Time Series Analyisis –

The Concept of Co-integration using R

Contents

Concept of Co-integration

- The stock prices are believed to be **non-stationary** in nature.

- If there exists a linear combination of 2 stock prices , such that errors of the long term relationship are stationary (mean reverting) in nature, then the 2 stocks are said to be co-integrated.

- **Engle - Granger test** is used to identify the co-integrated pairs in a particular sector.

Co-Integration

Examples of Co-integrated Series

- Income and Consumption
- Money, National Incomes, Price Rates, Interest Rates
- Price of a Commodity and Taxes Levied on that Commodity, Inflation Rate
- Temperatures, Electricity Consumption
- Prices of Two Stocks

Co-Integration-Formal Definition

- **"Order of integration"** tells you the **minimum number of differences** needed to get a stationary series.
- **A series of successive differences, d,** can transform the time series into one with stationarity. The differences are denoted by **I(d)**, where **d is the order of integration**.

- Suppose Y_t and X_t are two time series integrated of order d, then
  - Any linear combination of such two series will also be integrated of order d (denoted as I(d)). This is called 'Integration'
- However,

> If there exists a vector β such that $u_t = Y_t - \beta X_t$ is of a lower order of integration (I(d – b), where b > 0) then Y_t and X_t are defined as Cointegrated of order (d,b)

- The idea of cointegration was introduced by **Engle and Granger** in 1987

* Usually, the order of integration is either I(0) or I(1); It's rare to see values for d that are 2 or more.

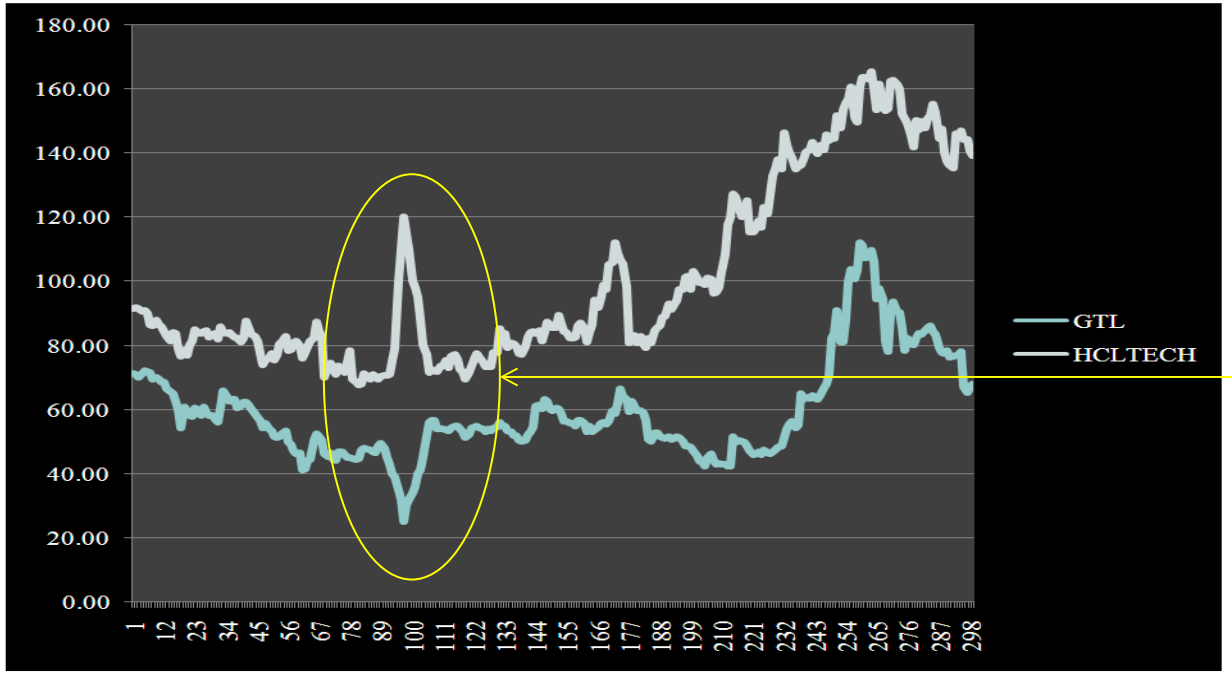
DATA SCIENCE INSTITUTE

What is Pairs Trading

- Pairs trading finds its roots in the area of securities trading
- A quest to generate returns, irrespective of market behavior, has led to the evolution of this strategy
- Certain securities, often competitors in the same sector, are associated in their day-to-day price movements. When the association breaks down, Traders take advantage of short term over pricing or under pricing of securities & invest accordingly, betting that the "spread" between the two would eventually converge
- Pairs Trading strategy works best in volatile market conditions

Simple Example of Pairs Trading

Fig. 1: Pairs Trading



This is the right opportunity for Pairs Trade

TRADER WILL GAIN PROFIT IF HE SELLS HCLTECH AND BUYS GTL

Co-Integration and Pairs Trading

- The **stock prices** are believed to be **non-stationary** in nature.
- **If there exists a linear combination of 2 stock prices ,** such that errors of the long term relationship are stationary (mean reverting) in nature, then the 2 **stocks are said to be co-integrated.**
- Pairs trading is in fact an extended application of the theory of co-integration.
- **Engle -Granger test** is used to identify the co-integrated pairs in a particular sector.

Engle Granger Test for Stock Prices

Let Y_t : TCS stock price at time t

Let X_t : IBM stock price at time t

Both the stock prices form non-stationary time series over time t.

If there exists a linear combination, such that the errors form stationary time series then TCS and IBM stocks are said to be co-integrated.

$$Y_t = \alpha + \beta X_t + \varepsilon_t$$

α = model intercept

β = regression coefficient

ε_t = error term

Engle Granger Test for Stock Prices

Regression coefficient β is estimated using ordinary least square (OLS) method

If $\beta < 0$ then the pair is discarded

Selection of regression equation is done by comparing β

$$Y_t = \alpha_1 + \beta_1 X_t + \varepsilon_t \quad \text{.....1}$$

$$X_t = \alpha_2 + \beta_2 Y_t + \varepsilon_t \quad \text{.....2}$$

The equation with higher β is selected

In case of TCS and IBM stocks $\beta 1 > \beta 2$ hence equation 1 was selected

Engle Granger Test

Engle Granger test is used to check stationarity of error time series.

Consider,

$$\Delta \varepsilon_t = \gamma^* \varepsilon_{t-1} + W_t$$

$\Delta \varepsilon_t = \varepsilon_t - \varepsilon_{t-1}$

ε_{t-1} = errors with lag 1

Objective	To test the **null hypothesis** that **time series is not stationary**

Null Hypothesis H0: Two stocks are not co-integrated

Alternate Hypothesis H1: Two stocks are co-integrated

Test Statistic	(y*/SE(y*)) Test statistic follows DF distribution under null
Decision Criteria	Reject the null hypothesis **if tcal < c(p)**

Engle Granger Test in R

```
# Installing and Loading Required Packages
install.packages("quantmod")
install.packages("egcm")
library(quantmod)
library(egcm)


#EG Test
getSymbols(c("tcs","ibm"), from =
                           env = .Gl
```

- **quantmod** is a very popular financial analysis package.
- **egcm** stands for Engle Granger Cointegration Models

- **getSymbols()** fetches data from various sources. The command requires us to specify start and end dates of the prices required.
- **Symbols** a character vector specifying the names of each symbol to be loaded
- **env** where to create objects.

- **ecgm()** performs a simple two-step EG test on a pair of time series.
- egcm(X,Y) considers Y on the LHS and X on the RHS

Engle Granger Test in R

```
# Output
IBM.Close[i] =   0.6258 TCS.Close[i] + 156.1917 + R[i], R[i] =   1.0000 R[i-1] + eps[i], eps ~ N(0,  1.6283^2)
               (0.8238)                (4.1724)               (0.0085)

R[2017-10-31] = -4.3906 (t = -0.364)

WARNING: TCS.Close and IBM.Close do not appear to be cointegrated.

Unit Root Tests of Residuals
                                                Statistic    p-value
  Augmented Dickey Fuller (ADF)                    -0.952    0.90049
  Phillips-Perron (PP)                             -2.319    0.91531
  Pantula, Gonzales-Farias and Fuller (PGFF)        0.988    0.80764
  Elliott, Rothenberg and Stock DF-GLS (ERSD)      -0.899    0.61297
  Johansen's Trace Test (JOT)                      -4.604    0.98910
  Schmidt and Phillips Rho (SPR)                   -2.290    0.93539

Variances
  SD(diff(TCS.Close))  =   0.200075
  SD(diff(IBM.Close))  =   1.634547
  SD(diff(residuals))  =   1.628260
  SD(residuals)        =  12.077787
  SD(innovations)      =   1.628260

Half life       = Infinite
R[last]         =  -4.390643 (t=-0.36)
```

Interpretation :

- Stocks TCS & IBM are not cointegrated.

?

What if we want to get data from other sources?

There are different packages using database APIs for data from other indices, eg. Quandl for India's National Stock Exchange (NSE) prices

Get an Edge!

There are a number of alternative approaches for testing co-integration, apart from the Engle Granger test, such as

- CRDW (Cointegrating Regression Durbin Watson) test ✉ Simple regression of one variable on the other, and the standard Durbin-Watson test on the residuals.
- Error Correction Test ✉ The tendency of cointegrated variables to revert to common stochastic trends is expressed in terms of error-correction
- Johansen's Multivariate VAR Approach ✉ Examines the number of independent linear combinations (k) for an m time series variables set that yields a stationary process

Quick Recap

In this session, we learnt about **co-integration in time series:**

Co-integration	• When the time series is integrated and linear combination of variables in that series is also integrated, but having an order lower than the whole series, then the variables are said to be co-integrated.
Pairs Trading	• When the association between two cointegrated stocks breaks down, Traders take advantage of short term over pricing or under pricing of securities & invest accordingly, betting that the "spread" between the two would eventually converge.
EG Test in R	• A combination of packages **quantmod** and **ecgm** is used to perform Engle Granger test in R, to check if a pair of time series is co-integrated.