# Lecture 4

## Univariate time series modelling and forecasting
## (Box-Jenkins Method)

Dr. Junsoo Lee

EC 413

Read

Handbook Ch 2; Enders Ch 2

## Univariate Time Series Models

Where we attempt to predict returns using ONE variable based on only information contained in their past values.

## Overview

### Stationary ARMA (p, q) Model

What is ARMA model?

$$y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + .. + a_p y_{t-p} + q_1 u_{t-1} + q_2 u_{t-2} + ... + q_q u_{t-q} + u_t$$

.. ARMA($p$, $q$) model

What is AR model?

- An autoregressive model of order $p$, an AR($p$) can be expressed as

$$y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + .. + a_p y_{t-p} + u_t$$

.. AR($p$) model

What is MA model?

- Let $u_t$ ($t$=1,2,3,...) be a white noise process, a sequence of independently and identically distributed (iid) random variables with E($u_t$)=0 and Var($u_t$) = $\sigma^2$. The $q$th order MA model is given as:

$$y_t = m + u_t + q_1 u_{t-1} + q_2 u_{t-2} + ... + q_q u_{t-q}$$

.. MA($q$) model

What is a White Noise Process, $u_t$?

- A white noise process is one with (virtually) no discernible structure. It is an independently and identically distributed (iid) random variables.

### Key Point of ARMA Models:

Allow for enough AR & MA terms so that *the error term looks like a white noise process*.

## ARIMA (p, d, q) Model

If $y_t$ is non-stationary, we take a first-difference of $y_t$ so that $\Delta y_t$ becomes stationary.

$$\Delta y_t = y_t - y_{t-1}$$

(d = 1 implies one time differencing. d = 1 is enough in most cases.)

$$\Delta y_t = c + a_1 \Delta y_{t-1} + a_2 \Delta y_{t-2} + .. + a_p \Delta y_{t-p} + q_1 u_{t-1} + q_2 u_{t-2} + ... + q_q u_{t-q} + u_t$$

.. ARIMA($p$, $1$, $q$) model

In unusual cases, we use a second difference, impying a firs difference of the differenced series

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$

**Questions:**

How many AR and MA terms are necessary?

How can we tell that the error term is a white noise process?

What do you mean by “stationary”?

→ Box-Jenkins provided procedures to estimate an optimal ARIMA model.

## Some Notation and Concepts

### A Weakly “Stationary” Process

If a series satisfies the next three equations, it is said to be weakly or covariance stationary: they do not depend on t.

1. $E(y_t) = m$ , $t = 1,2,...,\infty$ … Unconditional mean.
2. $Var(y_t) = s^2 <$ ¥ … Unconditional variance
3. $Cov(y_t, y_{t-s}) = g_s$ … Auto-covariance

- So if the process is covariance stationary, all the variances are the same and all the covariances depend on the difference between $t$ and $t$-$s$.

Ex) $Cov(y_t, y_{t-5}) = Cov(y_{t-1}, y_{t-6}) = Cov(y_{t-4}, y_{t-9}) = g_5$

*(All these are the same!)*

We say that $g_5$ is the autocovariance function with lag 5.

## Auto-Correlation Function (ACF)

- The covariances, $\gamma_s$, are known as autocovariances.

  One can find $\gamma_1$, $\gamma_2$, $\gamma_3$, $\gamma_4$, $\gamma_s$, ….. and so on.

  Here, $\gamma_0 = \text{Cov}(y_t,\ y_{t-0}) = \text{Var}(y_t) = \sigma^2$

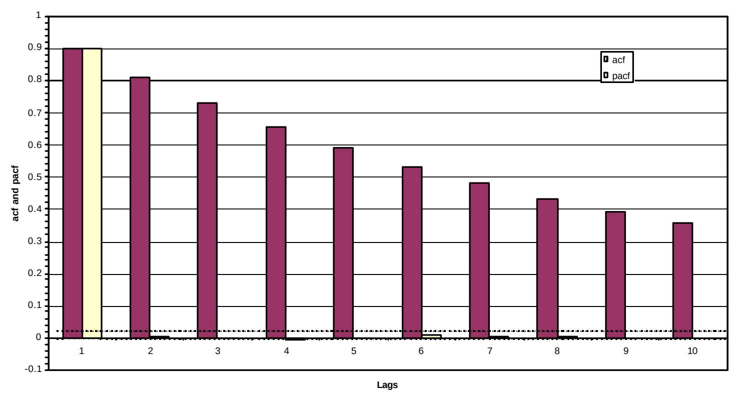
- However, the value of the autocovariances depend on the units of measurement of $y_t$. That is, if $y_t$ is multiplied by 100, $\gamma_s$ is also computed as 100 * $\gamma_s$.

- It is thus more convenient to use the autocorrelations which are the autocovariances normalised by dividing by the variance:

$$\boxed{\tau_s = \frac{\gamma_s}{\gamma_0}} \qquad -1 \le \tau_s \le 1, \qquad s = 0,1,2,\ldots$$

  $t_s$ is the autocorrelation between $y_t$ and $y_{t-s}$.

- If we plot $t_s$ against $s$=0,1,2,... then we obtain the **AutoCorrelation Function** (**ACF**) or called, **correlogram**.




## White Noise Process ($u_t$)

- A white noise process is one with (virtually) no discernible structure. It is an independently and identically distributed (iid) random variables.

  (i) $E(u_t)=0$
  (ii) $\text{Var}(u_t) = \sigma^2$
  (iii) $\gamma_s = 0$ if $s \neq 0$

  That is, $\gamma_0 = \sigma^2$, $\gamma_1 = \gamma_2 = \ldots = 0$ ($s > 0$)

Thus the autocorrelation function ($t_s$) will be zero apart from a single peak of 1 at $s = 0$.

- We have a distribution of $t_s$ under the null that a time series is a white noice.

  $t_s \sim$ approximately N(0,1/$T$) where $T$ = sample size

  *This implies:*

  $Var(t_s) = 1/T$

  $SE(t_s) = \sqrt{1/T}$

  If so, the 95% confidence interval of $t_s$ of a white noice process is

  $t_s = 0 \pm 1.96 * SE(t_s) = 0 \pm 1.96 * \sqrt{1/T}$

- **Point: If a time series is a white noise, $|\hat{t}_s|$ should not be bigger than the confidence interval, 1.96 * $\sqrt{1/T}$ for all values of s.**

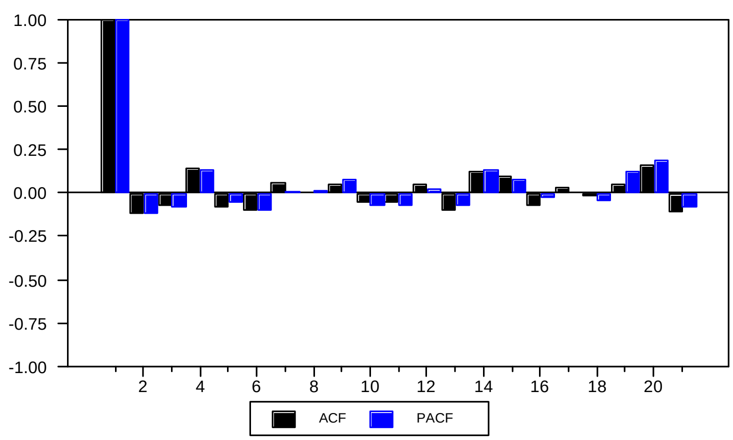
## Testing for a White Noise Process

*Does your data look like a white noise process?*

- We can use this to do significance tests for the autocorrelation coefficients by constructing a confidence interval.

  If $\hat{t}_s$ falls outside this region ($\pm$ **1.96 *** $\sqrt{1/T}$) for any value of $s$, then we reject the null hypothesis that the true value of the coefficient at lag $s$ is zero, which implies a white noise process.

  Ex) RATS example



Correlations of Series U
Autocorrelations

```
  1: -0.1154773 -0.0722445  0.1479023 -0.0814163 -0.0977566  0.0595666
  7:  0.0028219  0.0512692 -0.0584719 -0.0513372  0.0491615 -0.1019420
 13:  0.1220512  0.0973237 -0.0764919  0.0347748 -0.0208870  0.0551283
 19:  0.1662776 -0.1114049
```

Partial Autocorrelations (later on this)

```
  1: -0.1154773 -0.0867362  0.1315508 -0.0565580 -0.0969105  0.0106180
  7:  0.0179404  0.0821972 -0.0697137 -0.0695742  0.0229997 -0.0766431
 13:  0.1335587  0.0833780 -0.0288299 -0.0014239 -0.0451313  0.1250012
 19:  0.1921916 -0.0795414
```

- What if only one or a few $\hat{t}_s$ lie outside the confidence interval? Is there any test for joint insignificance up to lag s?

  -> *Joint restriction Test*

- We can also test the joint hypothesis that all *m* of the $t_k$ correlation coefficients are jointly equal to zero using the *Q*-statistic developed by Box and Pierce:

$$Q(m) = T \sum_{k=1}^{m} t_k^2$$

  where *T* = sample size, *m* = maximum lag length

  $H_o$: $t_1 = t_2 = t_3 = \ldots = t_m = 0$ (implying a white noise process)

  The *Q*-statistic is asymptotically distributed as a chi-square distribution ($\chi^2$)

  If $Q(m) > \chi^2_m$ (m = degree of freedom), the reject the null hypothesis and we say that the time series is NOT a white noice process

  If $Q(m) < \chi^2_m$, then we cannot reject the null of a white noise process using lags up to m. We often choose m = 10, 20, or 30.

- However, the Box Pierce test has poor small sample properties, so a variant has been developed, called the Ljung-Box statistic:

$$Q^*(m) = T(T+2) \sum_{k=1}^{m} t_k^2/(T-k)$$

  The *Q*-statistic is aso asymptotically distributed as a chi-square distribution ($\chi^2$) and $Q^*$ can be used instead of Q.

  This statistic is commonly used as a portmanteau (general) test of linear dependence (autocorrelation) in time series.

- <u>Question:</u>

Suppose that a researcher had estimated the first 5 autocorrelation coefficients using a series of length 100 observations, and found them to be (from 1 to 5): 0.207, -0.013, 0.086, 0.005, -0.022. Test each of the individual coefficient for significance, and use both the Box-Pierce and Ljung-Box tests to establish whether they are jointly significant.

- Solution:

A coefficient would be significant if it lies outside (-0.196, +0.196) at the 5% level, so only the first autocorrelation coefficient is significant. **(**Here, ±1.96 * $\sqrt{1/T}$ = ±0.196)

Next, the joint test is on:

$H_o$: $t_1 = t_2 = \ldots = t_5 = 0$ (implying a white noise process)

$$Q(5) = T \sum_{k=1}^{m} t_k^2 = 5.09 \text{ and}$$

$$Q^*(5) = T(T+2) \sum_{k=1}^{m} t_k^2/(T-k) = 5.26$$

Compared with a tabulated $c^2{}_c$ with df 5 is 11.1 at the 5% level (see $c^2{}_c$ chart), these are smaller than 11.1. So, we can say that the 5 coefficients are jointly insignificant. Then, the time series appears a white noise process.

Note: In RATS, we can also examine p-values. (if p-value > 5%, the null is not rejected. Thus, it is a white noise process. Otherwise, it is not a white oise process).

*Example) RATS reports:*

*Ljung-Box Q-Statistics*
*Q(20-0)= 18.0031. Significance Level 0.58720637*

® *Since the p-value> 0.05, we cannot reject the null that the series is a white noise process using 20 lags.*

- Most time series data are NOT a white noise process. The Box-Jenkins method is to find a model in such as way that by adding AR and MA terms the resulting residuals will look like a white noise process.

If $y_t$ is not a white noise process, add AR and MA terms until $u_t$ looks like a white noice process.

$$Y_t = (p \text{ \# of AR terms}) + (q \text{ \# of MA terms}) + u_t$$

where AR terms are $y_{t-j}$ (past values) and MA terms are $u_{t-j}$ (past errors).

If p and q are properly selected, $u_t$ will be a white noise process.

**Point**: After adding AR-MA terms, check if the residual ($\hat{u}_t$) is a white noise process. (more on this, later)

(i) Use ACF plots of $\hat{u}_t$ and check if any $\tau_k$ lies outside the CI.
(ii) Use Q* statistic on $\hat{u}_t$.

<u>RATS Example</u> (ARMA1.prg)

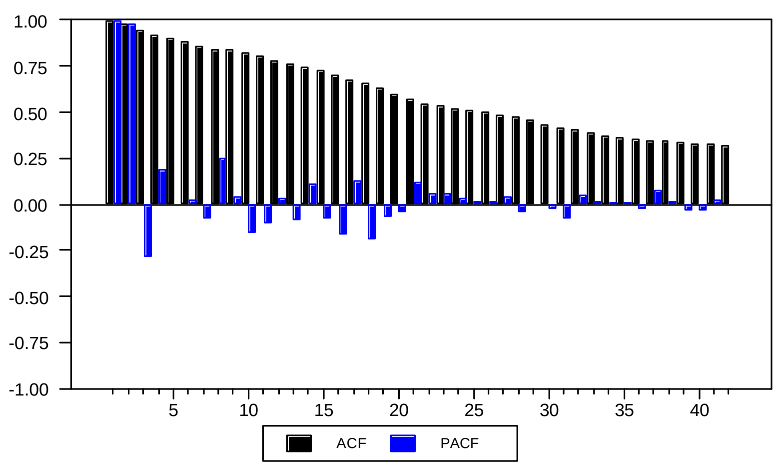
```
calendar 1960 1 12
allocate 1999:9

open data basics.wks
data(format=wks,org=cols) / rate

graph(key=upleft)
# rate

* Compute and graph autocorrelations:
correlate(partial=PACF, QSTATS) rate / ACF
graph(key=below,style=bar,nodates,min=-1.0,max=1.0,number=1) 2
# ACF
# PACF
```

*<u>Results</u>:*




*Ljung-Box Q-Statistics*
*Q(41-0)= 7378.5980. Significance Level 0.00000000*

→ Clearly, RATE is not a white noise process.

**MA (Moving Average) Models**

What is MA model?

- Let $u_t$ ($t$=1,2,3,...) be a white noise process, a sequence of independently and identically distributed (iid) random variables with $E(u_t)=0$ and $Var(u_t) = \sigma^2$. The $q$th order MA model is given as:

  $$y_t = m + u_t + q_1 u_{t-1} + q_2 u_{t-2} + ... + q_q u_{t-q}$$

  .. MA($q$) model

- The model is expressed in terms of past errors. We wish to estimate the coefficients $q_j$, $j=1,..,q$, and use the model for forecasting.

- Its properties are

  (i) Only q errors affect the current leve $y_t$ but higher order errors do not affect $y_t$.

  (ii) It is a short memory model.

  - For an MA(q) model, $\gamma_s = 0$ for s > q.

  Ex) MA(1) model

  - $y_t = m + u_t + q_1 u_{t-1}$
  - Of course, $\tau_1$ = ACF(1) is not 0 (it is not a white noise process). But, $\tau_2$ = ACF(2) = 0. Also, $\tau_3 = \tau_4 = .. = 0$.

    Ex) RATS example

  **Point**: If $\tau_k = \tau_{k+1} = \ldots = 0$, then the process may follow an MA(k-1) model.

Finding theoretical Variance and Covariances of MA models

Example 1)

Consider an MA(1) model, $y_t = m + u_t + q_1 u_{t-1}$

(i) Mean: $E(y_t)=m + 0 + 0 = \mu$.

(ii) Variance: $\gamma_0 = Var(y_t) = Var(m + u_t + q_1 u_{t-1}) = 0 + \sigma^2 + q_1^2 \sigma^2 = (1 + q_1^2)\sigma^2$

(iii) Covariances: $\gamma_s = Cov(y_t, y_{t-s})$

$\gamma_1 = \text{Cov}(y_t, y_{t-1}) = \text{Cov}(m + u_t + q_1 u_{t-1}, m + u_{t-1} + q_1 u_{t-2})$
$= E[y_t - E(y_t)][y_{t-1} - E(y_{t-1})]$
$= E[(u_t + q_1 u_{t-1})(u_{t-1} + q_1 u_{t-2})]$
since $y_t - E(y_t) = (m + u_t + q_1 u_{t-1}) - \mu = u_t + q_1 u_{t-1}$
$= q_1 \sigma^2$

Thus, $\tau_1 = \gamma_1 / \gamma_0 = q_1 \sigma^2 / [(1 + q_1^2)\sigma^2] = q_1/(1 + q_1^2)$

$\gamma_2 = \text{Cov}(y_t, y_{t-2}) = \text{Cov}(m + u_t + q_1 u_{t-1}, m + u_{t-2} + q_1 u_{t-3})$
$= E[(u_t + q_1 u_{t-1})(u_{t-2} + q_1 u_{t-3})]$
$= 0$
$\gamma_3 = E[(u_t + q_1 u_{t-1})(u_{t-3} + q_1 u_{t-4})] = 0$
$\gamma_4 = 0\ldots$

Thus, $\tau_s = 0$, s = 2, 3, 4..

Point: If $\tau_s = 0$, s = 2, 3, 4.., then it may be an MA(1) model.

Exercise) homework

For an MA(2) model, find $E(y_t)$, $\gamma_0$, $\gamma_1$, $\gamma_2$, $\gamma_3$, ..

## AR (Autoregressive) Models

What is AR model?

- An autoregressive model of order *p*, an AR(*p*) can be expressed as

$$y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + .. + a_p y_{t-p} + u_t$$

where $u_t$ is a white noise process

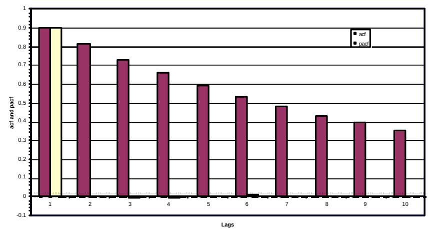
- The model is expressed in terms of past values. We wish to estimate the coefficients $a_j$, $j=1,..,p$, and use the model for forecasting.

- Its properties are

(i) *All previous values* will have cumulative effects on the current level $y_t$.

(ii) Thus, it is a long-run memory model.

  - $\tau_s$ = ACF(s) does not die out easily. It takes a longer time to have ACF close to zero.




- In theory, we say that there is a <u>persistent effect</u> over time.
- Many time series follow AR(p) models.

Ex) AR(1) model

- $y_t = c + \alpha\, y_{t-1} + u_t$
  - (Unconditional) Mean:

    $E(y_t) = E(c + \alpha\, y_{t-1} + u_t) = c + \alpha E(y_{t-1}) + 0$

    Letting $E(y_t) = E(y_{t-1}) = \mu$, (assuming weakly stationarity)

    $$\mu = c + \alpha\mu \rightarrow \mu = c / (1-\alpha).$$

  - Variance

    $Var(y_t) = Var(c + \alpha\, y_{t-1} + u_t) = 0 + \alpha^2 Var(y_{t-1}) + \sigma^2$

    Letting $Var(y_t) = Var(y_{t-1}) = \gamma_0$, (assuming weakly stationarity)

    $$\gamma_0 = \alpha^2\gamma_0 + \sigma^2 \rightarrow \gamma_0 = \sigma^2 / (1 - \alpha^2)$$

  - Covariance ($\gamma_s$)

    From $y_t = c + \alpha\, y_{t-1} + u_t$, let c = 0 for simplicity.

    We can have: $\boldsymbol{y_{t-1}} = \alpha\, y_{t-2} + u_{t-1}$

    $y_t = \alpha\, \boldsymbol{y_{t-1}} + u_t = \alpha\, (\alpha\, \boldsymbol{y_{t-2}} + u_{t-1}) + u_t$

    $= \alpha\, (\alpha(\alpha\, \boldsymbol{y_{t-3}} + u_{t-2}) + u_{t-1}) + u_t$

    = (keep substituting..)

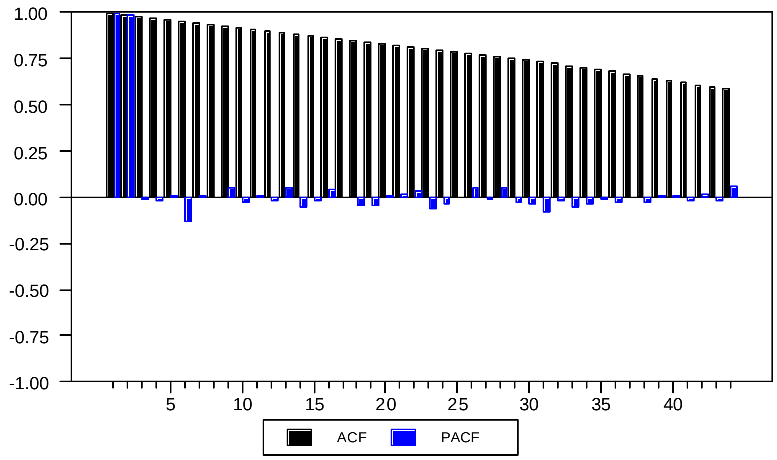
    $E(y_t y_{t-1}) = \alpha$

    $E(y_t y_{t-2}) = \alpha^2$

    $E(y_t y_{t-3}) = \alpha^3$

    .. so on

    Example) $\alpha = 0.99$




Since $|\alpha| < 1$ for stationary series, $\alpha^k \rightarrow 0$ as k increases, but it takes a while (k not small) to have $\alpha^k$ close to 0.

- Note: If $\alpha = 1$ (as in the random walk model, $y_t = y_{t-1} + u_t$), the unconditional mean does not exist, and $\alpha^k = 1$ for any k. This is the case of non-stationary data. If so, we need to do the first-difference. $\Delta y_t = y_t - y_{t-1}$. (more on this later.)

<u>Partial Autocorrelation Functions (PACF)</u>

- Measures the correlation between an observation *k* periods ago and the current observation, after controlling for observations at intermediate lags (i.e. all lags < *k*).

  PACF(k) = ACF(k) after controlling the effects of $(y_{t-1}, .., y_{t-k+1})$

  $\mathbf{y_t}$, $(y_{t-1}, .. , y_{t-k+1})$, $\mathbf{y_{t-k}}$

  At lag 1, the ACF(1) = PACF(1) always.

  At lag 2, $PACF(2) = (t_2 - t_{12}) / (1 - t_{12})$.

  For lags 3+, the formulae are more complex.

- PACF(k) can be found as the coefficient of $y_{t-k}$ in the regression:

  $Y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots.. + \alpha_{k-1} y_{t-k+1} + a_{\mathbf{k}} y_{t-k} + u_t$

  $\rightarrow \quad \alpha_k = \text{PACF(k)}$

- The PACF is useful for telling the maximum order of an AR process.

- For AR(q) models, PACF(q+s) = 0, s ≥ 1.

  Ex) If PACF(2) = PACF(3) = .. 0, then it may be an AR(1) model.

  Ex) If PACF(3) = PACF(4) = .. 0, then it may be an AR(2) model.

→ Point: For an AR(*p*), the theoretical pacf will be zero after lag *p*.

- For an MA(*q*), the theoretical PACF will be geometrically declining.

| Point: |
|---|
| 1. For an AR(*p*), the theoretical PACF will be zero after lag *p*. |
| 2. For an MA(*q*), the theoretical ACF will be zero after lag *q*. |

Summary

- An AR process has
  - a geometrically decaying ACF
  - number of spikes of PACF = AR order
- An MA process has
  - Number of spikes of ACF = MA order
  - a geometrically decaying PACF

## Stationarity Conditions for AR models

- The condition for stationarity of a general AR(*p*) model is that the roots of $1 - \alpha_1 z - \alpha_2 z^2 - \ldots - \alpha_p z^p = 0$ all lie outside the unit circle.

  Examples) Is $y_t$ stationary?

  (a) $y_t = y_{t-1} + u_t$

  $1 - z = 0. \quad \rightarrow \quad z = 1$

  The characteristic root is 1, so it is a unit root process (so non-stationary)

  (b) $y_t = 0.2y_{t-1} + 0.35y_{t-2} + u_t$

  $1 – 0.3\, z – 0.35\, z^2 = 0$ or $(1\text{-}0.7z)(1\text{+}0.5z) = 0$. Thus,

  $z = 1/0.7 = 1.43$ and $z = 1/0.5= 2$

The characteristic roots are 1.43 and 2. When any one of these lies outside the unit circle (bigger than 1), the process is stationary

- MA models are already stationary. Why?
- A stationary AR($p$) model can have an MA(∞) representation.

Ex) $y_t = 0.5\, y_{t-1} + u_t$

$y_t - 0.5y_{t-1} = u_t$ or $(1 - 0.5L)y_t = u_t$ using an lag operator, $Ly_t = y_{t-1}$

(in genelra, $L^k y_t = y_{t-k}$)

$$y_t = \frac{1}{1 - 0.5L} u_t = (1 + 0.5L + 0.5^2 L^2 + 0.5^3 L^3 + \ldots)u_t$$

$$= u_t + 0.5u_{t-1} + 0.5^2 u_{t-2} + 0.5^3 u_{t-3} + \quad .. \quad \text{MA}(\infty) \text{ form}$$

*Point: Any stationalry AR model can be seen as an MA model. Any (invertible) MA model can be seen as AR model. Thus, AR or MA models are interchangeable.*

## ARMA (Autoregressive MA) Models

An ARMA(p, q) is expressed as:

$$y_t = c + a_1 y_{t-1} + a_2 y_{t-2} + .. + a_p y_{t-p} + u_t + q_1 u_{t-1} + q_2 u_{t-2} + \ldots + q_q u_{t-q}$$

….. A Combination f AR and MA models.

- Either ACF or PACF cannot solely provide the information on the maximum orders of p or q.

## Building ARMA Models - The Box Jenkins Approach

- Box and Jenkins (1970) were the first to approach the task of estimating an ARMA model in a systematic manner. There are 3 steps to their approach:

  1. Identification

  2. Estimation

  3. Model diagnostic checking

  Step 1: (COR on the data and AIC/BIC)

- Involves determining the order of the model.
- Use of graphical procedures
- A better procedure is now available

Step 2: (BOXJENK)

- Estimation of the parameters
- Can be done using least squares or maximum likelihood depending on the model.

Step 3: (COR on residuals)

- Model checking on residuals

- Box and Jenkins suggest 2 methods:

  - Deliberate overfitting
  - Residual diagnostics

- Identification would typically not be done using ACF and PACF's.

- We want to form a parsimonious model.
  - variance of estimators is inversely proportional to the number of degrees of freedom.
  - models which are profligate might be inclined to fit to data specific features

- This gives motivation for using information criteria, which embody 2 factors

  - a term which is a function of the RSS
  - some penalty for adding extra parameters

- The object is to choose the number of parameters which minimises the information criterion. And the information criteria vary according to how stiff the penalty term is.

- The two most popular criteria are Akaike's (1974) information criterion (AIC), Schwarz's (1978) and Bayesian information criterion (SBIC).

$$AIC = \ln(\hat{\sigma}^2) + 2\,k/T$$

$$SBIC = \ln(\hat{\sigma}^2) + (k/T)\ln T$$

where $k = p + q + 1$, $T$ = sample size.

**Point: Choose a model with a LOWER values.**

- Which IC should be preferred if they suggest different model orders?

  - *SBIC* is strongly consistent but (inefficient).
  - *AIC* is not consistent, and will typically pick “bigger” models.

- Forecasting: based on the estimated ARMA models.

## Practical Applications

Handouts

Example 1) ARMA2.PRG
Practical Application

Example 2) ARMA3.PRG
Using bjident.src and bjfore.src in RATS

Homeworks