DATA ANALYTICS

Unit 3: Concept of ACF and PACF and Correlogram

Jyothi R.
Department of Computer Science and Engineering

Introduction

- Autocorrelation Functions and ARIMA Modelling
- Define what stationarity is and why it is so important to Econometrics
- Describe the Autocorrelation coefficient and its relationship to stationarity
- Evaluate the Q-statistic
- Describe the components of an Autoregressive Integrated Moving Average Model (ARIMA model)

Stationarity

- A strictly stationary process is one where the distribution of its values remains the same as time proceeds, implying that the probability lies in a particular interval is the same now as at any point in the past or the future.
- However we tend to use the criteria relating to a ‘weakly stationary process’ to determine if a series is stationary or not.

Weakly Stationary Series

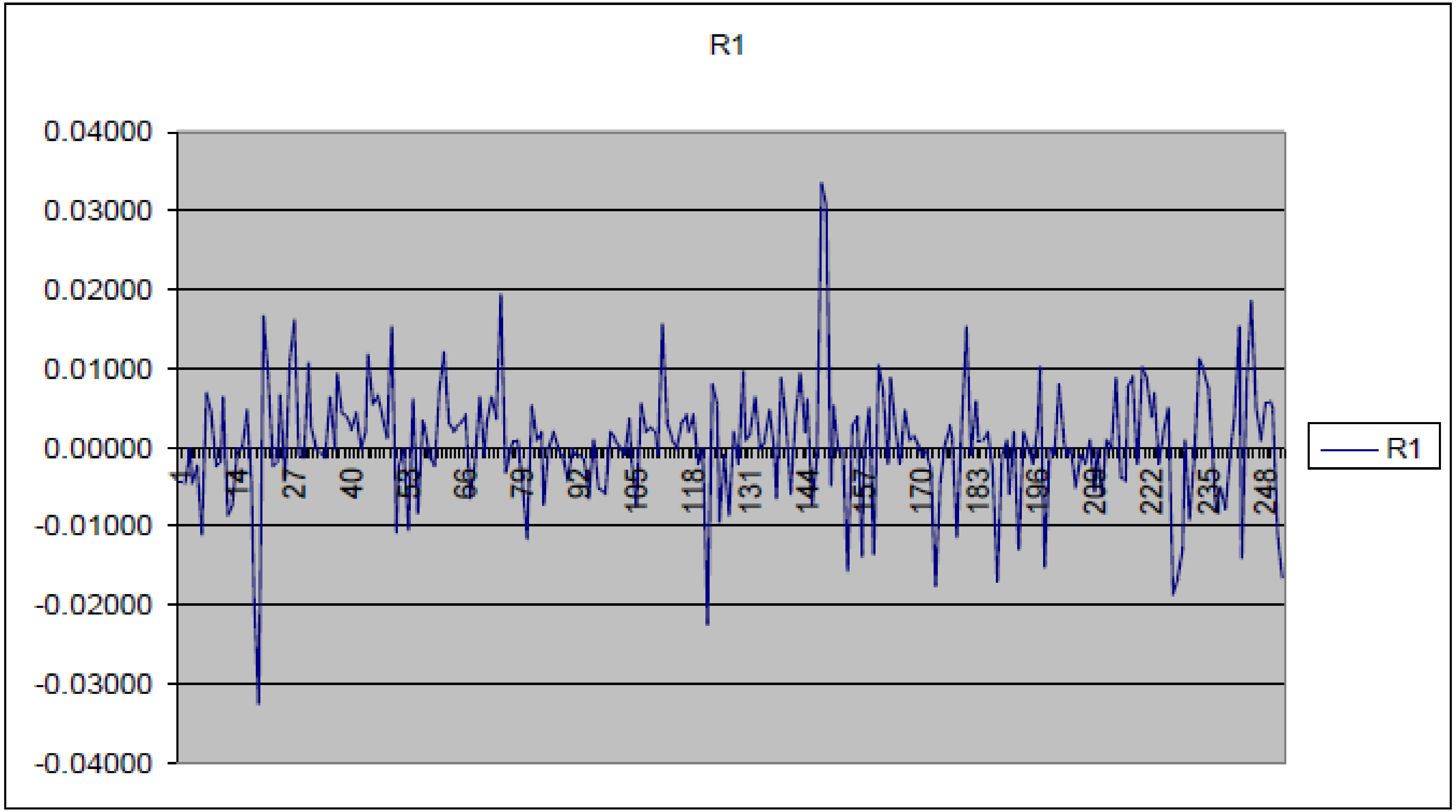
- A stationary process or series has the following properties:
  - constant mean
  - constant variance
  - constant auto covariance structure
- The latter refers to the covariance between y(t-1) and y(t-2) being the same as y(t-5) and y(t-6).

Stationary Series

- $E(y_t) = \mu$
- $E(y_t - \mu)^2 = \sigma^2$
- $E(y_{t1} - \mu)(y_{t2} - \mu) = \gamma_{t2-t1}, \forall t_1, t_2$

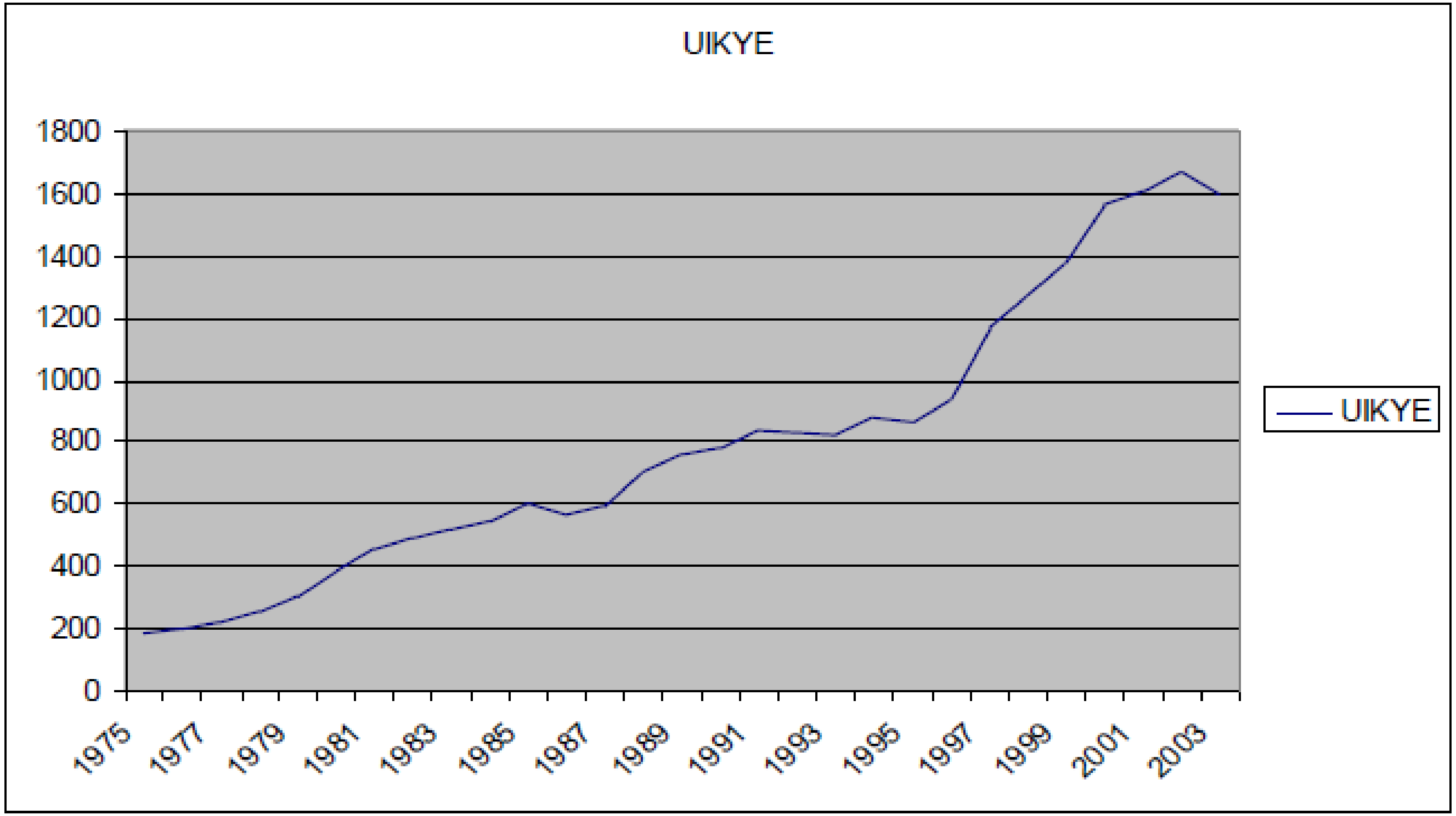
DATA ANALYTICS

Stationary Series



DATA ANALYTICS

Non-stationary Series



Implications of Non-stationary data

- If the variables in an OLS regression are not stationary, they tend to produce regressions with high R-squared statistics and low DW statistics, indicating high levels of autocorrelation.
- This is caused by the drift in the variables often being related, but not directly accounted for in the regression, hence the omitted variable effect.

Stationary Data

- It is important to determine if our data is stationary before the regression.
- This can be done in a number of ways:

  - plotting the data
  - assessing the autocorrelation function
  - Using a specific test on the significance of the autocorrelation coefficients.
  - Specific tests to be covered later.

Autocorrelation Function (ACF)

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{covariance at lag k}}{\text{variance}}$$

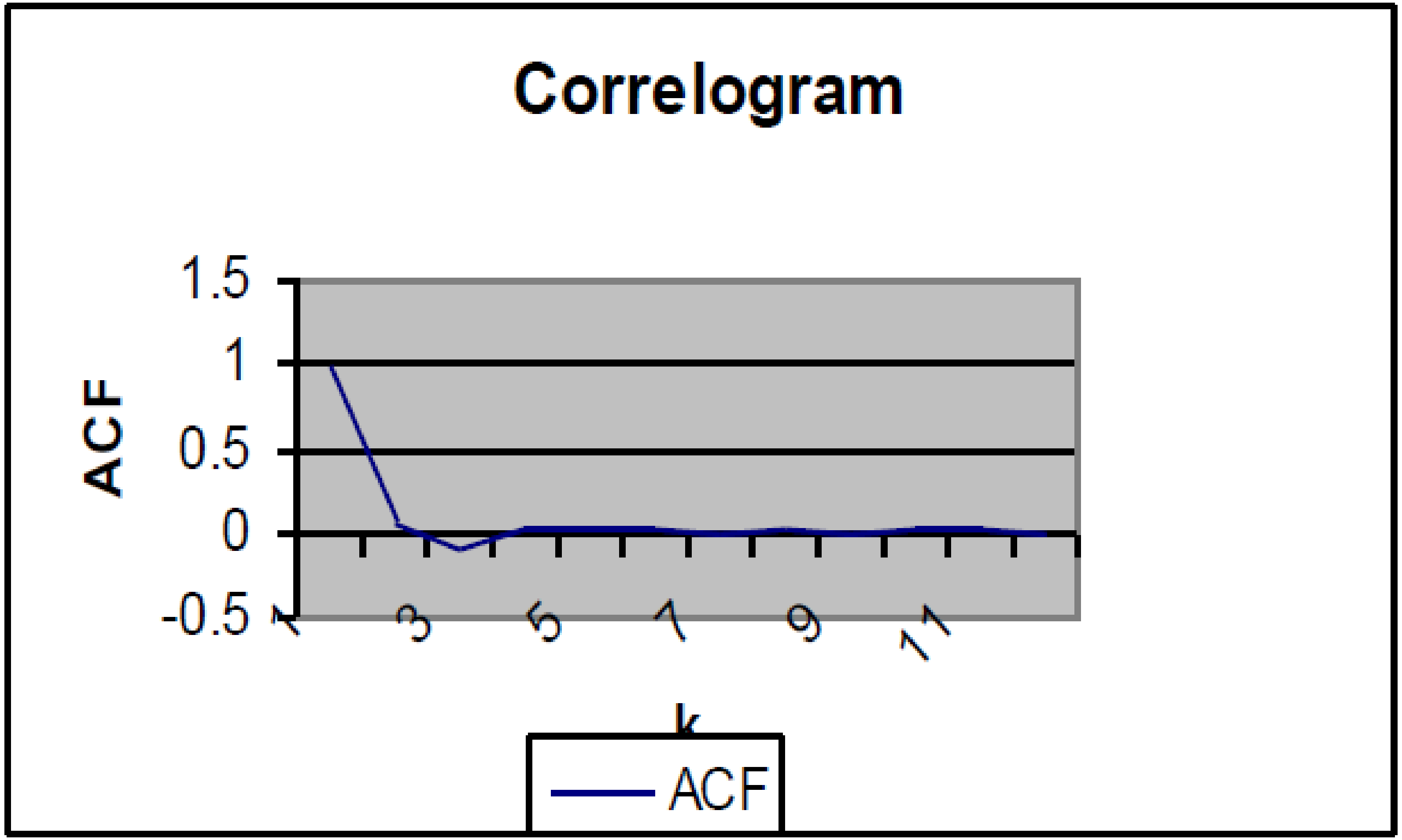
$$= \frac{\sum (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum (Y_t - \bar{Y})^2}$$

ρ_k – The ACF at lag k.

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.
- It can be used to determine stationarity, if the ACF falls immediately from 1 to 0, then equals about 0 thereafter, the series is stationary.
- If the ACF declines gradually from 1 to 0 over a prolonged period of time, then it is not stationary.

Stationary time series



Statistical Significance of the ACF

- The Q statistic can be used to determine if the sample ACFs are jointly equal to zero.
- If jointly equal to zero we can conclude that the series is stationary.
- It follows the chi-squared distribution, where the null hypothesis is that the sample ACFs jointly equal zero.

Q statistic

$$Q = n\sum_{k=1}^{m} \hat{\rho}_k^2$$

- n -> sample size
- m -> lag length
- $\chi^2(m)$ ->degrees of freedom

Ljung-Box Statistic

- This statistic is the same as the Q statistic in large samples, but has better properties in small samples.

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

Partial ACF

- The Partial Autocorrelation Function (PACF) is similar to the ACF, however it measures correlation between observations that are k time periods apart, after controlling for correlations at intermediate lags.
- This can also be used to produce a partial Correlogram, which is used in Box-Jenkins methodology (covered later).

Q-statistic Example

- The following information, from a specific variable can be used to determine if a time series is stationary or not.

$$\sum_{k=1}^{4} \hat{\rho}_k^2 = 0.32$$

$$n = 60$$

Q-statistic

$$Q = 60 * 0.32 = 19.2$$

$$\chi^2(4) = 9.488$$

$$19.2 > 9.488 \rightarrow reject - H_0$$

- The series is not stationary as the ACFs are jointly significantly different to 0.

Autoregressive Process

- An AR process involves the inclusion of lagged dependent variables.
- An AR(1) process involves a single lag, an AR(p) model involves p lags.
- AR(1) processes are often referred to as the random walk, or drift less random walk if we exclude the constant.

AR Process

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-1} + \ldots.\phi_p y_{t-p} + u_t$$

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + u_t$$

$$y_t = \mu + \sum_{i=1}^{p} \phi_i L^i y_t + u_t$$

L^{i} – lag operator

Moving Average (MA) process

- In this simple model, the dependent variable is regressed against lagged values of the error term.
- We assume that the assumptions on the mean of the error term being 0 and having a constant variance etc still apply.

MA process

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \ldots \theta_q u_{t-q}$$

$$y_t = \mu + \sum_{i=1}^{q} \theta_i u_{t-i} + u_t$$

$$y_t = \mu + \sum_{i=1}^{q} \theta_i L^i u_t + u_t$$

$$y_t = \mu + \theta(L) u_t$$

$$\text{Where}: \theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots \theta_q L^q$$

MA process

- The MA process has the following properties relating to its mean and variance:

$$E(y_t) = \mu$$

$$\operatorname{var}(y_t) = (1 + \theta_1^2 + \theta_2^2 + \ldots \theta_q^2)\sigma^2$$

Example of an MA process

$$\hat{y}_t = 0.7 + 0.8y_{t-1} + 0.3u_{t-1}$$
$$\quad (0.1) \quad (0.2) \quad\quad (0.1)$$

$$\overline{\mathrm{R}}^2 = 0.15, LM(1) = 2.47$$

$$y - output$$

$$u - error\,term$$

Example

- In the previous slide we have estimated a model using an AR(1) process and MA(1) process or ARMA(1,1) model, with a lag on the MA part to pick up any inertia in adjustment in output.
- The t-statistics are interpreted in the same way, in this case only one MA lag was significant.

Conclusion

- Before conducting a regression, we need to consider whether the variables are stationary or not.
- The ACF and Correlogram is one way of determining if a series is stationary, as is the Q- statistic
- An AR(p) process involves the use of p lags of the dependent variable as explanatory variables
- A MA(q) process involves the use of q lags of the error term.

References

Text Book:

“Business Analytics, The Science of Data-Driven Making”, U. Dinesh Kumar, Wiley 2017

Chapter-13

Image Courtesy

https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Time+Series+Analysis:+The+Basics

https://app.box.com/s/wr50f11slghr4vnvnqqrbjabnetehfyf

THANK YOU

Jyothi R
Assistant Professor, Department of Computer Science
jyothir@pes.edu