfied from the theorem above that if \( \sum a_j \) converges, then \( \lim a_j = 0 \).
Definition:
A sequence \( \{a_j\} \) is said to be **square-summable** if
\[
\sum_{j=0}^{\infty} a_j^2 < \infty,
\]
whereas a sequence \( \{a_j\} \) is said to be **absolute-summable** if
\[
\sum_{j=0}^{\infty} |a_j| < \infty.
\]

**Proposition:**
Absolute summability implies square-summability, but the converse **does not** hold.

**Proof:**
First we show that absolute summability implies square-summability. Suppose that \( \{a_j\}_{j=0}^{\infty} \) is absolutely summable. Then there exists an \( N < \infty \) such that \( |a_j| < 1 \) for all \( j \geq N, \)
6 implying that \( a_j^2 < |a_j| \) for all \( j \geq N \). Then
\[
\sum_{j=0}^{\infty} a_j^2 = \sum_{j=0}^{N-1} a_j^2 + \sum_{j=N}^{\infty} a_j^2 < \sum_{j=0}^{N-1} a_j^2 + \sum_{j=N}^{\infty} |a_j|.
\]
But \( \sum_{j=0}^{N-1} a_j^2 \) is finite, since \( N \) is finite, and \( \sum_{j=N}^{\infty} |a_j| \) is finite, since \( \{a_j\} \) is absolutely summable. Hence \( \sum_{j=0}^{\infty} a_j^2 < \infty \). It can verified that the converse is not true by considering \( \sum_{j=1}^{\infty} j^{-2} \).

**Proposition:**
Given two absolutely summable sequences \( \{a_j\} \) and \( \{b_j\} \), then the sequence \( \{a_j + b_j\} \) and \( \{a_jb_j\} \) are absolutely summable. It is also apparent that \( \sum |a_j| + \sum |b_j| < \infty \).

**Proof:**
\[
\sum_{j=0}^{\infty} |a_j + b_j| \leq \sum_{j=0}^{\infty} (|a_j| + |b_j|) = \sum_{j=0}^{\infty} |a_j| + \sum_{j=0}^{\infty} |b_j| < \infty, \quad (2)
\]
\[
\sum_{j=0}^{\infty} |a_jb_j| = \sum_{j=0}^{\infty} (|a_j||b_j|) \leq \sum_{j=0}^{\infty} (|a_j| + |b_j|)^2 < \infty. \quad (3)
\]

6Since by assumption \( \{a_j\} \) is absolute summable, then \( |a_j| \to 0 \).
Of course, $\sum_{j=0}^{\infty} (|a_j| + |b_j|)^2 < \infty$ since $(|a_j| + |b_j|)^2$ is the square of an absolutely summable sequence.

**Proposition:**
The *convolution* of two absolutely summable sequence $\{a_j\}$ and $\{b_j\}$ defined by

$$c_j = \sum_{k=0}^{\infty} a_k b_{j+k},$$

is absolutely summable.

**Proof:**

$$\sum_{j=0}^{\infty} |c_j| \leq \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} |a_k||b_{j+k}| = \sum_{k=0}^{\infty} |a_k| \sum_{s=0}^{\infty} |b_s| < \infty. \quad (4)$$

**Proposition:**
If $\sum |a_j|$ converges, so does $\sum a_j$.

### 1.3.2 Is This a Well Defined Random Sequence?

**Proposition:**
If the coefficients of the $MA(\infty)$ in (1) is square-summable, then $\sum_{j=0}^{T} \varphi_j \varepsilon_{t-j}$ converges in mean square to some random variable $Z_t$ (Say) as $T \to \infty$.

**Proof:**
The *Cauchy* criterion states that $\sum_{j=0}^{T} \varphi_j \varepsilon_{t-j}$ converges in mean square to some random variable $Z_t$ as $T \to \infty$ if and only if, for any $\varsigma > 0$, there exists a suitably large $N$ such that for any integer $M > N$

$$E \left[ \sum_{j=0}^{M} \varphi_j \varepsilon_{t-j} - \sum_{j=0}^{N} \varphi_j \varepsilon_{t-j} \right]^2 < \varsigma. \quad (5)$$

Since here we require that $Z_t$ being a covariance-stationary process, we use the convergence in mean square error of Cauchy criterion which guarantee the existence of second moments of $Y_t$. 

---

8 Copy Right by Chingnun Lee © 2006
In words, once $N$ terms have been summed, the difference between that sum and the one obtained from summing to $M$ is a random variable whose mean and variance are both arbitrarily close to zero.

Now the left hand side of (5) is simply

$$E \left[ \varphi_M \varepsilon_{t-M} + \varphi_{M-1} \varepsilon_{t-M+1} + \ldots + \varphi_{N+1} \varepsilon_{t-N-1} \right]^2$$

$$= (\varphi_M^2 + \varphi_{M-1}^2 + \ldots + \varphi_{N+1}^2) \sigma^2$$

$$= \left[ \sum_{j=0}^M \varphi_j^2 - \sum_{j=0}^N \varphi_j^2 \right] \sigma^2. \quad (6)$$

But if $\sum_{j=0}^\infty \varphi_j^2 < \infty$, then by the Cauchy criterion the right side of (6) may be made as small as desired by a suitable large $N$. Thus the $MA(\infty)$ is well defined sequence since the infinity series $\sum_{j=0}^\infty \varphi_j \varepsilon_{t-j}$ converges in mean squares. So the $MA(\infty)$ process $Y_t(= \mu + Z_t)$ is a well defined random variable with finite second moment.

### 1.3.3 Check Stationarity

Assume the $MA(\infty)$ process to be with absolutely summable coefficients.

The expectation of $Y_t$ is given by

$$E(Y_t) = \lim_{T \to \infty} E(\mu + \varphi_0 \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \ldots + \varphi_T \varepsilon_{t-T})$$

$$= \mu$$

The variance of $Y_t$ is

$$\gamma_0 = E(Y_t - \mu)^2$$

$$= \lim_{T \to \infty} E(\varphi_0 \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \ldots + \varphi_T \varepsilon_{t-T})^2$$

$$= \lim_{T \to \infty} (\varphi_0^2 + \varphi_1^2 + \varphi_2^2 + \ldots + \varphi_T^2) \sigma^2$$

$$< \infty. \ (From \ the \ assumption \ of \ absolutely \ summable \ coefficients)$$
For $j > 0$,
\[\gamma_j = E(Y_t - \mu)(Y_{t-j} - \mu)\]
\[= (\varphi_j\varphi_0 + \varphi_{j+1}\varphi_1 + \varphi_{j+2}\varphi_2 + \varphi_{j+3}\varphi_3 + \ldots)\sigma^2\]
\[= \sigma^2 \sum_{k=0}^{\infty} \varphi_{j+k}\varphi_k\]
\[\leq \sigma^2 \sum_{k=0}^{\infty} |\varphi_{j+k}\varphi_k|\]
\[< \infty. \quad \text{(from (3))}\]

Thus, $E(Y_t)$ and $\gamma_j$ are both finite and independent of $t$. The $MA(\infty)$ process with absolute-summable coefficients is weakly-stationary.

### 1.3.4 Check Ergodicity

**Proposition:**
The absolute summability of the moving average coefficients implies that the process is ergodic.

**Proof:**
From the results of (4) or recall the autocovariance of an $MA(\infty)$ is
\[\gamma_j = \sigma^2 \sum_{k=0}^{\infty} \varphi_{j+k}\varphi_k.\]
Then
\[|\gamma_j| = \sigma^2 \left| \sum_{k=0}^{\infty} \varphi_{j+k}\varphi_k \right|\]
\[\leq \sigma^2 \sum_{k=0}^{\infty} |\varphi_{j+k}\varphi_k|,\]
and
\[\sum_{j=0}^{\infty} |\gamma_j| \leq \sigma^2 \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} |\varphi_{j+k}\varphi_k|\]
\[= \sigma^2 \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} |\varphi_{j+k}| |\varphi_k|\]
\[= \sigma^2 \sum_{k=0}^{\infty} |\varphi_k| \sum_{j=0}^{\infty} |\varphi_{j+k}|.\]
But there exists an $M < \infty$ such that $\sum_{j=0}^{\infty} |\varphi_j| < M$, and therefore $\sum_{j=0}^{\infty} |\varphi_{j+k}| < M$ for $k = 0, 1, 2, \ldots$, meaning that

$$\sum_{j=0}^{\infty} |\gamma_j| < \sigma^2 \sum_{k=0}^{\infty} |\varphi_k| M < \sigma^2 M^2 < \infty.$$ 

Hence, the $MA(\infty)$ process with absolute-summable coefficients is ergodic.
2 Autoregressive Process

2.1 The First-Order Autoregressive Process

A stochastic process \{Y_t, \ t \in T\} is said to be a first order autoregressive process (\textit{AR}(1)) if it can be expressed in the form

\[ Y_t = c + \phi Y_{t-1} + \varepsilon_t, \]

where \(c\) and \(\phi\) are constants and \(\varepsilon_t\) is a white-noise process.

2.1.1 Check Stationarity and Ergodicity

Write the \textit{AR}(1) process in lag operator form:

\[ Y_t = c + \phi L Y_t + \varepsilon_t, \]

then

\[ (1 - \phi L)Y_t = c + \varepsilon_t. \]

In the case \(|\phi| < 1\), we know from the properties of lag operator in last chapter that

\[ (1 - \phi L)^{-1} = 1 + \phi L + \phi^2 L^2 + ...., \]

thus

\[ Y_t = (c + \varepsilon_t) \cdot (1 + \phi L + \phi^2 L^2 + ....) \]
\[ = (c + \phi c + \phi^2 c + ...) + (\varepsilon_t + \phi \varepsilon_t + \phi^2 \varepsilon_t + ...) \]
\[ = (c + \phi c + \phi^2 c + ...) + (\varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + ...) \]
\[ = \frac{c}{1 - \phi} + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + ... \]

This can be viewed as an \textit{MA}(\infty) process with \(\varphi_j\) given by \(\phi^j\). When \(|\phi| < 1\), this \textit{AR}(1) is an \textit{MA}(\infty) with absolute summable coefficient:

\[ \sum_{j=0}^{\infty} |\varphi_j| = \sum_{j=0}^{\infty} |\phi|^j = \frac{1}{1 - |\phi|} < \infty. \]

Therefore, the \textit{AR}(1) process is stationary and ergodic provided that \(|\phi| < 1\).
2.1.2 The Dependence Structure

The expectation of \( Y_t \) is given by\(^8\)

\[
E(Y_t) = E\left( \frac{c}{1 - \phi} + \varepsilon_t + \phi^1 \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \ldots \right)
\]
\[
= \frac{c}{1 - \phi} = \mu.
\]

The variance of \( Y_t \) is

\[
\gamma_0 = E(Y_t - \mu)^2
= E(\varepsilon_t + \phi^1 \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \ldots)^2
= (1 + \phi^2 + \phi^4 + \ldots) \sigma^2
= \left( \frac{1}{1 - \phi^2} \right) \sigma^2.
\]

For \( j > 0 \),

\[
\gamma_j = E(Y_t - \mu)(Y_{t-j} - \mu)
= E(\varepsilon_t + \phi^1 \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \ldots + \phi^j \varepsilon_{t-j} + \phi^{j+1} \varepsilon_{t-j-1} + \phi^{j+2} \varepsilon_{t-j-2} + \ldots) \times (\varepsilon_{t-j} + \phi^1 \varepsilon_{t-j-1} + \phi^2 \varepsilon_{t-j-2} + \ldots)
= (\phi^j + \phi^{j+2} \phi^{j+4} + \ldots) \sigma^2
= \phi^j (1 + \phi^2 + \phi^4 + \ldots) \sigma^2
= \left( \frac{\phi^j}{1 - \phi^2} \right) \sigma^2
= \phi \gamma_{j-1}.
\]

It follows that the autocorrelation function

\[
r_j = \frac{\gamma_j}{\gamma_0} = \phi^j,
\]

which follows a pattern of geometric decay as the plot on p.50.

\(^8\)Therefore, it is noted that while \( \mu \) is the mean of a \( MA \) process, the constant \( c \) is not the mean of a \( AR \) process.
2.1.3 An Alternative Way to Calculate the Moments of a Stationary AR(1) Process

Assume that the AR(1) process under consideration is weakly-stationary, then taking expectation on both side we have

\[ E(Y_t) = c + \phi E(Y_{t-1}) + E(\varepsilon_t). \]

Since by assumption that the process is stationary,

\[ E(Y_t) = E(Y_{t-1}) = \mu. \]

Therefore,

\[ \mu = c + \phi \mu + 0 \]

or

\[ \mu = \frac{c}{1 - \phi}, \]

reproducing the earlier result.

To find a higher moments of \( Y_t \) in an analogous manner, we rewrite this AR(1) as

\[ Y_t = \mu(1 - \phi) + \phi Y_{t-1} + \varepsilon_t \]

or

\[ (Y_t - \mu) = \phi(Y_{t-1} - \mu) + \varepsilon_t. \]  \hspace{1cm} (7)

For \( j \geq 0 \), multiply \((Y_{t-j} - \mu)\) on both side of (7) and take expectation:

\[
\gamma_j = E[(Y_t - \mu)(Y_{t-j} - \mu)]
= \phi E[(Y_{t-1} - \mu)(Y_{t-j} - \mu)] + E(Y_{t-j} - \mu)\varepsilon_t
= \phi \gamma_{j-1} + E(Y_{t-j} - \mu)\varepsilon_t.
\]

Next we consider the term \( E(Y_{t-j} - \mu)\varepsilon_t \). When \( j = 0 \), multiply \( \varepsilon_t \) on both side of (7) and take expectation:

\[ E(Y_t - \mu)\varepsilon_t = E[\phi(Y_{t-1} - \mu)\varepsilon_t] + E(\varepsilon_t^2). \]
Recall that $Y_{t-1} - \mu$ is a linear function of $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$:

$$Y_{t-1} - \mu = \varepsilon_{t-1} + \phi \varepsilon_{t-2} + \phi^2 \varepsilon_{t-3} + \ldots$$

we have

$$E[\phi(Y_{t-1} - \mu)\varepsilon_t] = 0.$$ 

Therefore,

$$E(Y_t - \mu)\varepsilon_t = E(\varepsilon_t^2) = \sigma^2;$$

and when $j > 0$, it is obvious that $E(Y_{t-j} - \mu)\varepsilon_t = 0$.

Therefore we the results that

$$\gamma_0 = \phi \gamma_1 + \sigma^2, \text{ for } j = 0$$

$$\gamma_1 = \phi \gamma_0, \text{ for } j = 1 \text{ and}$$

$$\gamma_j = \phi \gamma_{j-1}, \text{ for } j > 1.$$ 

That is

$$\gamma_0 = \phi \phi \gamma_0 + \sigma^2 = \frac{\sigma^2}{1 - \phi^2}.$$ 

Beside $\gamma_0$, we need first moment ($\gamma_1$) to solve $\gamma_0$.

### 2.2 The Second-Order Autoregressive Process

A stochastic process $\{Y_t, t \in T\}$ is said to be a **second order autoregressive process** ($AR(2)$) if it can be expressed in the form

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

where $c$, $\phi_1$ and $\phi_2$ are constants and $\varepsilon_t$ is a white-noise process.

#### 2.2.1 Check Stationarity and Ergodicity

Write the $AR(2)$ process in lag operator form:

$$Y_t = c + \phi_1 LY_t + \phi_2 L^2 Y_t + \varepsilon_t,$$
then
\[(1 - \phi_1 L - \phi_2 L^2)Y_t = c + \varepsilon_t.\]

In the case that all the roots of the polynomial \((1 - \phi_1 L - \phi_2 L^2) = 0\) lies outside the unit circle, we know from the properties of lag operator in last chapter that there exist a polynomial \(\varphi(L)\) such that
\[
\varphi(L) = (1 - \phi_1 L - \phi_2 L^2)^{-1} = \varphi_0 + \varphi_1 L + \varphi_2 L^2 + ....,
\]
with
\[
\sum_{j=0}^{\infty} |\varphi_j| < \infty.
\]

**Proof:** \(\varphi_j\) here is equal to \(c_1 \lambda_1^j + c_2 \lambda_2^j\), where \(c_1 + c_2 = 1\) and \(\lambda_1, \lambda_2\) are the reciprocal of the roots of the polynomial \((1 - \phi_1 L - \phi_2 L^2) = 0\). Therefore, \(\lambda_1\) and \(\lambda_2\) lie inside the unit circle. See Hamilton, p. 33, [2.3.23].

\[
\sum_{j=0}^{\infty} |\varphi_j| = \sum_{j=0}^{\infty} |c_1 \lambda_1^j + c_2 \lambda_2^j| \\
\leq \sum_{j=0}^{\infty} |c_1 \lambda_1^j| + \sum_{j=0}^{\infty} |c_2 \lambda_2^j| \\
\leq |c_1| \sum_{j=0}^{\infty} |\lambda_1^j| + |c_2| \sum_{j=0}^{\infty} |\lambda_2^j| \\
< \infty.
\]

Thus
\[
Y_t = (c + \varepsilon_t) \cdot (1 + \varphi_1 L + \varphi_2 L^2 + ....) \\
= (c + \varphi_1 L \varepsilon_t \varphi_2 L^2 \varepsilon_t + ...) + (\varepsilon_t + \varphi_1 \varepsilon_t \varepsilon_t + \varphi_2 L^2 \varepsilon_t + ...) \\
= (c + \varphi_1 \varepsilon_t + \varphi_2 L^2 \varepsilon_t + ...) + (\varepsilon_t + \varphi_1 \varepsilon_t + \varphi_2 \varepsilon_t - + ...) \\
= c(1 + \varphi_1 + \varphi_2 + ...) + \varepsilon_t + \varphi_1 \varepsilon_t + \varphi_2 \varepsilon_t - + ...
\]
where the constant term is from the fact that substituting 1 into the identity
\[
(1 - \phi_1 L - \phi_2 L^2)^{-1} = 1 + \varphi_1 L + \varphi_2 L^2 + ....
\]
This can be viewed as an $MA(\infty)$ process with absolute summable coefficient. Therefore, the $AR(2)$ process is stationary and ergodic provided that all the roots of $(1 - \phi_1 L - \phi_2 L^2) = 0$ lies outside the unit circle.

### 2.2.2 The Dependence Structure

Assume that the $AR(2)$ process under consideration is weakly-stationary, then taking expectation on both side we have

$$E(Y_t) = c + \phi_1 E(Y_{t-1}) + \phi_2 E(Y_{t-2}) + E(\varepsilon_t).$$

Since by assumption that the process is stationary,

$$E(Y_t) = E(Y_{t-1}) = E(Y_{t-2}) = \mu.$$

Therefore,

$$\mu = c + \phi_1 \mu + \phi_2 \mu + 0$$

or

$$\mu = \frac{c}{1 - \phi_1 - \phi_2}.$$

To find the higher moment of $Y_t$ in an analogous manner, we rewrite this $AR(2)$ as

$$Y_t = \mu(1 - \phi_1 - \phi_2) + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

or

$$(Y_t - \mu) = \phi_1 (Y_{t-1} - \mu) + \phi_2 (Y_{t-2} - \mu) + \varepsilon_t. \quad (8)$$

For $j \geq 0$, multiply $(Y_{t-j} - \mu)$ on both side of (8) and take expectation:

$$\gamma_j = E[(Y_t - \mu)(Y_{t-j} - \mu)]$$

$$= \phi_1 E[(Y_{t-1} - \mu)(Y_{t-j} - \mu)] + \phi_2 E[(Y_{t-2} - \mu)(Y_{t-j} - \mu)] + E(Y_{t-j} - \mu)\varepsilon_t$$

$$= \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + E(Y_{t-j} - \mu)\varepsilon_t.$$
Next we consider the term \( E(Y_{t-j} - \mu) \varepsilon_t \).
When \( j = 0 \), multiply \( \varepsilon_t \) on both side of (8) and take expectation:

\[
E(Y_t - \mu) \varepsilon_t = E[\phi_1(Y_{t-1} - \mu) \varepsilon_t] + E[\phi_2(Y_{t-2} - \mu) \varepsilon_t] + E(\varepsilon_t^2).
\]

Recall that \( Y_{t-1} - \mu \) is a linear function of \( \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots \); we have

\[
E[\phi_1(Y_{t-1} - \mu) \varepsilon_t] = 0
\]
and obviously \( E[\phi_2(Y_{t-2} - \mu) \varepsilon_t] = 0 \) also.

Therefore,

\[
E(Y_t - \mu) \varepsilon_t = E(\varepsilon_t^2) = \sigma^2,
\]
and when \( j > 0 \), it is obvious that \( E(Y_{t-j} - \mu) \varepsilon_t = 0 \).

Therefore we the results that

\[
\begin{align*}
\gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2, \text{ for } j = 0; \\
\gamma_1 &= \phi_1 \gamma_0 + \phi_2 \gamma_1, \text{ for } j = 1; \\
\gamma_2 &= \phi_1 \gamma_1 + \phi_2 \gamma_0, \text{ for } j = 2, \text{ and} \\
\gamma_j &= \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2}, \text{ for } j > 2.
\end{align*}
\]

That is

\[
\begin{align*}
\gamma_1 &= \frac{\phi_1}{1 - \phi_2} \gamma_0, \quad (9) \\
\gamma_2 &= \frac{\phi_2^2}{1 - \phi_2} \gamma_0 + \frac{\phi_2}{1 - \phi_2} \gamma_0 \\
\gamma_0 &= \left[ \frac{\phi_1^2}{1 - \phi_2} + \frac{\phi_2^2}{1 - \phi_2} + \phi_2^2 \right] + \sigma^2 \quad (10)
\end{align*}
\]

and therefore

\[
\gamma_0 = \left[ \frac{\phi_1^2}{1 - \phi_2} + \frac{\phi_2 \phi_1^2}{1 - \phi_2} + \phi_2^2 \right] + \sigma^2
\]
or

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

Substituting this result to (9) and (10), we obtains \( \gamma_1 \) and \( \gamma_2 \). Beside \( \gamma_0 \), we need first two moments (\( \gamma_1 \) and \( \gamma_2 \)) to solve \( \gamma_0 \).
2.3 The $p$th-Order Autoregressive Process

A stochastic process $\{Y_t, \ t \in T\}$ is said to be a $p-$th order autoregressive process ($AR(p)$) if it can be expressed in the form

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t,$$

where $c, \phi_1, \phi_2, \ldots$, and $\phi_p$ are constants and $\varepsilon_t$ is a white-noise process.

2.3.1 Check Stationarity and Ergodicity

Write the $AR(p)$ process in lag operator form:

$$Y_t = c + \phi_1 LY_t + \phi_2 L^2 Y_t + \ldots + \phi_p L^p Y_t + \varepsilon_t,$$

then

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)Y_t = c + \varepsilon_t.$$

In the case all the roots of the polynomial $(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) = 0$ lies outside the unit circle, we know from the properties of lag operator in last chapter that there exist a polynomial $\varphi(L)$ such that

$$\varphi(L) = (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)^{-1} = \varphi_0 + \varphi_1 L + \varphi_2 L^2 + \ldots,$$

with

$$\sum_{j=0}^{\infty} |\varphi_j| < \infty.$$

Thus

$$Y_t = (c + \varepsilon_t) \cdot (1 + \varphi_1 L + \varphi_2 L^2 + \ldots)$$

$$= (c + \varphi_1 c + \varphi_2 c + \ldots) + (\varepsilon_t + \varphi_1 \varepsilon_t + \varphi_2 \varepsilon_t + \ldots)$$

$$= c(1 + \varphi_1 + \varphi_2^2 + \ldots) + (\varepsilon_t + \varphi_1 \varepsilon_t - 1 + \varphi_2 \varepsilon_{t-2} + \ldots)$$

$$= \frac{c}{1 - \phi_1 - \phi_2 - \ldots - \phi_p} + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \ldots,$$

where the constant term is from the fact that substituting 1 into the identity

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)^{-1} = 1 + \varphi_1 L + \varphi_2 L^2 + \ldots.$$

This can be viewed as an $MA(\infty)$ process with absolute summable coefficient. Therefore, the $AR(p)$ process is stationary and ergodic provided that all the roots of $(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) = 0$ lies outside the unit circle.
2.3.2 The Dependence Structure

Assume that the \( AR(p) \) process under consideration is weakly-stationary, then taking expectation on both side we have

\[
E(Y_t) = c + \phi_1 E(Y_{t-1}) + \phi_2 E(Y_{t-2}) + ... + \phi_p E(Y_{t-p}) + E(\varepsilon_t).
\]

Since by assumption that the process is stationary,

\[
E(Y_t) = E(Y_{t-1}) = E(Y_{t-2}) = ... = E(Y_{t-p}) = \mu.
\]

Therefore,

\[
\mu = c + \phi_1 \mu + \phi_2 \mu + ... + \phi_p + 0
\]
or

\[
\mu = \frac{c}{1 - \phi_1 - \phi_2 - ... - \phi_p}.
\]

To find the higher moment of \( Y_t \) in an analogous manner, we rewrite this \( AR(p) \) as

\[
Y_t = \mu(1 - \phi_1 - \phi_2 - ... - \phi_p) + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t
\]
or

\[
(Y_t - \mu) = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + ... + \phi_p(Y_{t-p} - \mu) + \varepsilon_t. \tag{11}
\]

For \( j \geq 0 \), multiply \( (Y_{t-j} - \mu) \) on both side of (11) and take expectation:

\[
\gamma_j \quad = \quad E[(Y_t - \mu)(Y_{t-j} - \mu)]
\]
\[
\quad = \quad \phi_1 E[(Y_{t-1} - \mu)(Y_{t-j} - \mu)] + \phi_2 E[(Y_{t-2} - \mu)(Y_{t-j} - \mu)] + ...
\]
\[
\quad + \phi_p E[(Y_{t-p} - \mu)(Y_{t-j} - \mu)] + E(Y_{t-j} - \mu)\varepsilon_t
\]
\[
\quad = \quad \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + ... + \phi_p \gamma_{j-p} + E(Y_{t-j} - \mu)\varepsilon_t.
\]

Next we consider the term \( E(Y_{t-j} - \mu)\varepsilon_t \).

When \( j = 0 \), multiply \( \varepsilon_t \) on both side of (11) and take expectation:

\[
E(Y_t - \mu)\varepsilon_t = E[\phi_1(Y_{t-1} - \mu)\varepsilon_t] + E[\phi_2(Y_{t-2} - \mu)\varepsilon_t] + ... + E[\phi_p(Y_{t-p} - \mu)\varepsilon_t] + E(\varepsilon_t^2).
\]
Recall that $Y_{t-1} - \mu$ is a linear function of $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$, we have

$$E[\phi_1(Y_{t-1} - \mu)\varepsilon_t] = 0$$

and obviously, $E[\phi_i(Y_{t-i} - \mu)\varepsilon_t] = 0$, $i = 2, \ldots, p$ also.

Therefore,

$$E(Y_t - \mu)\varepsilon_t = E(\varepsilon_t^2) = \sigma^2,$$

and when $j > 0$, it is obvious that $E(Y_{t-j} - \mu)\varepsilon_t = 0$.

Therefore we the results that

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \ldots + \phi_p\gamma_p + \sigma^2, \text{ for } j = 0;$$

and

$$\gamma_j = \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \ldots + \phi_p\gamma_{j-p}, \text{ for } j = 1, 2, \ldots$$

(12)

Divided (12) by $\gamma_0$ produce the Yule–Walker equation:

$$r_j = \phi_1r_{j-1} + \phi_2r_{j-2} + \ldots + \phi_p r_{j-p}, \text{ for } j = 1, 2, \ldots$$

(13)

Intuitively, beside $\gamma_0$, we need first $p$ moments ($\gamma_1, \gamma_2, \ldots, \gamma_p$) to solve $\gamma_0$.

**Exercise:**

Use the same realization of $\varepsilon$’s to simulate and plot the following Gaussian process $Y_t$ (set $\sigma^2 = 1$) in a sample of size $T=500$:

1. $Y_t = \varepsilon_t$,
2. $Y_t = \varepsilon_t + 0.8\varepsilon_{t-1}$,
3. $Y_t = \varepsilon_t - 0.8\varepsilon_{t-1}$,
4. $Y_t = 0.8Y_{t-1} + \varepsilon_t$,
5. $Y_t = -0.8Y_{t-1} + \varepsilon_t$,
6. $\tilde{Y}_t = \tilde{\varepsilon}_t + 1.25\tilde{\varepsilon}_{t-1}$, where $Var(\tilde{\varepsilon}_t) = 0.64$.

Finally, please also plot their sample and population autocograms.
3 Mixed Autoregressive Moving Average Process

The dependence structure described by a $MA(q)$ process is truncated after the first $q$ period, meanwhile it is geometrically decaying in an $AR(p)$ process, depending on it $AR$ coefficients. A richer flexibility in the dependence structure in the first few lags model is called for to meet the real phenomena. An $ARMA(p,q)$ model meets this requirement.

A stochastic process $\{Y_t, t \in T\}$ is said to be a **autoregressive moving average process of order $(p,q)$ ($ARMA(p,q)$)** if it can be expressed in the form

$$
Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q},
$$

where $c, \phi_1, \phi_2, \ldots, \phi_p$, and $\theta_1, \ldots, \theta_q$ are constants and $\epsilon_t$ is a white-noise process.

### 3.1 Check Stationarity and Ergodicity

Write the $ARMA(p,q)$ process is lag operator form:

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) Y_t = c + (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q) \epsilon_t.$$  

In the case all the roots of the polynomial $(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p) = 0$ lies **outside the unit circle**, we know from the properties of lag operator in last chapter that there exist a polynomial $\varphi(L)$ such that

$$
\varphi(L) = (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)^{-1} (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)
$$

with

$$
\sum_{j=0}^{\infty} |\varphi_j| < \infty.
$$

Thus

$$
Y_t = \mu + \varphi(L) \epsilon_t,
$$

where

$$
\varphi(L) = \frac{(1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)}{(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)},
$$

$$
\sum_{j=0}^{\infty} |\varphi_j| < \infty,
$$

$$
\mu = \frac{c}{1 - \phi_1 - \phi_2 - \ldots - \phi_p}.
$$
This can be viewed as an $MA(\infty)$ process with absolute summable coefficient. Therefore, the $ARMA(p, q)$ process is stationary and ergodic provided that all the roots of $(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) = 0$ lies outside the unit circle. Thus, the stationarity of an $ARMA(p, q)$ process depends entirely on the autoregressive parameters ($\phi_1, \phi_2, ..., \phi_p$) and not on the moving average parameters ($\theta_1, \theta_2, ..., \theta_q$).

### 3.2 The Dependence Structure

Assume that the $ARMA(p, q)$ process under consideration is weakly-stationary, then taking expectation on both side we have

$$E(Y_t) = c + \phi_1 E(Y_{t-1}) + \phi_2 E(Y_{t-2}) + ... + \phi_p E(Y_{t-p}) + E(\varepsilon_t) + \theta_1 E(\varepsilon_{t-1}) + ... + \theta_q E(\varepsilon_{t-q}).$$

Since by assumption that the process is stationary,

$$E(Y_t) = E(Y_{t-1}) = E(Y_{t-2}) = ... = E(Y_{t-p}) = \mu.$$

Therefore,

$$\mu = c + \phi_1 \mu + \phi_2 \mu + ... + \phi_p + 0 + 0 + ... + 0$$

or

$$\mu = \frac{c}{1 - \phi_1 - \phi_2 - ... - \phi_p}.$$

To find the higher moment of $Y_t$ in an analogous manner, we rewrite this $ARMA(p, q)$ as

$$Y_t = \mu(1 - \phi_1 - \phi_2 - ... - \phi_p) + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$$

or

$$(Y_t - \mu) = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + ... + \phi_p(Y_{t-p} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} \quad (14)$$

For $j \geq 0$, multiply $Y_{t-j} - \mu$ on both side of (14) and take expectation:

$$\gamma_j = E[(Y_t - \mu)(Y_{t-j} - \mu)]$$

$$= \phi_1 E[(Y_{t-1} - \mu)(Y_{t-j} - \mu)] + \phi_2 E[(Y_{t-2} - \mu)(Y_{t-j} - \mu)] + ... + \phi_p E[(Y_{t-p} - \mu)(Y_{t-j} - \mu)] + (Y_{t-j} - \mu) \varepsilon_t + \theta_1 E(Y_{t-j} - \mu)(\varepsilon_{t-1}) + ... + \theta_q E(Y_{t-j} - \mu)(\varepsilon_{t-q})$$

$$= \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + ... + \phi_p \gamma_{j-p} + E(Y_{t-j} - \mu) \varepsilon_t + \theta_1 E(Y_{t-j} - \mu)(\varepsilon_{t-1}) + ... + \theta_q E(Y_{t-j} - \mu)(\varepsilon_{t-q}).$$
It is obvious that the term \( E(Y_{t-j} - \mu)\varepsilon_t + \theta_1 E(Y_{t-j} - \mu)(\varepsilon_{t-1}) + \ldots + \theta_q E(Y_{t-j} - \mu)(\varepsilon_{t-q}) = 0 \) when \( j > q \).

Therefore we the results that

\[ \gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \ldots + \phi_p \gamma_{j-p}, \quad \text{for} \ j = q + 1, q + 2, \ldots \quad (15) \]

Thus, after \( q \) lags the autocovariance function \( \gamma_j \) follow the \( p \)th order difference equation governed by the autoregressive coefficients. However, (15) does not hold for \( j \leq q \), owing to correlation between \( \theta_j \varepsilon_{t-j} \) and \( Y_{t-j} \). Hence, an ARMA\((p, q)\) process will have more complicated autocovariance function from lag 1 through \( q \) than would the corresponding AR\((p)\) process.

### 3.3 Common Factor

Therefore is a potential for redundant parameterizations with ARMA process. Consider factoring the lag polynomial operator in an ARMA\((p, q)\) process:

\[
(1 - \lambda_1 L)(1 - \lambda_2 L)\ldots(1 - \lambda_p L)Y_t = (1 - \eta_1 L)(1 - \eta_2 L)\ldots(1 - \eta_q L)\varepsilon_t. \quad (16)
\]

We assume that \(|\lambda_i| < 1\) for all \( i \), so that the process is covariance-stationary. If the autoregressive operator \((1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)\) and the moving average operator \((1 + \theta_1 L + \theta_2 L^2 - \ldots + \theta_q L^q)\) have any roots in common, say, \( \lambda_i = \eta_j \) for some \( i \) and \( j \), then both side of (16) can be divided by \((1 - \lambda_i L)\):

\[
\prod_{k=1, k \neq i}^p (1 - \lambda_k)Y_t = \prod_{k=1, k \neq j}^q (1 - \eta_k)\varepsilon_t,
\]

or

\[
(1 - \phi_1^* L - \phi_2^* L^2 - \ldots - \phi_p^* L^{p-1})Y_t = (1 + \theta_1^* L + \theta_2^* L^2 - \ldots + \theta_q^* L^{q-1})\varepsilon_t. \quad (17)
\]

where

\[
(1 - \phi_1^* L - \phi_2^* L^2 - \ldots - \phi_p^* L^{p-1}) = (1 - \lambda_1 L)(1 - \lambda_2 L)\ldots(1 - \lambda_i L)(1 - \lambda_i L)\ldots(1 - \lambda_p L),
\]

and

\[
(1 + \theta_1^* L + \theta_2^* L^2 - \ldots + \theta_q^* L^{q-1}) = (1 - \eta_1 L)(1 - \eta_2 L)\ldots(1 - \eta_j L)(1 - \eta_j L)\ldots(1 - \eta_q L).
\]

The stationary ARMA\((p, q)\) process satisfying (16) is clearly identical to the stationary ARMA\((p-1, q-1)\) process satisfying (17) which is more parsimonious in parameters.
4 The Autocovariance-Generating Function

If \{Y_t, t \in T\} is a stationary process with autocovariance function \(\gamma_j\), then its autocovariance-generating function is defined by

\[
g_Y(z) = \sum_{j=-\infty}^{\infty} \gamma_j z^j.
\]

**Proposition**

If two different process share the same autocovariance-generating function, then the two processes exhibit the identical sequence of autocovariance.

As an example of calculating an autocovariance-generating function, consider the \(MA(1)\) process. Its autocovariance-generating function is:

\[
g_Y(z) = \sigma^2 (1 + \theta z) (1 + \theta z^{-1}). \tag{18}
\]

The form of expression (18) suggests that for the \(MA(q)\) process, its autocovariance-generating function might be calculated as

\[
g_Y(z) = \sigma^2 (1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q) (1 + \theta_1 z^{-1} + \theta_2 z^{-2} + \ldots + \theta_q z^{-q}). \tag{19}
\]

This conjecture can be verified by carrying out the multiplication in (19) and collecting terms by power of \(z\):

\[
\sigma^2 (1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q) (1 + \theta_1 z^{-1} + \theta_2 z^{-2} + \ldots + \theta_q z^{-q})
\]

\[
= (\theta_q) z^q + (\theta_{q-1} + \theta_q \theta_1) z^{q-1} + (\theta_{q-2} + \theta_{q-1} \theta_1 + \theta_q \theta_2) z^{q-2}
\]

\[
+ \ldots + (\theta_1 + \theta_2 \theta_1 + \theta_3 \theta_2 + \ldots + \theta_q \theta_{q-1}) z^1
\]

\[
= \sigma^2 + (1 + \theta_1^{2} + \theta_2^{2} + \ldots + \theta_q^{2}) z^0
\]

\[
+ (\theta_1 + \theta_2 \theta_1 + \theta_3 \theta_2 + \ldots + \theta_q \theta_{q-1}) z^{-1} + \ldots + (\theta_q) z^{-q}.
\]

The coefficient on \(z^j\) is indeed the \(j\)th autocovariance in an \(MA(q)\) process.

The method for finding \(g_Y(z)\) extends to the \(MA(\infty)\) case. If

\[
Y_t = \mu + \varphi(L) \varepsilon_t
\]

with

\[
\varphi(L) = \varphi_0 + \varphi_1 L + \varphi_2 L^2 + \ldots
\]
and
\[ \sum_{j=0}^{\infty} |\varphi_j| < \infty, \]
then
\[ g_Y(z) = \sigma^2 \varphi(z) \varphi(z^{-1}). \]

**Example:**
An AR(1) process:
\[ Y_t - \mu = (1 - \phi L)^{-1} \varepsilon_t, \]
is in this form with \( \varphi(L) = 1/(1 - \phi L). \) Thus, the autocovariance-generating function can be calculated from
\[ g_Y(z) = \frac{\sigma^2}{(1 - \phi z)(1 - \phi z^{-1})}. \]

**Example:**
The autocovariance-generating function of an ARMA\((p, q)\) process is therefore be written:
\[ g_Y(z) = \frac{\sigma^2(1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q)(1 + \theta_1 z^{-1} + \theta_2 z^{-2} + \ldots + \theta_q z^{-q})}{(1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p)(1 - \phi_1 z^{-1} - \phi_2 z^{-2} - \ldots - \phi_p z^{-p})}. \]

4.1 Filters

Sometimes the data are *filtered*, or treated in a particular way before they are analyzed, and we would like to summarize the effect of this treatment on the autocovariance. This calculation is particularly simple using the autocovariance-generating function. For example, suppose that the original data, \( Y_t \) were generated from an MA(1) process:
\[ Y_t = (1 + \theta L) \varepsilon_t, \]
which has autocovariance-generating function

\[ g_Y(z) = \sigma^2 \cdot (1 + \theta z)(1 + \theta z^{-1}). \]

Let the data be analyzed is \( X_t \) which is taking first difference of \( Y_t \):

\[ X_t = Y_t - Y_{t-1} = (1 - L)Y_t = (1 - L)(1 + \theta L)\epsilon_t. \]

Regarding this \( X_t \) as an \( MA(2) \) process, then it has the autocovariance-generating function as

\[ g_X(z) = \sigma^2 \cdot [(1 - z)(1 + \theta z)][(1 - z^{-1})(1 + \theta z^{-1})] \]

\[ = \sigma^2 \cdot [(1 - z)(1 - z^{-1})][(1 + \theta z)(1 + \theta z^{-1})] \]

\[ = [(1 - z)(1 - z^{-1})]g_Y(z). \]

Therefore, applying the filter \((1 - L)\) to \( Y_t \) thus resulting in multiplying its autocovariance-generating function by \((1 - z)(1 - z^{-1})\).

This principle readily generalizes. Let the original data series be \( Y_t \) and it is filtered according to

\[ X_t = h(L)Y_t, \]

with

\[ h(L) = \sum_{j=-\infty}^{\infty} h_j L^j, \quad \text{and} \]

\[ \sum_{j=-\infty}^{\infty} |h_j| < \infty. \]

The autocovariance-generating function of \( X_t \) can according be calculated as

\[ g_X(z) = h(z)h(z^{-1})g_Y(z). \]
5 Invertibility

5.1 Invertibility for the \( MA(1) \) Process

Consider an \( MA(1) \) process,

\[
Y_t - \mu = (1 + \theta L)\epsilon_t,
\]
with

\[
E(\epsilon_t\epsilon_s) = \begin{cases} 
\sigma^2 & \text{for } t = s \\
0 & \text{otherwise}.
\end{cases}
\]

Provided that \(|\theta| < 1\) both side of (20) can be multiplied by \((1 + \theta L)^{-1}\) to obtain

\[
(1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + ...) (Y_t - \mu) = \epsilon_t,
\]

which could be viewed as an \( AR(\infty) \) representation. If a moving average representation such as (20) can be rewritten as an \( AR(\infty) \) representation such as (21) simply by inverting the moving average operator \((1 + \theta L)\), then the moving average representation is said to be invertible. For an \( MA(1) \) process, invertibility requires \(|\theta| < 1\); if \(|\theta| \geq 1\), then the infinite sequence in (21) would not be well defined.\(^9\)

Let us investigate what invertibility means in terms of the first and second moments. Recall that the \( MA(1) \) process (21) has mean \( \mu \) and autocovariance-generating function

\[
g_Y(z) = \sigma^2(1 + \theta z)(1 + \theta z^{-1}).
\]

Now consider a seemingly different \( MA(1) \) process,

\[
\tilde{Y}_t - \mu = (1 + \hat{\theta} L)\tilde{\epsilon}_t,
\]
with \( \tilde{\epsilon}_t \) a white noise sequence having different variance

\[
E(\tilde{\epsilon}_t\tilde{\epsilon}_s) = \begin{cases} 
\tilde{\sigma}^2 & \text{for } t = s \\
0 & \text{otherwise}.
\end{cases}
\]

\(^9\)When \(|\theta| \geq 1\), \((1 + \theta L)^{-1} = \frac{\theta^{-1}L^{-1}}{1 - \theta^{-1}L^{-1}} = \theta^{-1}L^{-1}(1 - \theta^{-1}L^{-1} + \theta^{-2}L^{-2} - \theta^{-3}L^{-3} + \ldots)\)
Note that $\tilde{Y}_t$ has the same mean ($\mu$) as $Y_t$. Its autocovariance-generating function is

$$g_Y(z) = \tilde{\sigma}^2 (1 + \tilde{\theta}z)(1 + \tilde{\theta}z^{-1})$$

$$= \tilde{\sigma}^2 \{(\tilde{\theta}^{-1}z^{-1} + 1)(\tilde{\theta}z)\} \{(\tilde{\theta}^{-1}z + 1)(\tilde{\theta}z^{-1})\}$$

$$= (\tilde{\sigma}^2 \tilde{\theta}^2)(1 + \tilde{\theta}^{-1}z)(1 + \tilde{\theta}^{-1}z^{-1}).$$

Suppose that the parameters of (22), $(\tilde{\theta}, \tilde{\sigma}^2)$ are related to those of (20) by the following equations:

$$\theta = \tilde{\theta}^{-1}, \quad \sigma^2 = \tilde{\theta}^2 \tilde{\sigma}^2. \quad (23)$$

Then the autocovariance-generating function $g_Y(z)$ and $g_{\tilde{Y}}(z)$ would be the same, meaning that $Y_t$ and $\tilde{Y}_t$ would have identical first and second moments.

Notice from (23) that if $|\theta| < 1$, then $|\tilde{\theta}| > 1$. In other words, for any invertible $MA(1)$ representation (20), we have found a non-invertible $MA(1)$ representation (22) with the same first and second moments as the invertible representation. Conversely, given any non-invertible representation with $|\tilde{\theta}| > 1$, there exist an invertible representation with $\theta = (1/\tilde{\theta})$ that has the same first and second moments as the noninvertible representation.

Not only do the invertible and noninvertible representation share the same moments, either representation (20) or (22) could be used as an equally valid description of any given $MA(1)$ process. Suppose a computer generated an infinite sequence of $\tilde{Y}_t$’s according to the noninvertible $MA(1)$ process:

$$\tilde{Y}_t - \mu = (1 + \tilde{\theta}L)\tilde{\varepsilon}_t$$

with $|\tilde{\theta}| > 1$. In what sense could these same data be associated with a invertible $MA(1)$ representation?

Imagine calculating a series $\{\varepsilon_t\}_{t=-\infty}^\infty$ defined by

$$\varepsilon_t \equiv (1 + \theta L)^{-1}(\tilde{Y}_t - \mu) \quad (25)$$

$$= (\tilde{Y}_t - \mu) - \theta(\tilde{Y}_{t-1} - \mu) + \theta^2(\tilde{Y}_{t-2} - \mu) - \theta^3(\tilde{Y}_{t-3} - \mu) + ...., \quad (26)$$

where $\theta = (1/\tilde{\theta})$. 
The autocovariance-generating function of \( \varepsilon_t \) is
\[
    g_{\varepsilon}(z) = (1 + \theta z^{-1})(1 + \theta z)^{-1} g_{\tilde{Y}}(z) \\
    = (1 + \theta z^{-1})(1 + \theta z)^{-1}(\tilde{\sigma}^2 \tilde{\theta}^2)(1 + \tilde{\theta}^{-1} z)(1 + \tilde{\theta}^{-1} z^{-1}) \\
    = (\tilde{\sigma}^2 \tilde{\theta}^2),
\]
where the last equality follows from the fact that \( \tilde{\theta}^{-1} = \theta \). Since the autocovariance-generating function is a constant, it follows that \( \varepsilon_t \) is a white noise process with variance \( \sigma^2 = \tilde{\sigma}^2 \tilde{\theta}^2 \).

Multiplying both side of (25) by \( (1 + \theta L) \),
\[
    \tilde{Y}_t - \mu = (1 + \theta L) \varepsilon_t
\]
is a perfectly valid invertible MA(1) representation of data that were actually generated from the noninvertible representation (22).

The converse proposition is also true—suppose that the data were generated really from (20) with \( |\theta| < 1 \), an invertible representation. Then there exists a noninvertible representation with \( \tilde{\theta} = 1/\theta \) that describes these data with equal validity.

Either the invertible or the noninvertible representation could characterize any given data equally well, though there is a practical reason for preferring the invertible representation. To find the value of \( \varepsilon \) for date \( t \) associated with the invertible representation as in (20), we need to know current and past value of \( Y \). By contrast, to find the value of \( \tilde{\varepsilon} \) for date \( t \) associated with the noninvertible representation, we need to use all of the future value of \( Y \). If the intention is to calculate the current value of \( \varepsilon_t \) using real-world data, it will be feasible only to work with the invertible representation.
5.2 Invertibility for the MA($q$) Process

Consider the MA($q$) process,

$$Y_t - \mu = (1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)\varepsilon_t$$

$$E(\varepsilon_t \varepsilon_s) = \begin{cases} \sigma^2 & \text{for } t = s \\ 0 & \text{otherwise.} \end{cases}$$

Provided that all the roots of

$$(1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q) = 0$$

lie outside the unit circle, this MA($q$) process can be written as an AR($\infty$) simply by inverting the MA operator,

$$(1 + \eta_1 L + \eta_2 L^2 + ...)(Y_t - \mu) = \varepsilon_t,$$

where

$$(1 + \eta_1 L + \eta_2 L^2 + ...) = (1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)^{-1}.$$