Lecture 14
ARIMA – Identification, Estimation & Seasonalities

ARMA Process

• We defined the ARMA(p, q) model:

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

Let \( x_t = y_t - \mu \).

Then,

\[ \phi(L)x_t = \theta(L)\epsilon_t \]

\( \Rightarrow x_t \) is a demeaned ARMA process.

• In this lecture, we will study:
  - Identification of \( p, q \).
  - Estimation of ARMA(p,q)
  - Non-stationarity of \( x_t \)
  - Differentiation issues – ARIMA(p,d,q)
  - Seasonal behavior – SARIMA(p,d,q)_S
Autocovariance Function

- We define the autocovariance function, \( \gamma(t-j) \) as: \( \gamma(t-j) = E[y_t y_{t-j}] \)

- For an AR(\( p \)) process, WLOG with \( \mu = 0 \) (or demeaned \( y_t \)), we get:
  \[
  \gamma(t-j) = E[y_t y_{t-j}] = E[(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t) y_{t-j}]
  = \phi_1 \gamma(j-1) + \phi_2 \gamma(j-2) + \ldots + \phi_p \gamma(j-p)
  \]

Notation: \( \gamma(k) \) or \( \gamma_k \) are commonly used. Sometimes, \( \gamma(k) \) is referred as “covariance at lag \( k \).

- The \( \gamma(t-j) \) determine a system of equations:

\[
\begin{align*}
\gamma(0) &= E[y_t y_t] = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \ldots + \phi_p \gamma(p) + \sigma^2 \\
\gamma(1) &= E[y_t y_{t-1}] = \phi_1 \gamma(0) + \phi_2 \gamma(1) + \phi_3 \gamma(2) + \ldots + \phi_p \gamma(p-1) \\
\gamma(2) &= E[y_t y_{t-2}] = \phi_1 \gamma(1) + \phi_2 \gamma(0) + \phi_3 \gamma(1) + \ldots + \phi_p \gamma(p-2) \\
& \vdots \\
\gamma(p) &= E[y_t y_{t-p}] = \phi_1 \gamma(p-1) + \phi_2 \gamma(p-2) + \ldots + \phi_p \gamma(0)
\end{align*}
\]

- These are the Yule-Walker equations, can be solved numerically. MM can be used (replace population moments with sample moments).

- Properties for a stationary time series
  1. \( \gamma(0) \geq 0 \) (from definition of variance)
  2. \( \gamma(k) \leq \gamma(0) \) (from Cauchy-Schwarz)
  3. \( \gamma(k) = \gamma(-k) \) (from stationarity)
  4. \( \Gamma \), the auto-correlation matrix, is psd \( (a' \Gamma a \geq 0) \)

Moreover, any function \( \gamma: Z \to R \) that satisfies (3) and (4) is the autocovariance of some stationary time series.
Autocovariance Function – Example: ARMA(1,1)

• For an ARMA(1,1) we have:

\[
\gamma_k = E[(y_t - \mu)(y_{t-k} - \mu)] \\
= E[\phi(y_{t-1} - \mu) \epsilon_t - \theta \epsilon_{t-1} (y_{t-k} - \mu)] \\
= \phi E[(y_{t-1} - \mu)(y_{t-k} - \mu)] + E[\epsilon_t (y_{t-k} - \mu)] - \theta E[\epsilon_{t-1} (y_{t-k} - \mu)] \\
= \phi \gamma_{k-1} + E[\epsilon_t (y_{t-k} - \mu)] - \theta E[\epsilon_{t-1} (y_{t-k} - \mu)]
\]

\[
\gamma_0 = \phi \gamma_{-1} + E[\epsilon_t (y_t - \mu)] - \theta E \left[ \frac{\epsilon_{t-1} (y_{t-1} - \mu)}{\phi (y_{t-1} - \mu) + \theta \epsilon_{t-2}} \right] \\
= \phi \gamma_{-1} + \sigma^2 - \theta E \left[ \frac{\epsilon_{t-1} \phi (y_{t-1} - \mu) + \epsilon_t - \theta \epsilon_{t-2}}{\phi (y_{t-1} - \mu) + \theta \epsilon_{t-2}} \right] \\
= \phi \gamma_{-1} + \sigma^2 - \theta (\phi \sigma^2 - \theta \sigma^2)
\]

Autocovariance Function – Example: ARMA(1,1)

• Similarly:

\[
\gamma_1 = \phi \gamma_0 + E[\epsilon_t (y_{t-1} - \mu)] - \theta E \left[ \frac{\epsilon_{t-1} (y_{t-1} - \mu)}{\phi (y_{t-1} - \mu) + \theta \epsilon_{t-2}} \right] \\
= \phi \gamma_0 - \theta \sigma^2
\]

• Two equations for \(\gamma_0\) and \(\gamma_1\):

\[
\Rightarrow \gamma_0 = \phi (\phi \gamma_0 - \theta \sigma^2) + \sigma^2 (1 - \phi \theta + \theta^2) \\
\gamma_0 = \frac{1 + \theta^2 - 2 \phi \theta}{1 - \theta^2}
\]

• In general:

\[
\gamma_k = \phi \gamma_{k-1} = \phi^{k-1} \gamma_1, \quad k > 1
\]
**Autocorrelation Function (ACF)**

- Now, we define the autocorrelation function (ACF):
  \[ \rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{covariance at lag } k}{\text{variance}} \]
  
The ACF lies between -1 and +1, with \( \rho(0) = 1 \).

- Dividing the autocovariance system by \( \gamma(0) \), we get:
  \[
  \begin{bmatrix}
  \rho(0) & \rho(1) & \cdots & \rho(p-1) \\
  \rho(1) & \rho(0) & \cdots & \rho(p-2) \\
  \vdots & \vdots & \ddots & \vdots \\
  \rho(p-1) & \rho(p-2) & \cdots & \rho(0)
  \end{bmatrix}
  \begin{bmatrix}
  \phi_1 \\
  \phi_2 \\
  \vdots \\
  \phi_p
  \end{bmatrix}
  =
  \begin{bmatrix}
  \rho(1) \\
  \rho(2) \\
  \vdots \\
  \rho(p)
  \end{bmatrix}
  \]

- These are “Yule-Walker” equations, which can be solved numerically.

---

**Autocorrelation Function (ACF)**

- Easy estimation: Use sample moments to estimate \( \gamma(k) \) and plug in formula:
  \[ r_k = \sum (Y_i - \bar{Y})(Y_{i+k} - \bar{Y}) \sum (Y_i - \bar{Y})^2 \]

- Distribution: For a linear, stationary, \( y_t = \mu + \sum \psi_j \varepsilon_{t-j}, \) with \( \text{E}[\varepsilon_t^4] < \infty \):
  \[ r \rightarrow Np, V/T, \]
  \( V \) is the covariance matrix with elements:
  \[ w_{ij} = \sum_{k=1}^{\infty} \{ \rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k) \} \times \{ \rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k) \} \]
  If \( \rho_k = 0 \), for all \( k \neq 0 \), \( V = I \) \( \Rightarrow \text{Var}[r(k)] = 1/T \).

- The standard errors under \( H_0, 1/\sqrt{T} \), are sometimes referred as \( \text{Bartlett's SE} \).
**Autocorrelation Function (ACF)**

- **Example:** Sample ACF for an MA(1) process.
  \[
  \begin{align*}
  \rho(0) &= 1, \\
  \rho(k) &= \frac{\theta}{1+\theta^2}, \quad \text{for } k = 1, -1 \\
  \rho(k) &= 0 \quad \text{for } |k| > 1.
  \end{align*}
  \]

  \[
  V_{1,1} = \sum_{h=1}^{\infty} (\rho(h+1) + \rho(h-1) - 2\rho(1)\rho(h))^2 = (\rho(0) - 2\rho(1)^2)^2 + \rho(1)^2
  \]

  \[
  V_{2,2} = \sum_{h=1}^{\infty} (\rho(h+2) + \rho(h-2) - 2\rho(2)\rho(h))^2 = \sum_{h=-1}^{1} \rho(h)^2.
  \]

  And if \( \hat{\rho} \) is the sample ACF from a realization of this MA(1) process, then with probability 0.95,

  \[
  |\hat{\rho}(h) - \rho(h)| \leq 1.96\sqrt{\frac{V_{hh}}{n}}.
  \]

**Autocorrelation Function (ACF)**

- **Example:** Sample ACF for an MA(\( q \)) process:
  \[
  y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}
  \]

  \[
  \begin{align*}
  \gamma(0) &= \mu^2 + \varepsilon_t^2 + \theta_1 \varepsilon_{t-1}^2 + \theta_2 \varepsilon_{t-2}^2 + \ldots + \theta_q \varepsilon_{t-q}^2 \\
  \gamma(1) &= \mu \varepsilon_t + \mu \varepsilon_{t-1} + (\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}) \\
  \gamma(2) &= \mu \varepsilon_{t-1} + \mu \varepsilon_{t-2} + (\theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \ldots + \theta_q \varepsilon_{t-q}) \\
  \gamma(q) &= 0
  \end{align*}
  \]

  In general, \( \gamma(k) = \sigma^2 \sum_{j=k}^{q} \theta_j \theta_{j-k} \quad k \leq q \quad \text{(with } \theta_0 = 1) \).

  \[
  = 0 \quad \text{otherwise.}
  \]

  Then, \[
  \rho(k) = \frac{\sigma^2 \sum_{j=k}^{q} \theta_j \theta_{j-k}}{(1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)} \quad k \leq q
  \]

  \[
  = 0 \quad \text{otherwise.}
  \]
Autocorrelation Function - Example

- **Example**: US Monthly Returns (1800 – 2013, $T=2534$)

Autocorrelations and Ljung Box Tests (SAS: Check for White Noise)

<table>
<thead>
<tr>
<th>Lag</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
<th>Autocorrelations</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>38.29</td>
<td>6</td>
<td>&lt;.0001</td>
<td>0.096 -0.017</td>
</tr>
<tr>
<td>12</td>
<td>46.89</td>
<td>12</td>
<td>&lt;.0001</td>
<td>0.013 0.032</td>
</tr>
<tr>
<td>18</td>
<td>75.07</td>
<td>18</td>
<td>&lt;.0001</td>
<td>-0.042 -0.084</td>
</tr>
<tr>
<td>24</td>
<td>89.77</td>
<td>24</td>
<td>&lt;.0001</td>
<td>-0.009 -0.054</td>
</tr>
</tbody>
</table>

$SE(r_k) = 1/\sqrt{T} = 1/\sqrt{2534} \approx .02.$

- **Note**: Small correlations, but because $T$ is large, many of them significant (even at $k=20$ months). Joint significant too!

ACF: Correlogram

- **The sample correlogram is the plot of the ACF against $k$.**

As the ACF lies between -1 and +1, the correlogram also lies between these values.

- **Example**: Correlogram for US Monthly Returns (1800 – 2013)
ACF: Joint Significance Tests

• Recall the Q statistic as:

\[ Q = T \sum_{k=1}^{m} \hat{\rho}_k^2 \]

It can be used to determine if the first \( m \) sample ACFs are jointly equal to zero. Under \( H_0: \rho_1=\rho_2=...=\rho_m=0 \), then \( Q \) converges in distribution to a \( \chi^2(m) \).

• The Ljung-Box statistic is similar, with the same asymptotic distribution, but with better small sample properties:

\[ LB = T ( T + 2 ) \sum_{k=1}^{m} \left( \frac{\hat{\rho}_k^2}{T - k} \right) \]

Partial ACF

• The Partial Autocorrelation Function (PACF) is similar to the ACF. It measures correlation between observations that are \( k \) time periods apart, after controlling for correlations at intermediate lags.

Definition: The PACF of a stationary time series \( \{y_t\} \) is

\[ \phi_{11} = \text{Corr}(y_t, y_{t+1}) = \rho(1) \]
\[ \phi_b = \text{Corr}(y_t - E[y_t | F_{t-b}], y_{t-b} - E[y_{t-b} | F_{t-b}]) \quad \text{for} \quad b = 2, 3, ... \]

This removes the linear effects of \( y_{t-1}, y_{t-2}, ..., y_{t-b} \).

• The PACF \( \phi_b \) is also the last coefficient in the best linear prediction of \( y_t \) given \( y_{t-1}, y_{t-2}, ..., y_{t-b} \): \( \Gamma_k \phi_b = \gamma(k) \)

where \( \phi_b = (\phi_{b1}, \phi_{b2}, ..., \phi_{bb}) \).
**Partial ACF**

**Example:** AR(p) process:

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

\[ E[y_t | F_{t-1}] = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} \]

Then, \[ \phi_{bb} = \phi_b \quad \text{if } 1 \leq b \leq p \]

\[ = 0 \quad \text{otherwise} \]

- Estimation: \[ \hat{\phi}_h = [\hat{\Phi}]^{-1}\hat{\gamma}(k) \quad \Rightarrow \text{a recursive system.} \]

A recursive algorithm, Durbin-Levinson, can be used.

- We can also produce a *partial correlogram*, which is used in Box-Jenkins methodology (covered later).

---

**Inverse ACF (IACF)**

- The IACF of the ARMA\((p,q)\) model

\[ \phi(L)y_t = \theta(L)\epsilon_t \]

is defined to be (assuming invertibility) the ACF of the *inverse* (or *dual*) process

\[ \theta(L)y_t^{-1} = \phi(L)\epsilon_t \]

- The IACF has the same property as the PACF: AR\((p)\) is characterized by an IACF that is nonzero at lag \(p\) but zero for larger lags.

- The IACF can also be used to detect over-differencing. If the data come from a nonstationary or nearly nonstationary model, the IACF has the characteristics of a noninvertible moving-average.
ACF, Partial ACF & IACF: Example

Example: Monthly USD/GBP 1st differences (1800-2013)

Non-Stationary Time Series Models

- The ACF is as a rough indicator of whether a trend is present in a series. A slow decay in ACF is indicative of a large characteristic root; a true unit root process, or a trend stationary process.

- Formal tests can help to determine whether a system contains a trend and whether the trend is deterministic or stochastic.

- We will analyze two situations faced in ARMA models:
  (1) Deterministic trend – Simple model: \( y_t = \alpha + \beta t + \epsilon_t \)
    - Solution: Detrending – i.e., regress \( y_t \) on \( t \). Then, keep residuals for further modeling.

  (2) Stochastic trend – Simple model: \( y_t = \epsilon + y_{t-1} + \epsilon_t \)
    - Solution: Differencing – i.e., apply \( \Delta = (1 - L) \) operator to \( y_t \). Then, use \( \Delta y_t \) for further modeling.
Non-Stationary Models: Deterministic Trend

• Suppose we have the following model:
  \[ y_t = \alpha + \beta t + \varepsilon_t \]
  \[ \Rightarrow \Delta y_t = y_t - y_{t-1} = \beta t - \beta (t-1) + \varepsilon_t - \varepsilon_{t-1} = \beta + \varepsilon_t - \varepsilon_{t-1} \]
  \[ \Rightarrow E[\Delta y_t] = \beta \]

• \( \{y_t\} \) will show only temporary departures from the trend line \( \alpha + \beta t \). This type of model is called a trend stationary (TS) model.

• If a series has a deterministic time trend, then we simply regress \( y_t \) on an intercept and a time trend \((t = 1, 2, \ldots, T)\) and save the residuals. The residuals are the detrended \( y_t \) series.

• If \( y_t \) is stochastic, we do not necessarily get stationary series.

Non-Stationary Models: Deterministic Trend

• Many economic series exhibit “exponential trend/growth”. They grow over time like an exponential function over time instead of a linear function. In this cases, it is common to work with logs
  \[ \ln(y_t) = \alpha + \beta t + \varepsilon_t \]
  \[ \Rightarrow \text{The average growth rate is: } E[\Delta \ln(y_t)] = \beta \]

• We can have a more general model:
  \[ y_t = \alpha + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \beta_1 t + \beta_2 t^2 + \ldots + \beta_{2k} t^k + \varepsilon_t, \]

• Estimation:
  - OLS.
  - Frish-Waugh method:
    (1) Detrend \( y_t \), get the residuals;
    (2) Use residuals to estimate the AR(\( p \)) model.
Non-Stationary Models: Deterministic Trend

• This model has a short memory.

• If a shock (big $\varepsilon_t$) hits $y_t$, it goes back to trend level in a short time. Thus, the best forecasts are not affected.

• In practice, it is not a popular model. A more realistic model involves stochastic (local) trend.

Non-Stationary Models: Stochastic Trend

• The more modern approach is to consider trends in time series as a variable.

• A variable trend exists when a trend changes in an unpredictable way. Therefore, it is considered as stochastic.

• Recall the AR(1) model: $y_t = c + \phi y_{t-1} + \varepsilon_t$.

• As long as $|\phi| < 1$, everything is fine: OLS is consistent, t-stats are asymptotically normal, etc.

• Now consider the extreme case where $\phi = 1$, $\Rightarrow y_t = c + y_{t-1} + \varepsilon_t$.

• Where is the (stochastic) trend? No $t$ term.
Non-Stationary Models: Stochastic Trend

• Let us replace recursively the lag of \( y_t \) on the right-hand side:
\[
y_t = \mu + y_{t-1} + \epsilon_t \\
= \mu + (\mu + y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
\vdots \\
= y_0 + t\mu + \sum_{i=0}^{t-1} \epsilon_{t-i}
\]

• This is what we call a "random walk with drift". The series grows with \( t \).

• Each \( \epsilon_t \) shock represents a shift in the intercept. All values of \{\epsilon_t\} have a 1 as coefficient \( \Rightarrow \) each shock never vanishes (permanent).

Non-Stationary Models: Stochastic Trend

• \( y_t \) is said to have a stochastic trend (ST), since each \( \epsilon_t \) shock gives a permanent and random change in the conditional mean of the series.

• For these situations, we use Autoregressive Integrated Moving Average (ARIMA) models.

• Q: Deterministic or Stochastic Trend?
They appear similar: Both lead to growth over time. The difference is how we think of \( \epsilon_t \). Should a shock today affect \( y_{t+1} \)?

- TS: \( y_{t+1} = \epsilon + \beta(t+1) + \epsilon_{t+1} \) \( \Rightarrow \epsilon_t \) does not affect \( y_{t+1} \).

- ST: \( y_{t+1} = \epsilon + y_t + \epsilon_{t+1} = \epsilon + [\epsilon + y_{t-1} + \epsilon] + \epsilon_{t+1} \) \( \Rightarrow \epsilon_t \) affects \( y_{t+1} \).

(In fact, the shock will have a permanent impact.)
**ARIMA(p,d,q) Models**

- For $p$, $d$, $q \geq 0$, we say that a time series $\{y_t\}$ is an ARIMA $(p,d,q)$ process if $w_t = \Delta^d y_t = (1 - L)^d y_t$ is ARMA$(p,q)$. That is,
  \[ \phi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t \]

- Applying the $(1 - L)$ operator to a time series is called differencing.

- **Notation:** If $y_t$ is non-stationary, but $\Delta^d y_t$ is stationary, then $y_t$ is integrated of order $d$, or $I(d)$. A time series with unit root is $I(1)$. A stationary time series is $I(0)$.

- **Examples:**
  - **Example 1:** RW: $y_t = y_{t-1} + \varepsilon_t$.
    $y_t$ is non-stationary, but $(1 - L) y_t = \varepsilon_t \Rightarrow$ white noise!
    Now, $y_t \sim$ ARIMA$(0,1,0)$.
  - **Example 2:** AR(1) with time trend: $y_t = \mu + \delta t + \phi y_{t-1} + \varepsilon_t$.
    $y_t$ is non-stationary, but $w_t = (1 - L) y_t = \delta + \phi w_{t-1} + \varepsilon_t - \varepsilon_{t-1}$
    Now, $y_t \sim$ ARIMA$(1,1,1)$.

- We call both process first difference stationary.

- **Note:**
  - Example 1: Differencing a series with a unit root in the AR part of the model reduces the AR order.
  - Example 2: Differencing can introduce an extra MA structure. We introduced non-invertibility. This happens when we difference a TS series. Detrending should be used in these cases.
**ARIMA(p,d,q) Models**

- In practice:
  - A root near 1 of the AR polynomial \( \Rightarrow \) differencing
  - A root near 1 of the MA polynomial \( \Rightarrow \) over-differencing

- In general, we have the following results:
  - Too little differencing: not stationary.
  - Too much differencing: extra dependence introduced.

- Finding the right \( d \) is crucial. For identifying preliminary values of \( d \):
  - Use a time plot.
  - Check for slowly decaying (persistent) ACF/PACF.

---

**Example: ACF & Partial ACF**

*Example: Monthly USD/GBP levels (1800-2013)*

Autocorrelation Check for White Noise

<table>
<thead>
<tr>
<th>Lag</th>
<th>Chi-Square</th>
<th>DF</th>
<th>ChiSq</th>
<th>Pr &gt;</th>
<th>Autocorrelations</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>9999.99</td>
<td>6</td>
<td>&lt;.0001</td>
<td>0.995</td>
<td>0.989 0.984 0.979</td>
</tr>
<tr>
<td>12</td>
<td>9999.99</td>
<td>12</td>
<td>&lt;.0001</td>
<td>0.963</td>
<td>0.957 0.952</td>
</tr>
<tr>
<td>18</td>
<td>9999.99</td>
<td>18</td>
<td>&lt;.0001</td>
<td>0.937</td>
<td>0.934 0.931</td>
</tr>
<tr>
<td>24</td>
<td>9999.99</td>
<td>24</td>
<td>&lt;.0001</td>
<td>0.920</td>
<td>0.917 0.914</td>
</tr>
</tbody>
</table>
Example: ACF & Partial ACF

Example: Monthly USD/GBP levels (1800-2013)

• A random walk (RW) is defined as a process where the current value of a variable is composed of the past value plus an error term defined as a white noise (a normal variable with zero mean and variance one).

• RW is an ARIMA(0,1,0) process

$$y_t = y_{t-1} + e_t \Rightarrow \Delta y_t = (1 - L)y_t = e_t, \quad e_t \sim WN\left(0, \sigma^2\right)$$

• Popular model. Used to explain the behavior of financial assets, unpredictable movements (Brownian motions, drunk persons).

• It is a special case (limiting) of an AR(1) process.

• Implication: $E[y_{t+1} | F_t] = y_t \Rightarrow \Delta y_t$ is absolutely random.

• Thus, a RW is nonstationary, and its variance increases with $t$. 

ARIMA(p,d,q) Models – Random Walk
**ARIMA(p,d,q) Models – Random Walk**

- Examples:

\[ y_t - y_{t-1} = \Delta y_t = \mu + \epsilon_t \]

- It can also be written as

\[ y_t = y_0 + t\mu + \sum_{i=0}^{\infty} \epsilon_{t+i} \]

\( \Rightarrow \epsilon_t \) has a permanent effect on the mean of \( y_t \).

- \( \text{E}[y_t] = y_0 + t\mu \) (Unconditional forecast)
- \( \text{E}[y_{t+s} | y_t] = y_t + s\mu \) (Conditional forecast)

**ARIMA(p,d,q) Models – RW with Drift**

- Change in \( Y_t \) is partially deterministic and partially stochastic.

\[ y_t - y_{t-1} = \Delta y_t = \mu + \epsilon_t \]

- It can also be written as

\[ y_t = y_0 + t\mu + \sum_{i=0}^{\infty} \epsilon_{t+i} \]

\( \Rightarrow \epsilon_t \) has a permanent effect on the mean of \( y_t \).

- \( \text{E}[y_t] = y_0 + t\mu \) (Unconditional forecast)
- \( \text{E}[y_{t+s} | y_t] = y_t + s\mu \) (Conditional forecast)
ARIMA(\(p,d,q\)) Models – RW with Drift

- Two series: 1) True JPY/USD 1997-2000 series; 2) A simulated RW (same drift and variance). Can you pick the true series?

ARIMA Models: Box-Jenkins

- An effective procedure for building empirical time series models is the Box-Jenkins approach, which consists of three stages:
  1. Model specification or identification (of ARIMA order),
  2. Estimation
  3. Diagnostics testing.

- Two main approaches to (1) Identification.
  - Correlation approach, mainly based on ACF & PACF.
  - Information criteria, based on the maximized likelihood \((x^2)\) plus a penalty function. For example, model selection based on the AIC.
ARIMA Models: Box-Jenkins

- We have a family of ARIMA models, indexed by $p$, $q$, and $d$.
  Q: How do we select one?

- Box-Jenkins Approach
  1) Make sure data is stationary – check a time plot. If not, differentiate.
  2) Using ACF & PACF, guess small values for $p$ & $q$.
  3) Estimate order $p$, $q$.
  4) Run diagnostic tests on residuals.

  Q: Are they white noise? If not, add lags ($p$ or $q$, or both).

- If order choice not clear, use AIC, AIC Corrected (AICc), BIC, or HQC (Hannan and Quinn (1979)).

- Value parsimony. When in doubt, keep it simple (KISS).

ARIMA Models: Identification - Correlations

- Correlation approach.
  Basic tools: sample ACF and sample PACF.
  - ACF identifies order of MA: Non-zero at lag $q$; zero for lags $> q$.
  - PACF identifies order of AR: Non-zero at lag $p$; zero for lags $> p$.
  - All other cases, try ARMA$(p,q)$ with $p > 0$ and $q > 0$.

For $p > 0$ and $q > 0$:

<table>
<thead>
<tr>
<th></th>
<th>AR$(p)$</th>
<th>MA$(q)$</th>
<th>ARMA$(p,q)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACF</td>
<td>Tails off</td>
<td>Cuts off after lag $q$</td>
<td>Tails off</td>
</tr>
<tr>
<td>PACF</td>
<td>Cuts off after lag $p$</td>
<td>Tails off</td>
<td>Tails off</td>
</tr>
<tr>
<td>IACF</td>
<td>Cuts off after lag $p$</td>
<td>Tails off</td>
<td>Tails off</td>
</tr>
</tbody>
</table>
ARIMA Models: Identification – AR(1)

ARIMA Models: Identification – AR(2)
ARIMA Models: Identification – MA(1)

FIGURE: ACF and PACF of MA(1) processes: $Z_t = (1 - \theta)\epsilon_t$.

ARIMA Models: Identification – MA(2)

FIGURE: ACF and PACF of MA(2) processes: $Z_t = (1 - \theta_1 - \theta_2)\epsilon_t$. 
ARIMA Models: Identification – ARMA(1,1)

FIGURE ACF and PACF of ARMA(1,1) model \((1 - \phi B) \Delta x_t = (1 - \theta D) \epsilon_t\).
ARIMA Models: Identification – ARMA(1,1)


- **Note**: Identification is not clear.

ARIMA Model: Identification - IC

- Popular information criteria
  
  \[ \text{AIC} = -2 \ln(L_{\text{likelihood}}) + 2 \ M \]

  \[ \ln L = -\frac{T}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} S(\theta_\rho, \theta_\theta, \mu) \]

  \[ \ln \hat{L} = -\frac{T}{2} \ln \hat{\sigma}^2 - \frac{T}{2} \left(1 + \ln 2\pi \right) \]

  \[ AIC = T \ln \hat{\sigma}^2 + 2 \ M \]

  \[ BIC = T \ln \hat{\sigma}^2 + M \ln T \]

  \[ HQC = T \ln \hat{\sigma}^2 + 2M \ln (\ln T) \]

- Note: Identification is not clear.
ARIMA Model: Identification - IC

• There is an AIC corrected statistic, that corrects AIC for finite sample sizes:

\[ AIC_c = T \ln \hat{\sigma}^2 + \frac{2M(M+1)}{T-M-1} \]

Burnham & Anderson (2002) recommend using AICc, rather than AIC, if \( T \) is small or \( M \) is large. Since AICc converges to AIC as \( T \) gets large, B&A advocate the AICc.

• For AR(\( p \)), other criteria are possible: Akaike’s final prediction error (FPE), Akaike’s BIC, Parzen’s CAT.

• Hannan and Rissanen’s (1982) \textit{minic}: Calculate the BIC for different \( p \)'s (estimated first) and different \( q \)'s. Select the best model.

Note: Box, Jenkins, and Reinsel (1994) proposed using the AIC above.

ARIMA Model: Identification - IC

**Example:** Monthly US Returns (1800 - 2013) Hannan and Rissanen (1982)’s minic.

**Minimum Information Criterion**

<table>
<thead>
<tr>
<th>Lags</th>
<th>MA 0</th>
<th>MA 1</th>
<th>MA 2</th>
<th>MA 3</th>
<th>MA 4</th>
<th>MA 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR 0</td>
<td>-6.1889</td>
<td><strong>-6.19573</strong></td>
<td>-6.19273</td>
<td>-6.19177</td>
<td>-6.18872</td>
<td>-6.18886</td>
</tr>
<tr>
<td>AR 1</td>
<td>-6.19511</td>
<td>-6.193</td>
<td>-6.19001</td>
<td>-6.18929</td>
<td>-6.18632</td>
<td>-6.18678</td>
</tr>
<tr>
<td>AR 2</td>
<td>-6.19271</td>
<td>-6.18993</td>
<td>-6.1911</td>
<td>-6.18802</td>
<td>-6.18536</td>
<td>-6.1839</td>
</tr>
<tr>
<td>AR 3</td>
<td>-6.19121</td>
<td>-6.18916</td>
<td>-6.18801</td>
<td>-6.18562</td>
<td>-6.18256</td>
<td>-6.18082</td>
</tr>
<tr>
<td>AR 4</td>
<td>-6.18853</td>
<td>-6.18609</td>
<td>-6.18523</td>
<td>-6.18254</td>
<td>-6.17983</td>
<td>-6.17774</td>
</tr>
<tr>
<td>AR 5</td>
<td>-6.18794</td>
<td>-6.18671</td>
<td>-6.18408</td>
<td>-6.18099</td>
<td>-6.1779</td>
<td>-6.17564</td>
</tr>
</tbody>
</table>

**Note:** Best Model is ARMA(0,1).
ARIMA Model: Identification - IC

• There is no agreement on which criteria is best. The AIC is the most popular, but others are also used. (Diebold, for instance, recommends the BIC.)

• Asymptotically, the BIC is consistent –i.e., it selects the true model if, among other assumptions, the true model is among the candidate models considered. For example, Hannan (1980) shows that in the case of common roots in the AR and MA polynomials, the BIC (& HQC) still select the correct orders $p$ and $q$ consistently. But, it is not efficient.

• The AIC is not consistent, generally producing too large a model, but is more efficient –i.e., when the true model is not in the candidate model set the AIC asymptotically chooses whichever model minimizes the MSE/MSPE.

ARIMA Process – Estimation

• We assume:
  - The model order ($d$, $p$ and $q$) is known. Make sure $y_t$ is I(0).
  - The data has zero mean ($\mu=0$). If this is not reasonable, demean $y$.

Fit a zero-mean ARMA model to the demeaned $y_t$:

$$
\phi(L)(y_t - \bar{y}) = \theta(L)\varepsilon_t
$$

• Several ways to estimate an ARMA($p$, $q$) model:

1) MLE. Assume a distribution, usually normal, and do ML.
3) Innovations algorithm for MA($q$).
4) Hannan-Rissanen algorithm for ARMA($p,q$).
ARIMA Process – MLE

- Steps:
  1) Assume a distribution for the errors. Typically, *i.i.d.* normal, say:
     \( \varepsilon_t \sim i.i.d. N(0, \sigma^2) \)
  2) Write down the joint pdf for \( \varepsilon_t \):
     \[
     f(\varepsilon_1, ..., \varepsilon_T) = f(\varepsilon_1) ... f(\varepsilon_T)
     \]
     Note: we are not writing the joint pdf in terms of the \( y_t \)'s, as a multiplication of the marginal pdfs because of the dependency in \( y_t \).
  3) Get \( \varepsilon_t \). For the general stationary ARMA\((p,q)\) model:
     \[
     \varepsilon_t = y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}
     \](if \( \mu \neq 0 \), demean \( y_t \).)
  4) The joint pdf for \( \{\varepsilon_1, ..., \varepsilon_T\} \) is:
     \[
     f(\varepsilon_1, ..., \varepsilon_T | \mu, \phi, \theta, \sigma^2) = \left(2\pi\sigma^2\right)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^{T} \varepsilon_t^2\right\}
     \]

ARIMA Process – MLE

- Steps:
  5) Let \( Y = (y_1, ..., y_T) \) and assume that initial conditions \( Y_* = (y_{1-p}, ..., y_0)' \) and \( \varepsilon_* = (\varepsilon_{1-q}, ..., \varepsilon_0)' \) are known.
  6) The conditional log-likelihood function is given by
     \[
     \ln L(\mu, \phi, \theta, \sigma^2) = -\frac{T}{2} \ln(2\pi\sigma^2) - \frac{S_*(\mu, \phi, \theta)}{2\sigma^2}
     \]
     where \( S_*(\mu, \phi, \theta) = \sum_{t=1}^{T} \varepsilon_t^2(\mu, \phi, \theta | y, y_*, \varepsilon_*) \) is the conditional SS.

Note: Usual Initial conditions: \( y_* = \bar{y} \) and \( \varepsilon_* = E[\varepsilon_t] = 0 \).

- Numerical optimization problem. Initial values (\( y_* \)) matter.
ARIMA Process – MLE: AR(1)

- **Example:** AR(1)
  \[ y_t = \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \]
  - Write down joint for \( \varepsilon_t \)
    \[ f(\varepsilon_1, \ldots, \varepsilon_n) = (2\pi\sigma)^{n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^n \varepsilon_t^2 \right\} \]
  - Solve for \( \varepsilon_t \):
    \[ Y_1 = \phi Y_0 + \varepsilon_1 \rightarrow \text{Let's take } Y_0 = 0 \]
    \[ Y_2 = \phi Y_1 + \varepsilon_2 \Rightarrow \varepsilon_2 = Y_2 - \phi Y_1 \]
    \[ Y_3 = \phi Y_2 + \varepsilon_3 \Rightarrow \varepsilon_3 = Y_3 - \phi Y_2 \]
    \[ \vdots \]
    \[ Y_n = \phi Y_{n-1} + \varepsilon_n \Rightarrow \varepsilon_n = Y_n - \phi Y_{n-1} \]

- **Example:**
  - To change the joint from \( \varepsilon_t \) to \( y_t \), we need the Jacobian \( J_{1,2} \)
    \[ \begin{vmatrix} \frac{\partial \varepsilon_2}{\partial Y_2} & \frac{\partial \varepsilon_2}{\partial Y_3} & \ldots & \frac{\partial \varepsilon_2}{\partial Y_n} \\ \frac{\partial \varepsilon_3}{\partial Y_2} & \frac{\partial \varepsilon_3}{\partial Y_3} & \ldots & \frac{\partial \varepsilon_3}{\partial Y_n} \\ \vdots & \vdots & \ldots & \vdots \\ \frac{\partial \varepsilon_n}{\partial Y_2} & \frac{\partial \varepsilon_n}{\partial Y_3} & \ldots & \frac{\partial \varepsilon_n}{\partial Y_n} \end{vmatrix} \]
    \[ = \begin{vmatrix} 1 & 0 & \cdots & 0 \\ -\phi & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{vmatrix} = 1 \]
  - Then, the likelihood function can be written as
    \[ L(\phi, \sigma^2) = f(Y_1, \ldots, Y_n) = f(Y_1)f(Y_2, \ldots, Y_n|Y_1) = f(Y_1)f(\varepsilon_2, \ldots, \varepsilon_n) \]
    \[ = \left( \frac{1}{2\pi\gamma^0} \right)^{1/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^n \varepsilon_t^2 \right\} \]
    \[ \quad \times \frac{\gamma^0}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_1-\phi y_0)^2}{2\sigma^2}}, \text{ where } Y_1 \sim N\left(0, \frac{\sigma^2}{1-\phi^2} \right) \]
ARIMA Process – MLE: AR(1)

- Example:
  - Then,
  
  \[
  L(\phi, \sigma^2) = \frac{1 - \phi^2}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=2}^{n} (Y_t - \phi Y_{t-1})^2 + (1 - \phi^2) Y_1^2 \right\}
  \]

  - Then, the log likelihood function:

  \[
  \log L(\phi, \sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2} \ln \left(1 - \phi^2\right) - \frac{1}{2\sigma^2} \sum_{t=2}^{n} \left( Y_t - \phi Y_{t-1} \right)^2 + \left(1 - \phi^2\right) Y_1^2
  \]

  - \(S^*(\phi)\) is the conditional SS and \(S(\phi)\) is the unconditional SS.

ARIMA Process – MLE: AR(1)

- Example:
  - F.o.c.'s:

  \[
  \frac{\partial \ln L(\phi, \sigma^2)}{\partial \phi} = 0
  \]

  \[
  \frac{\partial \ln L(\phi, \sigma^2)}{\partial \sigma} = 0
  \]

  Note: If we neglect \(\ln(1-\phi^2)\), then MLE = Conditional LSE.

  \[
  \max_{\phi} L(\phi, \sigma^2) = \min S(\phi).
  \]

  If we neglect both \(\ln(1-\phi^2)\) and \((1 - \phi^2) Y_1^2\), then

  \[
  \max_{\phi} L(\phi, \sigma^2) = \min S(\phi^*).
  \]
ARIMA Process – Yule-Walker

2) Yule-Walker for AR(p): Regress $y_t$ against $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$.


Example: For an AR(p), we the Yule-Walker equations are

$$
\begin{bmatrix}
\gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\
\gamma(1) & \gamma(0) & \cdots & \gamma(p-2) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma(p-1) & \gamma(p-2) & \cdots & \gamma(0)
\end{bmatrix}
\begin{bmatrix}
\phi_1 \\
\phi_2 \\
\vdots \\
\phi_p
\end{bmatrix}
= \begin{bmatrix}
\gamma(1) \\
\gamma(2) \\
\vdots \\
\gamma(p)
\end{bmatrix}
$$

- MM Estimation: Equate sample moments to population moments, and solve the equation. In this case, we use:

$$
E(Y_t) = \frac{1}{T} \sum_{t=1}^{T} Y_t \Rightarrow \mu = \bar{Y}
$$

$$
E((Y_t - \mu)(Y_{t-k} - \mu)) = \frac{1}{T} \sum_{t=1}^{T} (Y_t - \mu)(Y_{t-k} - \mu) \Rightarrow \gamma_k = \hat{\gamma}_k \quad (\& \rho_k = \hat{\rho}_k)
$$

ARIMA Process – Yule-Walker

- Then, the Yule-Walker estimator for $\phi$ is given by solving:

$$
\begin{bmatrix}
\hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(p-1) \\
\hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(p-2) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{\gamma}(p-1) & \hat{\gamma}(p-2) & \cdots & \hat{\gamma}(0)
\end{bmatrix}
\begin{bmatrix}
\hat{\phi}_1 \\
\hat{\phi}_2 \\
\vdots \\
\hat{\phi}_p
\end{bmatrix}
= \begin{bmatrix}
\hat{\gamma}(1) \\
\hat{\gamma}(2) \\
\vdots \\
\hat{\gamma}(p)
\end{bmatrix}
= \begin{bmatrix}
\hat{\phi} = \hat{\Gamma}_p^{-1}\hat{\gamma}_p \\
\sigma^2 = \hat{\gamma}_0 - \hat{\phi}\hat{\gamma}_p
\end{bmatrix}
$$

- If $\hat{\gamma}_0 > 0$, then $\hat{\Gamma}_m$ is nonsingular.
- If $\{Y_t\}$ is an AR(p) process, $\hat{\phi}$ $\overset{d}{\longrightarrow}$ $N\left(\phi, \frac{\sigma^2}{T}\Gamma_p^{-1}\right)$

$$
\hat{\sigma}^2 \overset{d}{\longrightarrow} \sigma^2
$$

$$
\mathbf{\hat{\phi}_{kk}} \overset{d}{\longrightarrow} N\left(0, \frac{1}{T}\right) \text{ for } k > p.
$$

- Thus, we can use the sample PACF to test for AR order, and we can calculate approximated C.I. for $\phi$.  

ARIMA Process – Yule-Walker

• **Distribution:**
  If $y_t$ is an AR($p$) process, and $T$ is large,
  \[ \sqrt{T} (\hat{\phi} - \phi) \sim N(0, \hat{\sigma}^2 \hat{\Gamma}_p^{-1}) \]

• 100(1-$\alpha$)% approximate C.I. for $\phi_j$ is
  \[ \hat{\phi}_j \pm z_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{T}} (\hat{\Gamma}_p^{-1})_{jj}^{1/2} \]

• **Note:** The Yule-Walker algorithm requires $\Gamma^{-1}$.

• For AR($p$). The Levinson-Durbin (LD) algorithm avoids $\Gamma^{-1}$. It is a recursive linear algebra prediction algorithm. It takes advantage that $\Gamma$ is a symmetric matrix, with a constant diagonal (Toeplitz matrix). Use LD replacing $\gamma$ with $\gamma_0$.

• Side effect of LD: automatic calculation of PACF and MSPE.

ARIMA Process – Yule-Walker: AR(1)

**Example:** AR(1) (MM) estimation:

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

It is known that $\rho_1 = \phi$. Then, the MME of $\phi$ is

\[ \Rightarrow \rho_1 = \hat{\rho}_1 \]

\[ \hat{\phi} = \hat{\rho}_1 = \frac{\sum_{i=1}^{T} (y_i - \bar{y})(y_{i-1} - \bar{y})}{\sum_{i=1}^{T} (y_i - \bar{y})^2} \]

• Also, $\sigma^2$ is unknown:
  \[ \gamma_0 = \frac{\sigma^2}{1 - \phi^2} \Rightarrow \hat{\sigma}^2 = \hat{\gamma}_0 \left(1 - \hat{\phi}^2\right) \]
**ARIMA Process – Yule-Walker: MA(1)**

Example: MA(1) (MM) estimation:

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

Again using the autocorrelation of the series at lag 1,

\[ \rho_1 = \frac{0}{\left(1 + \theta^2\right)} = \hat{\rho}_1 \]
\[ 0^2 \hat{\rho}_1 + 0 + \hat{\rho}_1 = 0 \]
\[ \hat{\theta}_{1,2} = \frac{-1 \pm \sqrt{1 - 4\hat{\rho}_1^2}}{2\hat{\rho}_1} \]

- Choose the root satisfying the invertibility condition. For real roots:
  \[ 1 - 4\hat{\rho}_1^2 \geq 0 \Rightarrow 0.25 \geq \hat{\rho}_1^2 \Rightarrow -0.5 \leq \hat{\rho}_1 \leq 0.5 \]

If \( \hat{\rho}_1 = \pm 0.5 \), unique real roots but non-invertible.
If \( |\hat{\rho}_1| < 0.5 \), unique real roots and invertible.

**ARIMA Process – Yule-Walker**

- Remarks
  - The MMEs for MA and ARMA models are complicated.
  
  - In general, regardless of AR, MA or ARMA models, the MMEs are sensitive to rounding errors. They are usually used to provide initial estimates needed for a more efficient nonlinear estimation method.

  - The moment estimators are not recommended for final estimation results and should not be used if the process is close to being nonstationary or noninvertible.
**ARIMA Process – Estimation Hannan-Rissanen**

4) *Hannan-Rissanen algorithm for ARMA(p,q)*

Steps:
1. Estimate high-order AR.
2. Use Step (1) to estimate (unobserved) noise $\varepsilon_t$
3. Regress $y_t$ against $y_{t-1}, y_{t-2}, \ldots, y_{t-p}, \hat{\varepsilon}_{t-1}, \ldots, \hat{\varepsilon}_{t-q}$
4. Get new estimates of $\varepsilon_t$. Repeat Step (3).

---

**ARFIMA Process: Fractional Integration**

- Consider a simple ARIMA model: $(1-L)^d y_t = \varepsilon_t$

- We went over two cases for $d=0$ & 1. Granger and Joyeaux (1980) consider the model where $0 \leq d \leq 1$.

- Using the binomial series expansion of $(1-L)^d$:

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k = \sum_{k=0}^{\infty} \frac{d!}{(d-k)!k!} (-L)^k = \sum_{k=0}^{\infty} \prod_{a=0}^{k-1} \frac{(d-a)}{k!} (-L)^k$$

$$= 1 - dL + \frac{d(d-1)L^2}{2!} - ...$$

- We can invert $(1-L)^d$ to get $(1-L)^{-d} = 1 + dL + \frac{(d+1)dL^2}{2!} + ...$

- In the ARIMA(0,d,1): $y_t = (1-L)^{-d} \varepsilon_t = \varepsilon_t + d\varepsilon_{t-1} + \frac{(d+1)d\varepsilon_{t-2}}{2!} + ...$
ARFIMA Process: Fractional Integration

- In the ARIMA(0, d, 1): $y_t = (1 - L)^{-d} \varepsilon_t = \varepsilon_t + d \varepsilon_{t-1} + \frac{(d+1)d \varepsilon_{t-2}}{2!} + ...$

- The above MA can be approximated by:

$$y_t = (1 - L)^{-d} \varepsilon_t \approx \varepsilon_t + (1 + 1)^d \varepsilon_{t-1} + (1 + 2)^d \varepsilon_{t-2} + ... = \sum_{j=0}^{\infty} \beta_j \varepsilon_{t-j}$$

where $\beta_0 = 1$. Convergence depends on $d$.

- The series is covariance stationary if $d < 1/2$.

- ARFIMA models have slow (hyperbolic) decay patterns in the ACF. This type of slow decay patterns also show long memory for shocks.

ARFIMA Process: Estimation

- Estimation is complicated. Many methods have been proposed. The majority of them are two-steps procedures. First, we estimate $d$. Then, we fit a traditional ARMA process to the transformed.

Popular estimation methods:
- Based on the log periodogram regressions, due to Geweke and Porter-Hudak (1983), GPH.
- Approximated ML (AML), due to Beran (1995). In this case, all parameters are estimated simultaneously.
ARFIMA Process: Remarks

- In a general review paper, Granger (1999) concludes that ARFIMA processes may fall into the empty box category – i.e., models with stochastic properties that do not mimic the properties of the data.


ARFIMA Process: Example

- From Bhardwaj and Swanson (2004)
SARIMA Process: Seasonal Time Series

- A time series repeats itself after a regular period of time.

- “Business cycle effects” in macroeconomics “time of the day” in trading patterns, “Monday effect” for stock returns, “9 to 5 effect” for electricity demand, etc.

- The smallest time period for this repetitive phenomenon is called a seasonal period, $s$.

Seasonal Time Series

- Two types of seasonal behavior:
  - Deterministic – Usual treatment: Build a deterministic function,
    \[ f(t) = f(t + k \times s), \quad k = 0, \pm 1, \pm 2, \ldots \]
    We can include seasonal (means) dummies. Instead of dummies, trigonometric functions (sum of cosine curves) can be used. A linear time trend is often included in both cases.

  - Stochastic – Usual treatment: SARIMA model. For example:
    \[ (1 - L')(1 - L)^d y_t = (1 - \theta L)(1 - \Theta L') \varepsilon_t \]
    where $s$ the seasonal periodicity or frequency of $y_t$. 

Seasonal Time Series - SARIMA

- For stochastic seasonality, we use the Seasonal ARIMA model. In general, we have the SARIMA\((P, D, Q)_s\)
  \[ \Phi_p (B^s) (1 - B^s)^D y_t = \theta_0 + \Theta_q (B^s) \varepsilon_t \]
  where \(\theta_0\) is constant and
  \[ \Phi_p (B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_p B^{ps} \]
  \[ \Theta_q (B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_q B^{qs} \]

Example 1: SARIMA\((0,0,1)_12\) = SMA\((1)_12\)
  \[ y_t = \theta_0 + \varepsilon_t - \Theta \varepsilon_{t-12} \]
  - Invertibility Condition: \(|\Theta| < 1\).
  - E\((y_t) = \theta_0\)
  - \(ACF : \rho_k = \begin{cases} \frac{-\Theta}{1 + \Theta^2}, & |k| = 12 \\ 0, & \text{otherwise} \end{cases} \)
  - Var\((y_t) = (1 + \Theta^2)\sigma^2\)

Seasonal Time Series - SARIMA

Example 1: SARIMA\((1,0,0)_12\) = AR\((1)_12\)
  \[ (1 - \Phi B^{12}) y_t = \theta_0 + \varepsilon_t \]
  - This is a simple seasonal AR model.
  - Stationarity Condition: \(|\Phi| < 1\).
  - E\((Y_t) = \frac{\theta_0}{1 - \Phi} \)
  - Var\((Y_t) = \frac{\sigma^2}{1 - \Phi^2} \)
  - ACF : \(\rho_{12k} = \Phi^k, k = 0, \pm 1, \pm 2, \cdots \)

- When \(\Phi = 1\), the series is non-stationary. To test for a unit root, consider seasonal unit root tests.
Seasonal Time Series – Multiplicative SARIMA

• A special, parsimonious class of seasonal time series models that is commonly used in practice is the multiplicative seasonal model $ARIMA(p, d, q)(P, D, Q)_s$.

\[ \phi_p(L) \Phi_p(L^s)(1 - L)^d(1 - L^s)^D y_t = \theta_0 + \theta_q(L) \Theta_q(L^s) \epsilon_t \]

where all zeros of $\phi(L); \Phi(L^s); \theta(L) \text{ and } \Theta(L^s)$ lie outside the unit circle. Of course, there are no common factors between $\phi(L) \Phi(L^s)$ and $\theta(L) \Theta(L^s)$.

• When $\Phi = 1$, the series is non-stationary. To test for a unit root, consider seasonal unit root tests.

Seasonal Time Series – Multiplicative SARIMA

• The ACF and PACF can be used to discover seasonal patterns.
Seasonal Time Series – Multiplicative SARIMA

• The ACF and PACF can be used to discover seasonal patterns.
Seasonal Time Series – Multiplicative SARIMA

• We usually work with the RHS variable, \( W_t = (1 - B)(1 - B^{12})y_t \)

\[
W_t = (1 - \theta L)(1 - \Theta L^{12})a_t, \\
W_t = a_t - \theta a_{t-1} - \Theta a_{t-12} + \theta \Theta a_{t-13}, \\
W_t \sim I(0)
\]

\[
\gamma_k = \begin{cases} 
(1 + \theta^2)(1 + \Theta^2)\sigma^2, & k = 0 \\
-\theta(1 + \Theta^2)\sigma^2, & |k| = 1 \\
-\Theta(1 + \theta^2)\sigma^2, & |k| = 12 \\
\theta \Theta \sigma^2, & |k| = 11, 13 \\
0, & \text{otherwise}
\end{cases}
\]

\[
\rho_k = \begin{cases} 
-\theta \\
-\Theta \\
\frac{\theta \Theta}{(1 + \theta^2)(1 + \Theta^2)}, & |k| = 11, 13 \\
0, & \text{otherwise}
\end{cases}
\]

Seasonal Time Series – Multiplicative SARIMA

• Most used seasonal model in practice: SARIMA\((0,1,1)(0,1,1)_{12}\)

\[
(1 - L)(1 - L^{12})y_t = (1 - \theta L)(1 - \Theta L^{12})\epsilon_t
\]

where \(|\theta| < 1\) and \(|\Theta| < 1\).

• This model is the most used seasonal model in practice. It was used by Box and Jenkins (1976) for modeling the well-known monthly series of airline passengers. It is called the airline model.

• We usually work with the RHS variable, \( W_t = (1 - L)(1 - L^{12})y_t \).

\( (1 - L) \): “regular” difference

\( (1 - L^{12}) \): “seasonal” difference.
Seasonal Time Series – Seasonal Unit Roots

• If a series has seasonal unit roots, then standard ADF test statistic do not have the same distribution as for non-seasonal series.

• Furthermore, seasonally adjusting series which contain seasonal unit roots can alias the seasonal roots to the zero frequency, so there is a number of reasons why economists are interested in seasonal unit roots.


Non-Stationarity in Variance

• Stationarity in mean does not imply stationarity in variance

• Non-stationarity in mean implies non-stationarity in variance.

• If the mean function is time dependent:
  1. The variance, Var(y_t) is time dependent.
  2. Var[y_t] is unbounded as $t \rightarrow \infty$.
  3. Autocovariance functions and ACFs are also time dependent.
  4. If $t$ is large wrt $y_0$, then $p_k \approx 1$.

• It is common to use variance stabilizing transformations: Find a function $G(.)$ so that the transformed series $G(y_t)$ has a constant variance. For example, the Box-Cox transformation:

$$T (Y_t) = \frac{Y_t^\lambda - 1}{\lambda}$$
Variance Stabilizing Transformation - Remarks

- Variance stabilizing transformation is only for positive series. If a series has negative values, then we need to add each value with a positive number so that all the values in the series are positive.

- Then, we can search for any need for transformation.

- It should be performed before any other analysis, such as differencing.

- Not only stabilize the variance, but we tend to find that it also improves the approximation of the distribution by Normal distribution.