Lecture 8
Time Series:
Stationarity, AR(p) & MA(q)

Time Series: Introduction

• A time series \( y_t \) is a process observed in sequence over time, \( t = 1, \ldots, T \) \( \Rightarrow Y_t = \{y_1, y_2, y_3, \ldots, y_T\} \).

Examples: IBM monthly stock prices from 1973: January till today; or USD/GBP daily exchange rates from February 15, 1923 to March 19, 1938.
**Time Series: Introduction**

• Given the sequential nature of $Y_t$, we expect $y_t$ & $y_{t-1}$ to be dependent.

*Technical note:* Time series dependence creates statistical problems: the classical results (based on LLN & CLT) tend not to be valid. New assumptions and tools (stationarity, ergodicity, martingale difference sequences or MDS) are needed.

• We use different tools to identify models that capture this dependence.

• In a time series model, we describe how $y_t$ depends on past $y_i$’s. That is, the information set is $F_t = \{y_{t-1}, y_{t-2}, y_{t-3}, \ldots\}$

**Time Series: Introduction**

• We estimate time series models to forecast out-of-sample. For example, the $l$-step ahead forecast: $\hat{y}_{T+l} = E_t[y_{t+l} | F_t]$.

• In the 1970s it was found that very simple time series models outforecasted very sophisticated (big) economic models. Big shock to big multivariate models.
**Time Series: Introduction – Categories**

- Usually, time series models are separated into two categories:
  - Univariate \( (y_t \in \mathbb{R}, \text{it is a scalar}) \)
    
    **Example:** We are interested in the behavior of IBM stock prices as a function of its past.
    
    \( \Rightarrow \) Primary model: Autoregressions (ARs).

  - Multivariate \( (y_t \in \mathbb{R}^m, \text{it is a vector-valued}) \)
    
    **Example:** We are interested in the joint behavior of IBM stock and bond prices as a function of their joint past.
    
    \( \Rightarrow \) Primary model: Vector autoregressions (VARs).

**Time Series: Introduction – AR and MA models**

- Two popular models for \( E[y_t | F_{t-1}] \):
  - An autoregressive (AR) process models \( E[y_t | F_{t-1}] \) with lagged dependent variables:
    
    \[ E[y_t | F_{t-1}] = f(y_{t-1}, y_{t-2}, y_{t-3}, \ldots) \]

    **Example:** AR(1) process, \( y_t = \alpha + \beta y_{t-1} + \varepsilon_t \).

  - A moving average (MA) process models \( E[y_t | F_{t-1}] \) with lagged errors, \( \varepsilon_t \):
    
    \[ E[y_t | F_{t-1}] = f(\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \ldots) \]

    **Example:** MA(1) process, \( y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t \)

- There is a third model, ARMA, that combines lagged dependent variables and lagged errors.
Time Series: Introduction – White Noise

- In general, we assume the error term, $\varepsilon_t$, is uncorrelated with everything, with mean 0 and constant variance, $\sigma^2$. We call a process like this a white noise (WN) process.

- We denote a WN process as $\varepsilon_t \sim WN(0, \sigma^2)$

- White noise is the basic building block of all time series. $z_t = \sigma u_t, \quad u_t \sim i.i.d (0, 1) \implies z_t \sim WN(0, \sigma^2)$

- The $z_t$’s are random shocks, no dependence over time, representing unpredictable events. It represents a model of news.

Time Series: Introduction – Forecasting

- We want to select an appropriate time series model to forecast $y_t$. In this class, we will use linear model, with choices: AR(p), MA(q) or ARMA(p, q).

- Steps for forecasting:
  1. Identify the appropriate model. That is, determine p, q.
  2. Estimate the model.
  3. Test the model.
  4. Forecast.

- In this lecture, we go over the statistical theory (stationarity, ergodicity and MDS CLT), the main models (AR, MA & ARMA) and tools that will help us describe and identify a proper model.
**Time Series: Introduction – Conditionality**

- We make a key distinction: Conditional vs Unconditional moments. In time series we model the conditional mean as a function of its past, for example in an AR(1) process, we have:

\[ y_t = \alpha + \beta y_{t-1} + \varepsilon_t \]

Then, the conditional mean forecast at time \( t \), conditioning on information at time \( I_{t-1} \), is:

\[ E_t[y_t | I_{t-1}] = E_t[y_t] = \alpha + \beta y_{t-1} \]

Notice that the unconditional mean is given by:

\[ E[y_t] = \alpha + \beta E[y_{t-1}] = \alpha/(1 - \beta) = \text{constant} \]

The conditional mean is time varying; the unconditional mean is not!

**Key distinction:** Conditional vs. Unconditional moments.

---

**CLM Revisited: Time Series**

- With autocorrelated data, we get dependent observations. For example, with autocorrelated errors:

\[ \varepsilon_t = \rho \varepsilon_{t-1} + u_t \]

the independence assumption (\( A2' \)) is violated. The LLN and the CLT cannot be easily applied in this context. We need new tools.

- We will introduce the concepts of stationarity and ergodicity. The ergodic theorem will give us a counterpart to the LLN.

To get asymptotic distributions, we also need a CLT for dependent variables, using new technical concepts: mixing and stationarity. Or we can rely on a new CLT: The martingale CLT.

- We will not cover these technical points in detail.
### Time Series – Stationarity

- Consider the joint probability distribution of the collection of RVs:

\[
F(y_{t_1}, y_{t_2}, \ldots, y_{t_T}) = F(Y_{t_1} \leq y_{t_1}, Y_{t_2} \leq y_{t_2}, \ldots, Y_{t_T} \leq y_{t_T})
\]

Then, we say that a process is stationary of:

- 1st order if \( F(y_{t_1}) = F(y_{t_1+k}) \) for any \( t_1, k \)
- 2nd order if \( F(y_{t_1}, y_{t_2}) = F(y_{t_1+k}, y_{t_2+k}) \) for any \( t_1, t_2, k \)
- Nth-order if \( F(y_{t_1}, \ldots, y_{t_T}) = F(y_{t_1+k}, \ldots, y_{t_T+k}) \) for any \( t_1, \ldots, t_T, k \)

- Moments describe a distribution. We calculate moments as usual:

\[
E(Y_t) = \mu \\
\text{Var}(Y_t) = \sigma^2 = E[(Y_t - \mu)^2] \\
\text{Cov}(Y_{t_1}, Y_{t_2}) = E[(Y_{t_1} - \mu)(Y_{t_2} - \mu)] = \gamma(t_1 - t_2)
\]

- \( \text{Cov}(Y_{t_1}, Y_{t_2}) = \gamma(t_1 - t_2) \) is called the auto-covariance function.

### Time Series – Stationarity & Autocovariances

Note: \( \gamma(0) \) is the variance.

- The autocovariance function is symmetric. That is,

\[
\gamma(t_1 - t_2) = \text{Cov}(Y_{t_1}, Y_{t_2}) = \text{Cov}(Y_{t_2}, Y_{t_1}) = \gamma(t_2 - t_1)
\]

- From the autocovariances, we derive the autocorrelations:

\[
\text{Corr}(Y_{t_1}, Y_{t_2}) = \rho(Y_{t_1}, Y_{t_2}) = \frac{\gamma(t_1 - t_2)}{\sigma_{t_1} \sigma_{t_2}} = \frac{\gamma(t_1 - t_2)}{\gamma(0)}
\]

the last step takes into account stationarity: \( (\sigma_{t_1} = \sigma_{t_2} = \sqrt{\gamma(0)}) \)

- \( \text{Corr}(Y_{t_1}, Y_{t_2}) = \rho(Y_{t_1}, Y_{t_2}) \) is called the auto-correlation function (ACF),

- think of it as a function of \( k = t_1 - t_2 \). The ACF is also symmetric.

- Stationarity requires all these moments to be independent of time. If the moments are time dependent, we say the series is non-stationary.
Time Series – Stationarity & Constant Moments

• For strictly stationary process (constant moments), we need:
  \[ \mu_t = \mu \]
  \[ \sigma_t = \sigma \]
  because \( F(y_t) = F(y_{t+k}) \rightarrow \mu_t = \mu_{t+k} = \mu \)
  \[ \sigma_t = \sigma_{t+k} = \sigma \]

Then,
\[ F(y_{t_1}, y_{t_2}) = F(y_{t_1+k}, y_{t_2+k}) \Rightarrow \text{Cov}(y_{t_1}, y_{t_2}) = \text{Cov}(y_{t_1+k}, y_{t_2+k}) \]
\[ \Rightarrow \rho(t_1, t_2) = \rho(t_1 + k, t_2 + k) \]

Let \( t_1 = t - k \) & \( t_2 = t \)
\[ \Rightarrow \rho(t_1, t_2) = \rho(t - k, t) = \rho(t, t - k) = \rho(k) = \rho_k \]

The correlation between any two RVs depends on the time difference. Given the symmetry, we have \( \rho(k) = \rho(-k) \).

Time Series – Weak Stationary

• A Covariance stationary process (or 2nd order weakly stationary) has:
  - constant mean
  - constant variance
  - covariance function depends on time difference between RVs.

That is, \( Z_t \) is covariance stationary if:
\[ E(Z_t) = \text{constant} \]
\[ Var(Z_t) = \text{constant} \]
\[ \text{Cov}(Z_{t_1}, Z_{t_2}) = E[(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2})] = \gamma(t_1 - t_2) = f(t_1 - t_2) \]
Example: Assume $\varepsilon_t \sim WN(0, \sigma^2)$.

$$y_t = \phi y_{t-1} + \varepsilon_t$$  \hspace{1cm} \text{(AR(1) process)}

• Mean
Taking expectations on both side:

$$E[y_t] = \phi E[y_{t-1}] + E[\varepsilon_t]$$

$$\mu = \phi \mu + 0$$

$$E[y_t] = \mu = 0$$  \hspace{1cm} \text{(assuming $\phi \neq 1$)}

• Variance
Applying the variance on both side:

$$Var[y_t] = \gamma(0) = \phi^2 Var[y_{t-1}] + Var[\varepsilon_t]$$

$$\gamma(0) = \sigma^2/(1 - \phi^2)$$  \hspace{1cm} \text{(assuming $|\phi| < 1$)}

Example (continuation): $y_t = \phi y_{t-1} + \varepsilon_t$  \hspace{1cm} \text{(AR(1) process)}

• Covariance

$$\gamma(1) = Cov[y_t, y_{t-1}] = E[y_t y_{t-1}] = E[(\phi y_{t-1} + \varepsilon_t)y_{t-1}]$$

$$= \phi E[y_{t-1}]^2 = \phi Var[y_{t-1}] = \phi \gamma(0)$$

$$= \phi \left[\sigma^2/(1 - \phi^2)\right]$$

$$\gamma(2) = Cov[y_t, y_{t-2}] = E[y_t y_{t-2}] = E[(\phi y_{t-1} + \varepsilon_t)y_{t-2}]$$

$$= \phi E[y_{t-1}] E[y_{t-2}] = \phi \phi \gamma(0) = \phi^2 \gamma(0)$$

$$= \phi^2 \left[\sigma^2/(1 - \phi^2)\right]$$

$$\vdots$$

$$\gamma(k) = Cov[y_t, y_{t-k}] = \phi^k \gamma(0)$$

$\Rightarrow$ If $|\phi| < 1$, the process is covariance stationary: mean, variance and covariance are constant.
**Time Series – Stationarity: Example**

*Example (continuation)*: $y_t = \phi y_{t-1} + \varepsilon_t$ \hspace{1em} (AR(1) process)

- **Covariance**
  
  $\gamma(k) = \text{Cov}[y_t, y_{t-k}] = \phi^k \gamma(0)$

Note: From the autocovariance function, we can derive the correlation function:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\phi^k \gamma(0)}{\gamma(0)} = \phi^k$$

**Time Series – Non-Stationarity: Example**

*Example*: Assume $\varepsilon_t \sim \text{WN}(0, \sigma^2)$.

$$y_t = \mu + y_{t-1} + \varepsilon_t$$ \hspace{1em} (Random Walk with drift process)

Doing backward substitution:

1. $y_t = \mu + (\mu + y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$
2. $= 2\mu + y_{t-2} + \varepsilon_t + \varepsilon_{t-1}$
3. $= 2\mu + (\mu + y_{t-3} + \varepsilon_{t-2}) + \varepsilon_t + \varepsilon_{t-1}$
4. $= 3\mu + y_{t-3} + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2}$

$\Rightarrow y_t = \mu \cdot t + \sum_{j=0}^{t-1} \varepsilon_{t-j} + y_0$

- **Mean and Variance**

  $E[y_t] = \mu \cdot t + y_0$

  $\text{Var}[y_t] = \gamma(0) = \sum_{j=0}^{t-1} \sigma^2 = \sigma^2 \cdot t$

$\Rightarrow$ the process is non-stationary; that is, moments are time dependent.
Stationary Series

Examples:

\[ y_t = 0.08 + \epsilon_t + 0.4 \epsilon_{t-1} \quad \epsilon_t \sim WN \]
\[ y_t = 0.13 y_{t-1} + \epsilon_t \]

Non-Stationary Series

Examples: Assume \( \epsilon_t \sim WN(0, \sigma^2) \).

\[ y_t = \mu + t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \quad AR(2) \text{ with deterministic trend} \]
\[ y_t = \mu + y_{t-1} + \epsilon_t \quad \text{Random Walk with drift} \]
Time Series – Ergodicity

- Intuition behind Ergodicity:
  We go to a casino to play a game with 20% return, but on average, one gambler out of 100 goes bankrupt. If 100 gamblers play the game, there is a 99% chance of winning and getting a 20% return. This is the ensemble scenario. Suppose that gambler 35 is the one that goes bankrupt. Gambler 36 is not affected by the bankruptcy of gambler 35.

Suppose now that instead of 100 gamblers you play the game 100 times. This is the time series scenario. You keep winning 20% every day until day 35 where you go bankrupt. There is no day 36 for you.

Result: The probability of success from the group (ensemble scenario) does not apply to one person (time series scenario).

Ergodicity describes a situation where the ensemble scenario outcome applies to the time series scenario.

Time Series – Ergodicity of the Mean

- We want to estimate the mean of the process \( \{Z_t\}, \mu(Z_t) \). But, we need to distinguishing between ensemble average (with \( m \) observations) and time average (with \( n \) observations):
  - Ensemble Average \( \bar{z} = \frac{\sum_{i=1}^{m} Z_i}{m} \)
  - Time Series Average \( \bar{z} = \frac{\sum_{i=1}^{n} Z_i}{n} \)

Q: Which estimator is the most appropriate?
A: Ensemble Average. But, it is impossible to calculate. We only observe one \( Z_t \).

- Q: Under which circumstances we can use the time average (only one realization of \( \{Z_t\} \))? Is the time average an unbiased and consistent estimator of the mean? The Ergodic Theorem gives us the answer.
Time Series – Ergodicity of the Mean

- **Definition:** A covariance-stationary process is *ergodic* for the mean if
  \[ \text{plim } \bar{z} = \text{E}[Z_t] = \mu \]

- It can be shown that the \( \text{var} [\bar{Z}] \) can be written as a function of the autocorrelations, \( \rho_k \):
  \[ \text{var} [\bar{Z}] = \frac{\sigma^2}{n} \sum_k (1 - \frac{|k|}{n}) \rho_k \]

**Theorem:** A sufficient condition for ergodicity for the mean is that the autocorrelations \( \rho_k \) between two observations, say \( (y_{t_i}, y_{t_j}) \), \( \rho(t_i, t_j) = \rho_{t_i-t_j} \), go to zero as \( t_i \) & \( t_j \) grow further apart.

Condition for ergodicity: \( \rho_k \to 0 \), as \( k \to \infty \)

Time Series – Lag Operator

- Define the operator \( L \) as
  \[ L^k z_t = z_{t-k} \]

- It is usually called *Lag operator*. But it can produce lagged or forward variables (for negative values of \( k \)). For example:
  \[ L^{-3} z_t = z_{t+3} \]

- Also note that if \( c \) is a constant \( \Rightarrow L \, c = c \).

- Sometimes the notation for \( L \) when working as a lag operator is \( B \) (*backshift operator*), and when working as a forward operator is \( F \).

- Important application: Differencing
  \[ \Delta z_t = (1-L)z_t = z_t - z_{t-1} \]
  \[ \Delta^d z_t = (1-L)^d z_t \]
**Time Series – Useful Result: Geometric Series**

- The function \( f(x) = (1 - x)^{-1} \) can be written as an infinite geometric series (use a Maclaurin series around \( c=0 \)):
  \[
  f(x) = \frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \ldots = \sum_{n=0}^{\infty} x^n
  \]

- If we multiply \( f(x) \) by a constant, \( a \):
  \[
  \sum_{n=0}^{\infty} ax^n = \frac{a}{1-x} \quad \rightarrow \quad \sum_{n=1}^{\infty} ax^n = a \left( \frac{1}{1-x} - 1 \right)
  \]

- We will use this result when, under certain conditions, we invert a lag polynomial (say, \( \theta(L) \)) to convert an AR (MA) process into an infinite MA (AR) process.

**Example:** Suppose we have an MA(1) process:
\[
y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t = \mu + \theta(L) \varepsilon_t, \quad \text{with} \ \theta(L) = (1 + \theta_1 L)
\]

**Example (continuation):** \( y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t = \mu + \theta(L) \varepsilon_t \),
where we have defined the lag polynomial:
\[
\theta(L) = (1 + \theta_1 L).
\]

Recall,
\[
f(x) = \frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \ldots = \sum_{n=0}^{\infty} x^n
\]

Let \( x = -\theta_1 L \). Then,
\[
\theta(L)^{-1} = \frac{1}{1-(-\theta_1 L)} = 1 + (-\theta_1 L) + (-\theta_1 L)^2 + (-\theta_1 L)^3 + (-\theta_1 L)^4 + \ldots
\]
\[
= \sum_{n=0}^{\infty} (-\theta_1 L)^n = 1 - \theta_1 L + \theta_1^2 L^2 - \theta_1^3 L^3 + \theta_1^4 L^4 + \ldots
\]

That is, we get an AR(\( \infty \)), by multiplying both sides by \( \theta(L)^{-1} \):
\[
\theta(L)^{-1} y_t = \theta(L)^{-1} \mu + \varepsilon_t = \mu^* + \varepsilon_t.
\]
\[
\theta(L)^{-1} y_t = y_t - \theta_1 y_{t-1} + \theta_1^2 y_{t-2} - \theta_1^3 y_{t-3} + \theta_1^4 y_{t-4} + \ldots = \mu^* + \varepsilon_t.
\]
Example (continuation):
That is, we get an AR(∞):
\[ \theta(L)^{-1} y_t = y_t - \theta_1 y_{t-1} + \theta_1^2 y_{t-2} - \theta_1^3 y_{t-3} + \theta_1^4 y_{t-4} + \cdots = \mu^* + \varepsilon_t. \]

Solving for \( y_t \):
\[ y_t = \mu^* + \theta_1 y_{t-1} - \theta_1^2 y_{t-2} + \theta_1^3 y_{t-3} - \theta_1^4 y_{t-4} + \cdots + \varepsilon_t. \]

Moving Average Process

• An MA process models \( E[y_t | F_{t-1}] \) with lagged error terms. An MA(\( q \)) model involves \( q \) lags.

• We keep the white noise assumption for \( \varepsilon_t \); \( \varepsilon_t \sim WN(0, \sigma^2) \)

Example: A linear MA(\( q \)) model:
\[ y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t = \mu + \theta(L) \varepsilon_t, \]
where
\[ \theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \theta_3 L^3 + \cdots + \theta_q L^q. \]

• In time series, the constant does not affect the properties of AR and MA process. It is usually removed (think of the data analyze as demeaned). Thus, in this situation we say “without loss of generalization”, we assume \( \mu=0. \)
Moving Average Process – Stationarity

• Q: Is MA(q) stationary? Check the moments (assume $\mu = 0$).

\[
y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + ... + \theta_q \varepsilon_{t-q}
\]
\[
y_{t-1} = \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} + \theta_2 \varepsilon_{t-3} + ... + \theta_q \varepsilon_{t-q} + \theta_q \varepsilon_{t-(q+1)}
\]

• Mean

\[
E[y_t] = \mu + \theta_1 E[\varepsilon_{t-1}] + \theta_2 E[\varepsilon_{t-2}] + ... + \theta_q E[\varepsilon_{t-q}] + E[\varepsilon_t] = \mu = 0
\]

• Variance

\[
\text{Var}[y_t] = \theta_1^2 \text{Var}[\varepsilon_{t-1}] + \theta_2^2 \text{Var}[\varepsilon_{t-2}] + ... + \theta_q^2 \text{Var}[\varepsilon_{t-q}] + \text{Var}[\varepsilon_t]
\]
\[
= (1 - \theta_1^2 - \theta_2^2 - ... - \theta_q^2) \sigma^2.
\]

To get a positive variance, we require \((1 - \theta_1^2 - \theta_2^2 - ... - \theta_q^2) > 0\).

Moving Average Process – Stationarity

• Covariance

\[
y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + ... + \theta_q \varepsilon_{t-q}
\]
\[
y_{t-1} = \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} + \theta_2 \varepsilon_{t-3} + ... + \theta_q \varepsilon_{t-q} + \theta_q \varepsilon_{t-(q+1)}
\]

\[
\gamma(1) = \text{Cov}[y_t, y_{t-1}] = E[y_t y_{t-1}]
\]
\[
= E[(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q})(\varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} + \theta_2 \varepsilon_{t-3} + ... + \theta_q \varepsilon_{t-(q+1)})]
\]
\[
= \theta_1 E[\varepsilon_{t-1}\varepsilon_{t-1}] + \theta_1^2 E[\varepsilon_{t-1}\varepsilon_{t-2}] + ... + \theta_q E[\varepsilon_{t-1}\varepsilon_{t-(q+1)}] + E[\varepsilon_{t-1}^2]
\]
\[
= E[\varepsilon_{t-1}] + \theta_1 E[\varepsilon_{t-1} \varepsilon_{t-2}] + \theta_2 E[\varepsilon_{t-1} \varepsilon_{t-3}] + ... + \theta_q E[\varepsilon_{t-1} \varepsilon_{t-(q+1)}]
\]
\[
= \theta_1 \sigma^2 + \theta_2 \theta_1 \sigma^2 + \theta_3 \theta_2 \sigma^2 + ... + \theta_q \theta_{q-1} \sigma^2 + 0
\]
\[
= \sigma^2 \sum_{j=1}^{q} \theta_j \theta_{j-1} \quad \text{(where } \theta_0 = 1)\]
Moving Average Process – Stationarity

We continue with the derivations of the \( \gamma(k) \) function. It is easier to derive them rewriting \( y_t \) & \( y_{t-2} \):

\[
y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \ldots + \theta_q \varepsilon_{t-q}
y_{t-2} = \varepsilon_{t-2} + \theta_1 \varepsilon_{t-3} + \theta_2 \varepsilon_{t-4} + \theta_3 \varepsilon_{t-5} + \ldots + \theta_q \varepsilon_{t-(q+2)}
\]

\( \gamma(2) = \text{Cov}[y_t, y_{t-2}] = E[y_t y_{t-2}] \)

\[
= E[(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q})^2]
= E[y_t \varepsilon_{t-2}] + \theta_1 E[y_t \varepsilon_{t-3}] + \ldots + \theta_q E[y_t \varepsilon_{t-(q+2)}]
= \theta_2 \sigma^2 + \theta_3 \theta_1 \sigma^2 + \theta_4 \theta_2 \sigma^2 + \ldots + \theta_q \theta_{q-2} \sigma^2 + 0
= \sigma^2 \sum_{j=2}^{q} \theta_j \theta_{j-2} \quad (\text{where } \theta_0 = 1)
\]

\( \gamma(q) = E[y_t y_{t-q}] = \)

\[
= E[\varepsilon_t \varepsilon_{t-q}] + \theta_1 E[\varepsilon_t \varepsilon_{t-1} y_{t-q}] + \theta_2 E[\varepsilon_t \varepsilon_{t-2} y_{t-q}] + \ldots + \theta_q E[\varepsilon_t \varepsilon_{t-q} y_{t-q}]
= \theta_q \sigma^2
= \sigma^2 \sum_{j=q}^{q} \theta_j \theta_{j-q} \quad (\text{where } \theta_0 = 1)
\]

Moving Average Process – Stationarity

\( \gamma(q) = \sigma^2 \sum_{j=q}^{q} \theta_j \theta_{j-q} \quad (\text{where } \theta_0 = 1) \)

In general, for the \( k \) autocovariance:

\[
\gamma(k) = \sigma^2 \sum_{j=k}^{q} \theta_j \theta_{j-k} \quad \text{for } |k| \leq q
\]

\( \gamma(k) = 0 \quad \text{for } |k| > q \)

Remark: After lag \( q \), the autocovariance (and autocorrelation function) are 0.

- It is easy to verify that the sums \( \sum_{j=k}^{q} \theta_j \theta_{j-k} \) are finite. Then, mean, variance and covariance are constant.

\( \Rightarrow \text{MA}(q) \) is always stationary.
MA Process – Invertibility

• Assuming $\theta(L) \neq 1$, we can invert $\theta(L)$. Then, by inverting $\theta(L)$, an MA($q$) process can generate an AR process:

\[
y_t = \mu + \theta(L) \varepsilon_t \quad \Rightarrow \quad \theta(L)^{-1} y_t = \mu^* + \varepsilon_t
\]

Then, we have an infinite sum polynomial on $\theta L$. (Recall the geometric series result.) That is, we convert an MA($q$) into an AR($\infty$).

\[
\sum_{j=0}^{\infty} \pi_j(L) y_t = \mu^* + \varepsilon_t
\]

MA Process – Invertibility

• We need to make sure that $\theta(L)^{-1}$ is defined: We require $\theta(L) \neq 0$. When this condition is met, we can write $\varepsilon_t$ as a causal function of $y_t$. We say the MA is invertible. For this to hold, we require:

\[
\sum_{j=0}^{\infty} |\pi_j(L)| < \infty
\]

Technical note: An MA($q$) is typically required to have roots of the characteristic equation $\theta(\alpha) = 0$ greater than one in absolute value.
MA Process – MA(1)

**Example**: MA(1) process:

\[ y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t = \mu + \theta(L) \varepsilon_t, \text{ with } \theta(L) = (1 + \theta_1 L) \]

**Moments**

- \[ E[y_t] = \mu = 0 \]
- \[ \text{Var}[y_t] = \gamma(0) = \sigma^2 + \theta_1^2 \sigma^2 = \sigma^2 (1 + \theta_1^2) \]
- \[ \text{Cov}[y_t, y_{t-1}] = \gamma(1) = E[y_t y_{t-1}] = E[(\theta_1 \varepsilon_{t-1} + \varepsilon_t)(\theta_1 \varepsilon_{t-2} + \varepsilon_{t-1})] \]

\[ = \theta_1 \sigma^2 \]

\[ \gamma(k) = E[y_t y_{t+k}] = E[(\theta_1 \varepsilon_{t-1} + \varepsilon_t)(\theta_1 \varepsilon_{t+k-1} + \varepsilon_{t+k})] = 0 \quad (\text{for } k > 1) \]

That is, for \(|k| > 1, \gamma(k) = 0.\)

To get the ACF, we divide the autocovariances by \(\gamma(0):\)

MA Process – Example: MA(1)

**Example (continuation):**

Then, the autocorrelation function (ACF):

\[ \rho(1) = \gamma(1)/\gamma(0) = \theta_1 \sigma^2 / \sigma^2 (1 + \theta_1^2) = \theta_1 / (1 + \theta_1^2) \]

\[ \rho(k) = \gamma(k)/\gamma(0) = 0 \quad (\text{for } k > 1) \]

**Remark:** The autocovariance function is zero after lag 1. Similarly, the ACF is also zero after lag 1.

- **Invertibility:** If \(|\theta| < 1, we can write \((1 + \theta L)^{-1} y_t + \mu^* = \varepsilon_t \)

\[ \Rightarrow (1 - \theta_1 L + \theta_1^2 L^2 + \ldots + \theta_1^j L^j + \ldots) y_t + \mu^* + \sum_{i=1}^{\infty} \pi_i(L)y_t = \varepsilon_t \]

That is, \(\pi_i = (-\theta_1)^i\).
MA Process – MA(2)

Example: MA(2) process:

\[ y_t = \mu + \theta_2 \varepsilon_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t = \mu + \theta(L) \varepsilon_t, \]

with \( \theta(L) = (1 + \theta_1 L + \theta_2 L^2) \)

• Moments

\[
E(Y_t) = \mu \\
\gamma_k = \begin{cases} 
\sigma^2 (1 + \theta_1^2 + \theta_2^2), & k = 0 \\
-\theta_1 \sigma^2 (1 - \theta_2), & |k| = 1 \\
-\theta_2 \sigma^2, & |k| = 2 \\
0, & |k| > 2 
\end{cases}
\]

Remark: The autocovariance function is zero after lag 2. Similarly, the ACF is also zero after lag 2.

MA Process – Example: MA(2)

– Invertibility: The roots of \( \lambda^2 - \theta_1 \lambda - \theta_2 = 0 \) all lie inside the unit circle. It can be shown the invertibility condition for an MA(2) process is:

\[
\theta_1 + \theta_2 < 1 \\
\theta_2 - \theta_1 < 1 \\
-1 < \theta_2 < 1
\]
MA Process – Estimation

• MA processes are more complicated to estimate. In particular, there are nonlinearities. Consider an MA(1):

\[ y_t = \varepsilon_t + \theta \varepsilon_{t-1} \]

The auto-correlation is \( \rho_1 = \theta / (1 + \theta^2) \). Then, we can use the method of moments (MM), which uses the estimated \( \rho_1, r_1 \), to estimate of \( \theta \):

\[ r_1 = \frac{\theta}{(1 + \theta^2)} \Rightarrow \hat{\theta} = \frac{1 \pm \sqrt{1 - 4r_1^2}}{2r_1} \]

• A nonlinear solution and difficult to solve.

• Alternatively, if \( |\theta| < 1 \), we can try \( a \in (-1; 1) \),

\[ \varepsilon_i(a) = y_i + a \varepsilon_{i-1} + a^2 \varepsilon_{i-2} + \ldots \]

and look (numerically) for the least-square estimator

\[ \hat{\theta} = \arg \min_{a} \{ S_f(a) = \sum_{i=1}^{T} \varepsilon_i^2(a) \} \]

The Wold Decomposition

**Theorem** - Wold (1938).

Any covariance stationary \( \{y_t\} \) has infinite order, moving-average representation:

\[ y_t = \sum_{j=0}^{\infty} \psi_j L^j \varepsilon_t + \kappa_t, \quad \psi_0 = 1, \]

where \( \kappa_t \): deterministic term (perfectly forecastable). Say, \( \kappa_t = \mu \)

\[ \sum_{j=0}^{\infty} \psi_j^2 < \infty \]

\[ \varepsilon_t \sim WN(0, \sigma^2) \]

Thus, \( y_t \) is a linear combination of innovations over time plus a deterministic part.

• A stationary process can be decomposed into a sum of two parts, one represented as an MA(\( \infty \)) and the other a deterministic “trend.”
The Wold Decomposition

**Example:**
Let $x_t = y_t - \kappa_t$ ($x_t = \text{MA}(\infty)$ part). Then, check moments:

\[
E[x_t] = E[y_t - \kappa_t] = \sum_{j=0}^{\infty} \psi_j E[\varepsilon_{t-j}] = 0.
\]

\[
E[x_t^2] = \sum_{j=0}^{\infty} \psi_j^2 E[\varepsilon_{t-j}^2] = \sigma^2 \sum_{j=0}^{\infty} \psi_j < \infty.
\]

\[
E[x_t x_{t-j}] = E[(\varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \ldots) (\varepsilon_{t-j} + \psi_1 \varepsilon_{t-j-1} + \psi_2 \varepsilon_{t-j-2} + \ldots)]
= \sigma^2 (\psi_j + \psi_1 \psi_{j+1} + \psi_2 \psi_{j+2} + \ldots) = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+2}
\]

$X_t$ is a covariance stationary process.

Autoregressive (AR) Process

- We want to model the conditional expectation of $y_t$:
  \[
  E[y_t | F_{t-1}]
  \]
  where $F_{t-1} = \{y_{t-1}, y_{t-2}, y_{t-3}, \ldots\}$ is the past history of the series. We assume the error term, $\varepsilon_t = y_t - E[y_t | F_{t-1}]$, follows a $\text{WN}(0, \sigma^2)$.

- The most common models are AR models. An AR(1) model involves a single lag, while an AR($p$) model involves $p$ lags.

**Example:** A linear AR($p$) model (the most popular in practice):

\[
y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \text{WN}.
\]

Using lag operator we write: \[
\phi(L) y_t = \varepsilon_t
\]
with \[
\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p
\]
AR(1) Process

• An AR(1) model:
  \[ y_t = \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN. \]

Recall that in a previous example, under the stationarity condition \(|\phi_1| < 1\), we derived the moments:

\[ E[y_t] = \mu = 0 \quad \text{(assuming } \phi_1 \neq 1) \]
\[ \text{Var}[y_t] = \gamma(0) = \sigma^2/(1 - \phi_1^2) \quad \text{(assuming } |\phi_1| < 1) \]
\[ \gamma(1) = E[y_t y_{t-1}] = E[(\phi_1 y_{t-1} + \varepsilon_t) * y_{t-1}] = \phi_1 \gamma(0) \]
\[ \gamma(2) = E[y_t y_{t-2}] = E[(\phi_1 y_{t-1} + \varepsilon_t) * y_{t-2}] = \phi_1^2 \gamma(0) \]
\[ \gamma(3) = E[y_t y_{t-3}] = E[(\phi y_{t-1} + \varepsilon_t) * y_{t-3}] = \phi_1 \gamma(2) = \phi_1^3 \gamma(0) \]

\[ \ldots \]
\[ \gamma(k) = \phi_1 \gamma(k-1) = \phi_1^k \gamma(0) \]

AR(1) Process

• We want to derive the autocorrelation:
  \[ \rho(t_1, t_2) = \frac{\gamma(t_1 - t_2)}{\sigma_{t_1} \sigma_{t_2}} \]

If the process is stationary (\(\sigma_t = \sigma_{t-1} = \sqrt{\gamma(0)}\))

\[ \rho(1) = \rho(t, t - 1) = \frac{\gamma(1)}{\sigma_t \sigma_{t-1}} = \frac{\gamma(1)}{\gamma(0)} = \phi_1 \]
\[ \rho(2) = \frac{\gamma(2)}{\gamma(0)} = \phi_1^2 \]
\[ \ldots \]
\[ \rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi_1^k \]

Remark: The ACF decays with \(k\).

When we plot \(\rho(k)\) against \(k\), we plot also \(\rho(0)\) which is 1.

• Note that when \(\phi_1 = 1\), the AR(1) is non-stationary, \(\rho(k) = 1\), for all \(k\). The present and the past are always correlated!
Example: A process with $|\phi_1| < 1$ (actually, 0.065) is the monthly changes in the USD/GBP exchange rate. Below we plot its corresponding ACF:

![USD/GBP Exchange Rate: Monthly Changes Rates (1971-2020)](image)

Example: Below we plot the monthly changes in the USD/GBP exchange rate. Stationary series do not look smooth:

![USD/GBP Exchange Rate: Monthly Changes Rates (1971-2020)](image)
Example: A process with $\phi_1 \approx 1$ (actually, 0.99) is the nominal USD/GBP exchange rate. Below, we plot the ACF, it is not 1 all the time, but its decay is very slow (after 30 months, it is still .40 correlated!):

![ACF of USD/GBP Exchange Rate](image)

**AR(1) Process**

Example: Below we plot the nominal USD/GBP exchange rate. Stationary series look smooth, smooth enough that you can clearly spot trends:

![USD/GBP Exchange Rate](image)
AR(2) Process

• An AR(2) model:
  \[ y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \quad \varepsilon_t \sim \text{WN}. \]

• Moments:
  \[ E[y_t] = \mu/(1 - \phi_1 - \phi_2) = 0 \quad (\text{assuming } \phi_1 + \phi_2 \neq 1) \]
  \[ \text{Var}[y_t] = \sigma^2/(1 - \phi_1^2 - \phi_2^2) \quad (\text{assuming } \phi_1^2 + \phi_2^2 < 1) \]

  Stationarity condition: \( \phi_1^2 + \phi_2^2 < 1 \).

• Autocovariance function
  \[ \gamma(k) = \text{Cov}[y_{t+k}, y_t] = E[(\phi_1 y_{t+k} + \phi_2 y_{t+k-2} + \varepsilon_t) y_t] \]
  \[ = \phi_1 E[y_{t+k}] + \phi_2 E[y_{t+k-2}] + E[\varepsilon_t y_t] \]
  \[ = \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) + E[\varepsilon_t y_t] \]

  • We have a recursive formula.

AR(2) Process

• We have a recursive formula:
  \( k=0 \)
  \[ \gamma(0) = \phi_1 \gamma(-1) + \phi_2 \gamma(-2) + E[\varepsilon_t y_t] \]
  \[ = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \sigma^2 \]

  \( k=1 \)
  \[ \gamma(1) = \phi_1 \gamma(0) + \phi_2 \gamma(2) + E[\varepsilon_t y_{t-1}] \]
  \[ = \phi_1 \gamma(1) + \phi_2 \gamma(1) \]
  \[ \Rightarrow \gamma(1) = [\phi_1/(1 - \phi_2)] \gamma(0) \]

  \( k=2 \)
  \[ \gamma(2) = \phi_1 \gamma(1) + \phi_2 \gamma(0) + E[\varepsilon_t y_{t-2}] \]
  \[ = \phi_1 \gamma(1) + \phi_2 \gamma(0) \]
  \[ \Rightarrow \gamma(2) = [\phi_1^2/(1 - \phi_2)] + \phi_2 \gamma(0) \]

Replacing \( \gamma(1) \) and \( \gamma(2) \) back to \( \gamma(0) \):

\[ \gamma(0) = [\phi_1^2/(1 - \phi_2)] \gamma(0) + [\phi_2 \phi_1^2/(1 - \phi_2) + \phi_2^2] \gamma(0) + \sigma^2 \]

\[ = \left( \frac{\sigma^2}{1 - \phi_2} \right) \]

\[ = \left( \frac{1 - \phi_2}{(1 - \phi_2) - \phi_1^2(1 + \phi_2) + \phi_2^2(1 - \phi_2)} \right) \]
**AR(2) Process**

- Dividing the previous formulas by \( \gamma(0) \), we get the ACF:
  \[
  \rho(k) = \frac{\gamma(k)}{\gamma(0)} = \frac{\phi_1 \rho(k-1) + \phi_2 \rho(k-2) + \mathbb{E}[\varepsilon_t y_{t-k}]}{\gamma(0)}
  \]

\( (k=0) \) \( \rho(0) = 1 \)

\( (k=1) \) \( \rho(1) = \phi_1 / (1 - \phi_2) \)

\( (k=2) \) \( \rho(2) = \phi_1 \rho(1) + \phi_2 \rho(0) = \phi_1^2 / (1 - \phi_2) + \phi_2 \)

\( (k=3) \) \( \rho(3) = \phi_1 \rho(2) + \phi_2 \rho(1) = \phi_1^3 / (1 - \phi_2) + \phi_1 \phi_2 + \phi_2 \phi_1 / (1 - \phi_2) \)

**Remark:** Again, we see exponential decay in the ACF.

**AR Process – Stationarity: Technical Conditions**

- Let’s compute moments of \( y_t \) using the infinite sum (assume \( \mu = 0 \)):
  \[
  E[y_t] = \phi(L)^{-1} E[\varepsilon_t] = 0 \quad \Rightarrow \phi(L) \neq 0
  \]
  \[
  \text{Var}[y_t] = \phi(L)^{-2} \text{Var}[\varepsilon_t] \quad \Rightarrow \phi(L)^{-2} > 0
  \]
  \[
  E[y_t y_{t-j}] = \gamma(j) = E[(\phi_1 y_{t-1} y_{t-j} + \phi_2 y_{t-2} y_{t-j} + \ldots + \phi_p y_{t-p} y_{t-j} + \varepsilon_{t-j})]
  \]
  \[
  = \phi_1 \gamma(j-1) + \phi_2 \gamma(j-2) + \ldots + \phi_p \gamma(j-p)
  \]

where, abusing notation,

\[
\phi(L)^{-2} = 1 / (1 - \phi_1^2 L - L^2 \phi_2^2 L^2 - \ldots - \phi_p^2 L^p)
\]

Using the fundamental theorem of algebra, \( \phi(z) \) can be factored as

\[
\phi(z) = (1 - r_1^{-1} z)(1 - r_2^{-1} z)\ldots(1 - r_p^{-1} z)
\]

where the \( r_1, \ldots, r_k \in C \) are the roots of \( \phi(z) \). If the \( \phi \)'s coefficients are all real, the roots are either real or come in complex conjugate pairs.

**Example:** The root \( r_1 \) of polynomial \( \phi(z) = 1 - \phi_1 z \) satisfies \( |r_1| > 1 \).
Theorem: The linear AR(p) process is strictly stationary and ergodic if and only if \( |\tau_j| > 1 \) for all \( j \), where \( |\tau_j| \) is the modulus of the complex number \( \tau_j \).

- We usually say “all roots lie outside the unit circle.”

Note: If one of the \( \tau_j \)’s equals 1, \( \phi(L) \) (\( \& y_t \)) has a unit root —i.e., \( \phi(1) = 0 \). This is a special case of non-stationarity.

- Recall \( \phi(L)^{-1} \) produces an infinite sum on the \( \varepsilon_{t-j} \)’s. If this sum does not explode, we say the process is stable.

- If the process is stable, we can calculate \( \delta y_t / \delta \varepsilon_{t-j} \).

\( \delta y_t / \delta \varepsilon_{t-j} \) = How much \( y_t \) is affected today by an innovation (a shock) \( t-j \) periods ago. When expressed as a function of \( j \), we call this the impulse response function (IRF).

**AR Process – Stationarity: Technical Conditions**

**Example**: AR(1) process

\[ y_t = \mu + \phi y_{t-1} + \varepsilon_t \]

\[ E[y_t] = E[\mu + \varepsilon_t] = \frac{\mu}{1 - \phi} = \mu \quad \Rightarrow \phi \neq 1 \quad (r_1 
eq 0) \]

\[ Var[y_t] = Var[\varepsilon_t] = \frac{\sigma^2}{1 - \phi^2}; \quad \text{since} \quad \sigma^2 > 0 \Rightarrow |\phi| < 1 \quad (r_1 > 1) \]

Note: \( 1/(1 - \phi^i) = \sum_{j=0}^{\infty} \phi^j \quad i = 1, 2 \)

These infinite sums will not explode (stable process) if \( |\phi| < 1 \Rightarrow \text{stationarity condition} \).

Under this condition, we can calculate the impulse response function:

\[ \delta y_t / \delta \varepsilon_{t-j} = \phi^j \]
We derived the autocovariance function, $\gamma(k)$, before:

$$\gamma(k) = \phi_1 \gamma(k - 1) = \phi_1^k \gamma(0)$$

Again, when $|\phi| < 1$, the autocovariance do not explode as $k$ increases. There is an exponential decay towards zero.

Note:
- when $0 < \phi_1 < 1$ \Rightarrow All autocovariances are positive.
- when $-1 < \phi_1 < 0$ \Rightarrow The sign of $\gamma(k)$ shows an alternating pattern beginning a negative value.

**AR Process – Example: AR(1)**

**Example:** AR(2) process

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \Rightarrow (1 - \phi_1 L - \phi_2 L^2) y_t = \mu + \epsilon_t$$

We can invert $(1 - \phi_1 L - \phi_2 L^2)$ to get the MA($\infty$) process.

- Stationarity Check
  - $E[y_t] = \mu / (1 - \phi_1 - \phi_2) = \mu^* \Rightarrow \phi_1 + \phi_2 \neq 1$.
  - $\text{Var}[y_t] = \sigma^2 / (1 - \phi_1^2 - \phi_2^2) \Rightarrow \phi_1^2 + \phi_2^2 < 1$

Stationarity condition: $|\phi_1 + \phi_2| < 1$

- Things can be simplified by rewriting the AR(2) in matrix form as an AR(1):

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \mu \\ 0 \end{bmatrix} + \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix} \Rightarrow \tilde{y}_t = \tilde{\mu} + A \tilde{y}_{t-1} + \tilde{\epsilon}_t$$

**Note:** Now, we check $[I - A^i]$ ($i = 1, 2$) for stationarity conditions
AR Process – AR(2): Technical Condition

\[ \tilde{y}_t = \tilde{\mu} + A\tilde{y}_{t-1} + \tilde{\epsilon}_t \quad \Rightarrow \quad \tilde{y}_t = [I - AL]^{-1} \tilde{\epsilon}_t \]

Note: Recall the expansion:
\[ (I - F)^{-1} = \sum_{j=0}^{\infty} F^j = I + F + F^2 + \ldots \]

Checking that \([I - AL]\) is not singular, same as checking that \(A\) does not explode. The stability of the system can be determined by the eigenvalues of \(A\). That is, get the \(\lambda_i\)'s and check if \(|\lambda_i| < 1\) for all \(i\).

\[ A = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \Rightarrow |A - \lambda I| = \det \begin{bmatrix} \phi_1 - \lambda & \phi_2 \\ 1 & -\lambda \end{bmatrix} = -(\phi_1 - \lambda)\lambda - \phi_2 \]

\[ \lambda_i = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2} \]

- If \(|\lambda_i| < 1\) for all \(i=1, 2\), \(y_t\) is stable (it does not explode) and stationary. It can be shown:
  \[ \lambda_1\lambda_2 = \phi_2 \Rightarrow |\lambda_1\lambda_2| = |\phi_2| < 1 \]
  \[ \lambda_1 + \lambda_2 = \phi_1 \Rightarrow |\lambda_1 + \lambda_2| = |\phi_1| < 2 \]

AR Process – Stationarity

- We derived autocovariance function, \(\gamma(k)\), before, getting a recursive formula. Let’s write the first autocovariances:

  \[ \langle k=0 \rangle \quad \gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \sigma^2 \]
  \[ \langle k=1 \rangle \quad \gamma(1) = [\phi_1/(1 - \phi_2)] \gamma(0) \]
  \[ \langle k=2 \rangle \quad \gamma(2) = [\phi_1^2/(1 - \phi_2) + \phi_2] \gamma(0) \]

- With \(|\phi_2| < 1\), we get well defined \(\gamma(1), \gamma(2)\) & \(\gamma(0)\).
AR Process – Stationarity

• The AR(2) in matrix AR(1) form is called Vector AR(1) or VAR(1).

Nice property: The VAR(1) is Markov - i.e., forecasts depend only on today’s data.

• It looks complicated, but it is straightforward to apply the VAR formulation to any AR(p) processes. We can also use the same eigenvalue conditions to check the stationarity of AR(p) processes.

AR Process – Causality

• The AR(p) model:

\[ \phi(L)y_t = \mu + \varepsilon_t \]

where \( \phi(L) = 1 - \phi_1 L^1 - \phi_2 L^2 - \ldots - \phi_p L^p \)

Then, \( y_t = \phi(L)^{-1}(\mu + \varepsilon_t) \), \( \Rightarrow \) an MA(\( \infty \)) process!

• But, we need to make sure that we can invert the polynomial \( \phi(L) \).

• When \( \phi(L) \neq 0 \), we say the process \( y_t \) is causal (strictly speaking, a causal function of \{\( \varepsilon_t \}\}).

Definition: A linear process \{\( y_t \)\} is causal if there is a

\[ \psi(L) = 1 + \psi_1 L + \psi_2 L^2 + \ldots \]

with \( \sum_{j=0}^{\infty} |\psi_j(L)| < \infty \)

with \( y_t = \psi(L)\varepsilon_t \).
**AR Process – Causality**

**Example:** AR(1) process:

\[ \phi(L)y_t = \mu + \varepsilon_t, \quad \text{where } \phi(L) = 1 - \phi_1 L \]

Then, \( y_t \) is causal if and only if:

\[ |\phi_1| < 1 \]

or

the root \( r_1 \) of the polynomial \( \phi(z) = 1 - \phi_1 z \) satisfies \( |r_1| > 1 \).

**Q:** How do we calculate the \( \Psi \)'s coefficients for an AR\((p)\)?

**A:** Matching coefficients:

\[
(Y_t - \mu) = \frac{1}{(1 - \phi_1 L)} \varepsilon_t = \sum_{i=0}^{\psi_1} \phi_i^i L^i \varepsilon_t = \left( 1 + \phi_1 L + \phi_1^2 L^2 + \cdots \right) \varepsilon_t \quad \Rightarrow \quad \Psi_i = \phi_1^i, \quad i \geq 0
\]

**AR Process – Estimation and Properties**

- Define

\[ x_t = (1 \ y_{t-1} \ y_{t-2} \ldots y_{t-p}) \]
\[ \beta = (\mu \ \phi_1 \ \phi_2 \ \ldots \ \phi_p) \]

Then the model can be written as

\[ y_t = x_t' \beta + \varepsilon_t \]

- The OLS estimator is

\[ b = \hat{\beta} = (X'X)^{-1}X'y \]

- Properties:
  - Using the Ergodic Theorem, OLS estimator is consistent.
  - Using the MDS CLT, OLS estimator is asymptotically normal.

\[ \Rightarrow \text{asymptotic inference is the same.} \]

- The asymptotic covariance matrix is estimated just as in the cross-section case: The sandwich estimator.
**ARMA Process**

- A combination of AR(\(p\)) and MA(\(q\)) processes produces an ARMA(\(p, q\)) process:

\[
y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}
\]

\[
= \mu + \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{q} \theta_i L^i \epsilon_i + \epsilon_t
\]

\[
\Rightarrow \phi(L)y_t = \mu + \theta(L)\epsilon_t
\]

- Usually, we insist that \(\phi(L) \neq 0\), \(\theta(L) \neq 0\) & that the polynomials \(\phi(L), \theta(L)\) have no common factors. This implies it is not a lower order ARMA model.

**Example:** Common factors.

Suppose we have the following ARMA(2, 3) model \(\phi(L)y_t = \theta(L)\epsilon_t\) with

\[
\phi(L) = 1 - .6L + .3L^2
\]

\[
\theta(L) = 1 - 1.4L + .9L^2 - .3L^3 = (1 - .6L + .3L^2)(1 - L)
\]

This model simplifies to: \(y_t = (1-L)\epsilon_t \Rightarrow \text{an MA(1) process.}\)

- Pure AR Representation: \(\Pi(L)(y_t - \mu) = a_t \Rightarrow \Pi(L) = \frac{\phi_p(L)}{\theta_q(L)}\)

- Pure MA Representation: \((y_t - \mu) = \Psi(L)a_t \Rightarrow \Psi(L) = \frac{\theta_q(L)}{\phi_p(L)}\)

- Special ARMA(\(p, q\)) cases:
  \(- p = 0: \text{MA}(q)\)
  \(- q = 0: \text{AR}(p)\).
ARMA: Stationarity, Causality and Invertibility

**Theorem:** If $\phi(L)$ and $\theta(L)$ have no common factors, a (unique) stationary solution to $\phi(L)y_t = \theta(L)e_t$ exists if and only if

$$|z| = 1 \Rightarrow \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p \neq 0.$$ 

This ARMA($p$, $q$) model is causal if and only if

$$|z| \leq 1 \Rightarrow \phi(z) = 1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p \neq 0.$$ 

This ARMA($p$, $q$) model is invertible if and only if

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

*Note:* Real data cannot be exactly modeled using a finite number of parameters. We choose $p$, $q$ to create a good approximated model.

ARMA Process – Dynamic Multiplier

• The dynamic multiplier measures the effect of $\varepsilon_t$ on subsequent values of $x_t$: That is, the first derivative on the Wold representation:

$$\delta x_{t+j}/\delta \varepsilon_t = \delta x_j/\delta \varepsilon_0 = \psi_j.$$ 

For an AR(1) process:

$$\delta x_{t+j}/\delta \varepsilon_t = \delta x_j/\delta \varepsilon_0 = \phi^j.$$ 

• That is, the dynamic multiplier for any linear SDE depends only on the length of time $j$, not on time $t$. 
The impulse-response function (IRF) is a sequence of dynamic multipliers as a function of time from the one time change in the innovation, $\varepsilon_t$.

Usually, IRF are represented with a graph, that measures the effect of the innovation, $\varepsilon_t$, on $y_t$ over time:

$$\delta y_{t+j}/\delta \varepsilon_t + \delta y_{t+j+1}/\delta \varepsilon_t + \delta y_{t+j+2}/\delta \varepsilon_t + \ldots = \psi_j + \psi_{j+1} + \psi_{j+2} + \ldots$$

Once we estimate the ARMA coefficients, it is easy to draw an IRF.

**ARMA Process – Addition**

Q: We add two ARMA processes, what order do we get?

Adding MA processes

\[
\begin{align*}
  x_t &= A(L)\varepsilon_t \\
  z_t &= C(L)u_t \\
  y_t &= x_t + z_t = A(L)\varepsilon_t + C(L)u_t
\end{align*}
\]

Under independence:

\[
\gamma_y(j) = E[y_t y_{t-j}] = E[(x_t + z_t)(x_{t-j} + z_{t-j})] = E[(x_t x_{t-j} + z_t z_{t-j})] = \gamma_x(j) + \gamma_z(j)
\]

Then, $\gamma(j) = 0$ for $j > \text{Max}(q_x, q_z) \implies y_t$ is ARMA($0, \text{Max}(q_x, q_z)$)

Implication: MA(2) + MA(1) = MA(2)
ARMA Process – Addition

- Q: We add two ARMA process, what order do we get?

- Adding AR processes

\[
(1 - A(L))x_i = \epsilon_i \\
(1 - C(L))z_i = u_i \\
y_i = x_i + z_i = ?
\]

- Rewrite system as:

\[
(1 - C(L))(1 - A(L))x_i = (1 - C(L))\epsilon_i \\
(1 - A(L))(1 - C(L))z_i = (1 - A(L))u_i \\
(1 - A(L))(1 - C(L))y_i = (1 - C(L))\epsilon_i + (1 - A(L))u_i = \epsilon_i + u_i - [C(L)\epsilon_i + A(L)u_i]
\]

- Then, \( y_i \) is ARMA\((p_x + p_z, \max(p_x, p_z))\)