The Concept of Co-integration

Contents

DATA SCIENCE
INSTITUTE

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.

DATA SCIENCE
INSTITUTE

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

DATA SCIENCE INSTITUTE

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.
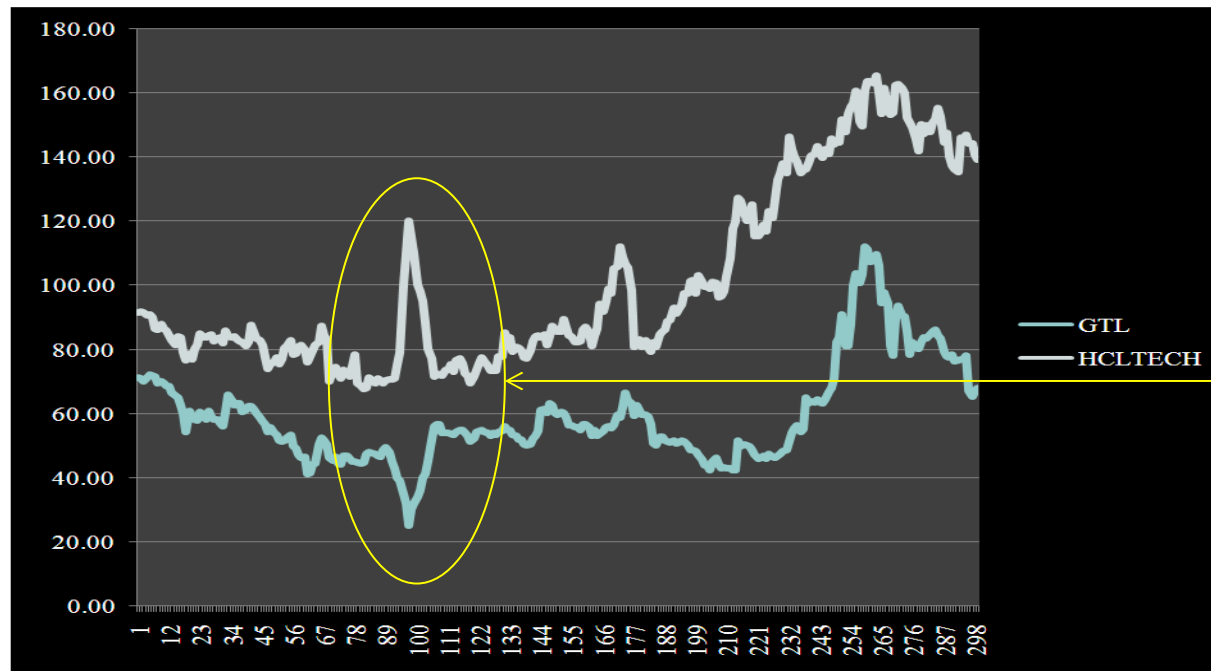
What is Pairs Trading

- 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

DATA SCIENCE
INSTITUTE

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

DATA SCIENCE INSTITUTE

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

DATA SCIENCE
INSTITUTE

Engle Granger Test for Stock Prices

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

If β < 0 then the pair is discarded

Selection of regression equation is done by comparing β

$$Y_t = \alpha_1 + \beta_1 X_t + \varepsilon_t \quad \ldots\ldots\ldots 1$$

$$X_t = \alpha_2 + \beta_2 Y_t + \varepsilon_t \quad \ldots\ldots\ldots 2$$

The equation with higher β is selected

In case of TCS and IBM stocks β1 > β2 hence equation 1 was selected

DATA SCIENCE
INSTITUTE

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	$(\gamma^*/SE(\gamma^*))$ Test statistic follows DF distribution under null
Decision Criteria	Reject the null hypothesis **if tcal < c(p)**

DATA SCIENCE INSTITUTE

Engle Granger Test in Python

```python
# Installing and Loading Required Libraries
import yfinance as yf
import pandas as pd


#EG Test

# Define the stock symbols for Wipro and Infosys
tickers = ["WIPRO.NS", "INFY.NS"]

start_date = "2022-01-01"
end_date = "2022-12-31"

# Download the historical data
data = yf.download(tickers, start=start_date, end=end_date)
adj_close_data = data['Adj Close']
adj_close_data.columns = ['INFY','WIPRO']

infy = adj_close_data.INFY
wipro = adj_close_data.WIPRO
```

- ❑ **yfinance** is a library that allows import of data directly from yahoo finance.

- ❑ **tickers** is a set symbols used to extract data.
- ❑ **start_date, end_date** is the start date and end date of the date range

Engle Granger Test in Python

```
# Import "coint" from "statsmodels"
from statsmodels.tsa.stattools import coint
coint(infy,wipro)
```

☐ **coint()** is a function that runs the engel garner test and returns t-statistic, p-value, Critical values for the test statistic at the 1 %, 5 %, and 10 % levels based on regression curve.

```
# Output
(-2.154422825452808,
 0.44797584761041354,
 array([-3.94132922, -3.36097908, -3.06166543]))
```

Interpretation :
➢ As tcal > c(p) at 5%, we accept null Hypothesis.
➢ Stocks INFY & WIPRO 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. nsepy for India's National Stock Exchange (NSE) prices

13

DATA SCIENCE INSTITUTE

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

DATA SCIENCE
INSTITUTE

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 Python	• **coint()** is used to perform Engle Granger test in Python, to check if a pair of time series is co-integrated.

DATA SCIENCE
INSTITUTE