

# Tech Mahindra - Solution 3

*Vasim Shaikh*

## Business Problem Framing

An investment company, XYZ Inc. is planning to make a substantial investment in Coca-Cola (KO) stock. The team need to understand quantitative returns and risk associated with the stock and overall market.

The team believes that higher trade volume relative to previous trading period tends to reduce the returns of S&P500 and need a scientific approach to their belief.

## Analytic problem framing

Returns and risks can be measured over different period. This analysis consists of daily, weekly, monthly and quarter period.

- 1) Analyze returns - Returns in its simplest terms means, how much money has being made?  
Arithmetic average or log returns are two of the preferred method. Others being trimmed mean, harmonic mean etc.
- 2) Analyze market risk - Market risk, is the uncertainty inherent to the entire market. Examples include interest rates, recession and wars.  
Beta is a numeric value that measures the fluctuations of a stock to changes in entire market.
- 3) Hypothesis testing - Higher trade volume relative to previous trading period tends to reduce the returns of S&P500.  
Ha = If there is an increase in volume from previous period, on average the returns of S&P500 decreases.  
Ho = If there is an decrease in volume from previous period, on average the returns of S&P500 increases.  
We will use the Welch t-test statistic for testing our hypothesis.

## Data

Two flat files are available - historical records for S&P500 and Coca-Cola (KO).

As OHLC figures prior to 1961 are same, they are removed.

Preprocessing steps such as adjustment for split and bonus, missing values, duplicate records and outlier has been validated.

## Analysis

### Returns

As discussed, the common approach for stock returns are arithmetic mean or log returns. The calculation are as follows:

Arithmetic mean =  $\text{sell price} - \text{buy price} / \text{buy price}$

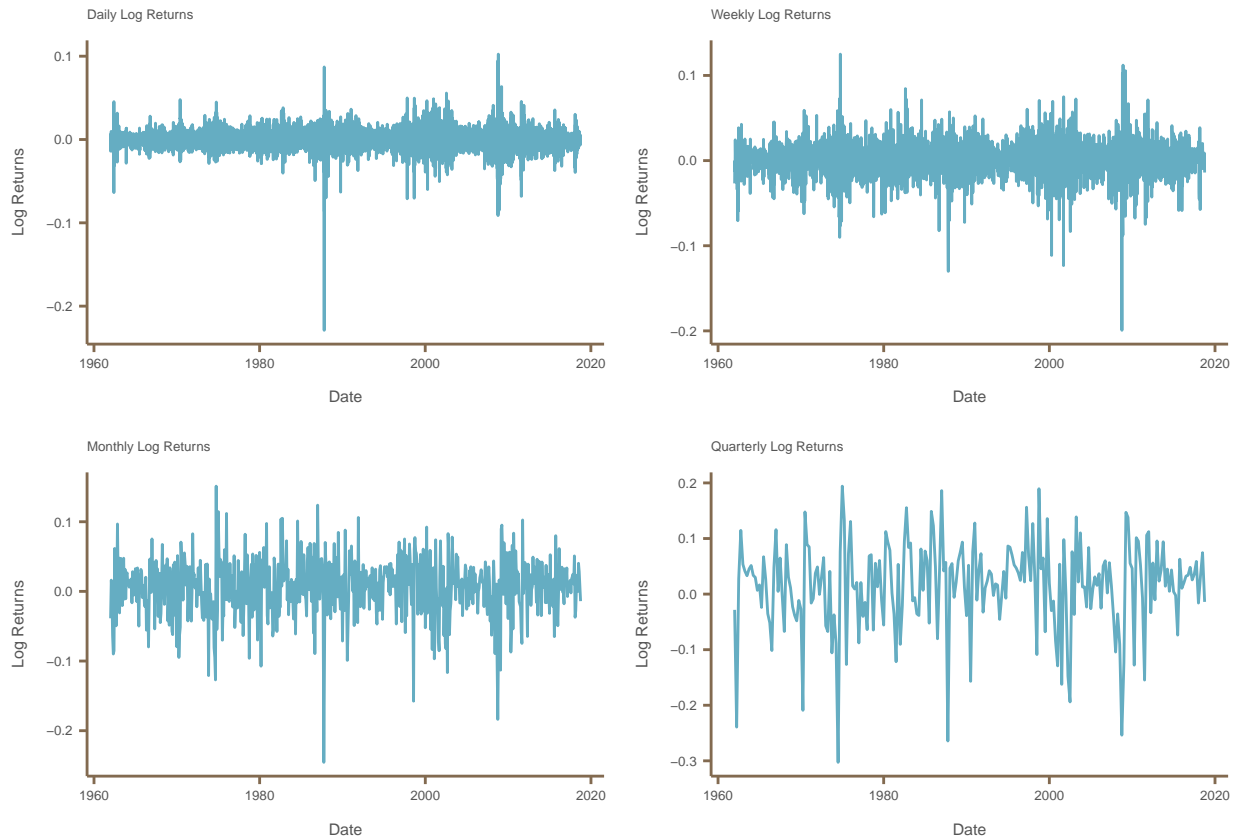
Log return = Natural log of  $(\text{sell price} / \text{buy price})$

In the following table we notice that arithmetic returns are biased towards positive, hence we use log returns.

Table 1: -

Sr no	Buy Price	Sell price	Arithmetic return	Log return
1	200	100	-100%	-30%
2	100	200	50%	30%

### Visualizing S&P500 returns



Note: Interactive visualization of S&P500 returns can be found in MS Excel file.

## Expected Price

Assuming stock returns follow a **normal distribution**.

We can be 68% confidence (one standard deviation from average) that S&P500 price will be between 2,644 and 3,608, for next year.

Alternatively, we can be 95% confidence (two standard deviation from average) that S&P500 price will be between 2,264 and 4,215, for next year.

On similar line, we can be 68% confident that Coca-Cola stock will be between 41.3 and 62.2 and 95% confident that it will be between 33.7 and 76.4 for the next year.

### Note

Based on 5th October 2018; S&P @ 2,903 and KO @ 45.9.

For calculation see appendix.

## Market Risk

A stock's beta is calculated as follows:

$$beta = \frac{Covariance(KO, S\&P500)}{Variance(SP)}$$

For every one rupee change in S&P500, Coca-Cola changes by following percent.

Table 2: -

Period	Beta
Daily	79%
Weekly	80%
Monthly	78%
Quarterly	90%

From the above table we conclude that Coca-Cola stock is less risky compared to overall market.

### Note

For calculation see appendix.

Interactive visualization of S&P500 returns can be found in MS Excel file.

## Hypothesis Testing

T-Test has the following three assumptions

- Normal Distribution - not standard normal distribution though.
- n greater than 30 - for both group
- Homoskedasticity (equal variance between group) - In Welch t-test this assumption is relaxed.

### T-test

On September 25, 1995 - NASDAQ started automated trading, thus giving birth to trading from desk. Therefore we split our dataset on this year and perform hypothesis testing for later period.

To iterate:

Ha = If there is an increase in volume from previous period, on average the returns of S&P500 decreases.

Ho = If there is an decrease in volume from previous period, on average the returns of S&P500 increases.

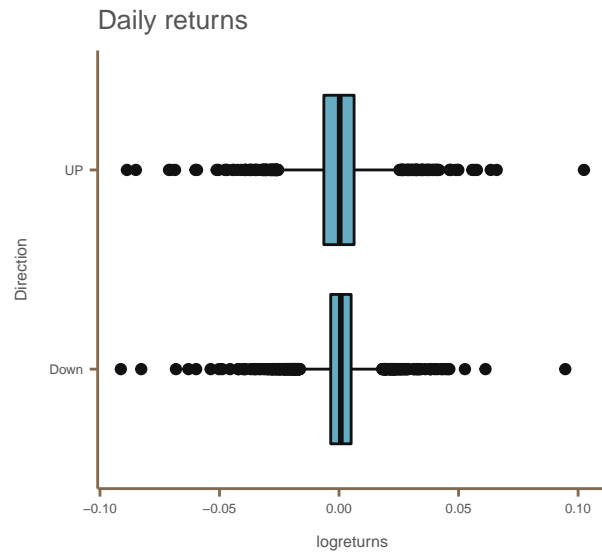
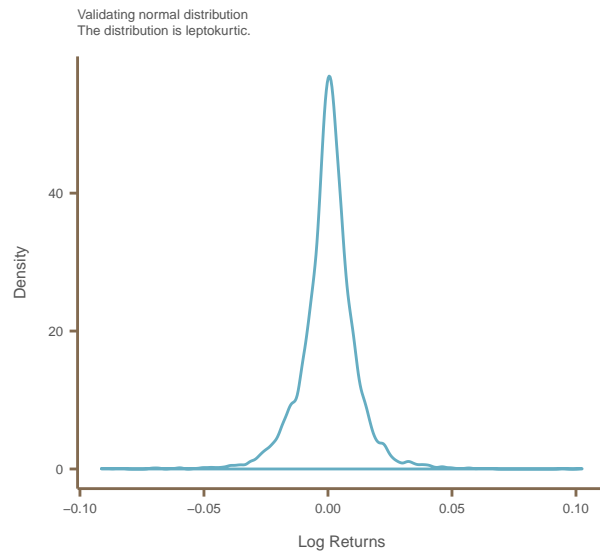


Table 3: T-Test Result

static	value	Period	Sig. Level	Test Result
p_value	0.0799	Daily	0.05	Fail to Reject Null Hypothesis
p_value	0.0179	Weekly	0.05	Reject Null Hypothesis
p_value	0.0242	Monthly	0.05	Reject Null Hypothesis
p_value	0.0350	Quarterly	0.05	Reject Null Hypothesis

From above table we conclude that higher volume over different period tends to reduce returns, except on daily basis. Even for daily volume, since p-value is near significance level we can't confidently fail to reject null hypothesis.

## Appendix

### Calculation of confidence interval of returns

```
# 252 - working days in a year  
# Change 252 period to 22 for monthly calc and so on.  
# currentprice is considered as last day of csv file.
```

```
SandP_68_perc <- sp %>%  
  select(Date, Open, Close) %>%  
  mutate(logreturns = log(Close/Open)) %>%  
  summarise(avg = mean(logreturns) * 252,  
            sd = sd(logreturns) * sqrt(252),  
            currentprice = 2903) %>%  
  mutate(Upper = exp(avg + sd) * currentprice,  
         Lower = exp(avg - sd) * currentprice)  
  
SandP_95_perc <- sp %>%  
  select(Date, Open, Close) %>%  
  mutate(logreturns = log(Close/Open)) %>%  
  summarise(avg = mean(logreturns) * 252,  
            sd = sd(logreturns) * sqrt(252),  
            currentprice = 2903) %>%  
  mutate(Upper = exp(avg + 2 * sd) * currentprice,  
         Lower = exp(avg - 2 * sd) * currentprice)  
  
coke_68_perc <- ko %>%  
  select(date, open, close) %>%  
  mutate(logreturns = log(close/open)) %>%  
  summarise(avg = mean(logreturns) * 252,  
            sd = sd(logreturns) * sqrt(252),  
            currentprice = 45.9) %>%  
  mutate(Upper = exp(avg + sd) * currentprice,  
         Lower = exp(avg - sd) * currentprice)  
  
coke_95_perc <- ko %>%  
  select(date, open, close) %>%  
  mutate(logreturns = log(close/open)) %>%  
  summarise(avg = mean(logreturns) * 252,  
            sd = sd(logreturns) * sqrt(252),  
            currentprice = 45.9) %>%  
  mutate(Upper = exp(avg + 2 * sd) * currentprice,  
         Lower = exp(avg - 2 * sd) * currentprice)
```

## Beta Calculation

An interactive visualization can be found in Excel file.

*#The code provides alpha (Intercept) and beta (spchange) figures.*

```
dailyBeta <- sp %>%
  left_join(ko, by = c("Date" = "date")) %>%
  mutate(spchange = (Close - lag(Close, 1)) / lag(Close, 1) ) %>%
  mutate(kochange = (close - lag(close, 1)) / lag(close, 1) ) %>%
  select(spchange, kochange) %>%
  lm(kochange ~ spchange, data = ., na.action = na.omit)

weeklyBeta <- sp %>%
  left_join(ko, by = c("Date" = "date")) %>%
  as_period(period = "weekly") %>%
  mutate(spchange = (Close - lag(Close, 1)) / lag(Close, 1) ) %>%
  mutate(kochange = (close - lag(close, 1)) / lag(close, 1) ) %>%
  select(spchange, kochange) %>%
  lm(kochange ~ spchange, data = ., na.action = na.omit)

monthlyBeta <- sp %>%
  left_join(ko, by = c("Date" = "date")) %>%
  as_period(period = "monthly") %>%
  mutate(spchange = (Close - lag(Close, 1)) / lag(Close, 1) ) %>%
  mutate(kochange = (close - lag(close, 1)) / lag(close, 1) ) %>%
  select(spchange, kochange) %>%
  lm(kochange ~ spchange, data = ., na.action = na.omit)

quarterlyBeta <- sp %>%
  left_join(ko, by = c("Date" = "date")) %>%
  as_period(period = "quarterly") %>%
  mutate(spchange = (Close - lag(Close, 1)) / lag(Close, 1) ) %>%
  mutate(kochange = (close - lag(close, 1)) / lag(close, 1) ) %>%
  select(spchange, kochange) %>%
  lm(kochange ~ spchange, data = ., na.action = na.omit)
```