JMP042: US Stock Indices

Differencing, Log Transformation, Stationarity, Augmented Dickey Fuller (ADF) Test

Produced by

M Ajoy Kumar, Associate Professor,
Siddaganga Institute of Technology
ajoy@sit.ac.in

Muralidhara A, JMP Global Academic Team
muralidhara.a@jmp.com

US Stock Indices

Differencing, Log Transformation, Stationarity, Augmented Dickey Fuller (ADF) Test

Key ideas

This Case Study requires the use of Time Series Analysis platform of JMP to understand the basic concepts related to time series data analysis. It explores the ways to practically understand the risks and rate of return related to the financial indices data.

Background

After completing graduation degree in Finance, James joined a leading Investment Advisory firm in New York as intern. He was reporting to Sarah, a Senior Analyst at the firm. James was involved in a research project on US Stock Markets. As the first assignment Sarah wanted James to analyse the performance of US Stock Markets during 2016-2019. The specific tasks given to James included analysing the rates of returns and risks of four major markets and comparing their performances. Since Sarah wanted to use this data for further analysis and forecasting, she wanted James to also analyse the stationarity of the indices of the markets.

The Task

James needs to analyse the performance of stock market for the financial indicators computed for four indices. He must derive the following metrices.

- Daily rates of returns
- Risk measured as standard deviation of daily returns
- Comparison of average returns and risk

After deriving, he wants to visually explore the values of indices and daily rates of returns, followed by examination of the stationarity of the data series.

Data US_INDICES.jmp

James collected daily data of four US stock indices, Dow Jones Industrial Average (DJIA), NASDAQ Composite Index, S&P 500 and NYSE Composite Index. The data was collected for a period of four years (01-01-2016 to 31-12-2019) from Yahoo Finance.

Date	Trading days for which the closing value of the indices are collected
DJIA	The Dow Jones Industrial Average, Dow Jones, or simply the Dow, is a stock market index that measures the stock performance of 30 large companies listed on stock exchanges in the United States.
NASDAQ	The NASDAQ Composite is a stock market index that includes almost all stocks listed on the Nasdaq stock market. This index is weighed more towards Technology related companies.
S&P 500	The S&P 500, or simply the S&P, is a stock market index that measures the stock performance of 500 large companies listed on stock exchanges in the United States.
NYSE	The NYSE Composite is a stock market index covering all common stock listed on the New York Stock Exchange

Except Date, all variables represent continuous time series data.

Analysis of Daily Rates of Returns

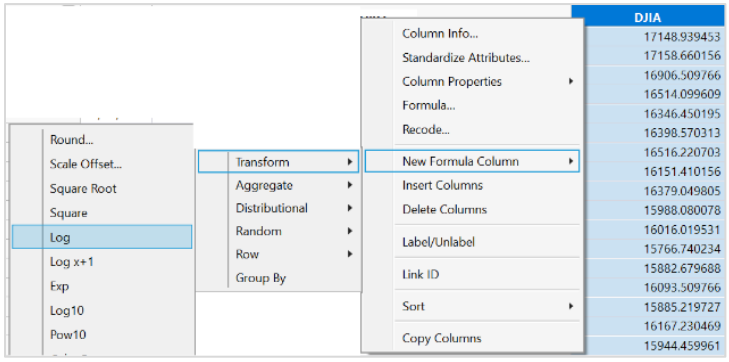
Rates of returns are defined as the difference between the natural log of prices (values) of two successive periods.

$$r = l_n(P_{t+1}) - l_n(P_t)$$

Here 'ln' indicates natural log, 'P' indicates price and 't' indicates time.

To compute the daily rates of returns, a new series representing log transformation of the base data of indices needs to be generated. Further, the daily rates of return series are generated as the first difference of the log series.

Exhibit 1 Computing Log Transformed values of the Index



(To compute natural log of the base data, select the series, *right click* > *New Formula Column* > *Transform* > *Log* as shown in Exhibit 1. This will create a new column with Log Transformed series of the base data)

To generate the first difference of the log series, select the newly formed log transformed series, right click>New Formula Column>Transform>Row>Difference. This will create a new column with the first difference of the log transformed data. This newly formed column with name Difference[Log[DJIA]] is Daily Rates of Return of DJIA. Rename the column to DRR_DJIA

One can also directly compute the daily rates of return using the formula option. Add a new column using Column >New Columns > Colum Properties > Formula > input the following formula (Dif(("Log[DJIA"))) to get the rates of returns.

Repeat the same steps to calculate the daily rates of return for NASDAQ and populate a new column DRR_NASDAQ.

Also create a new variable called 'year' using New formula Column > Transform > Date Time > Year. This will create a new variable Called Year. Change the Modeling type of Year variable to Ordinal by right clicking.

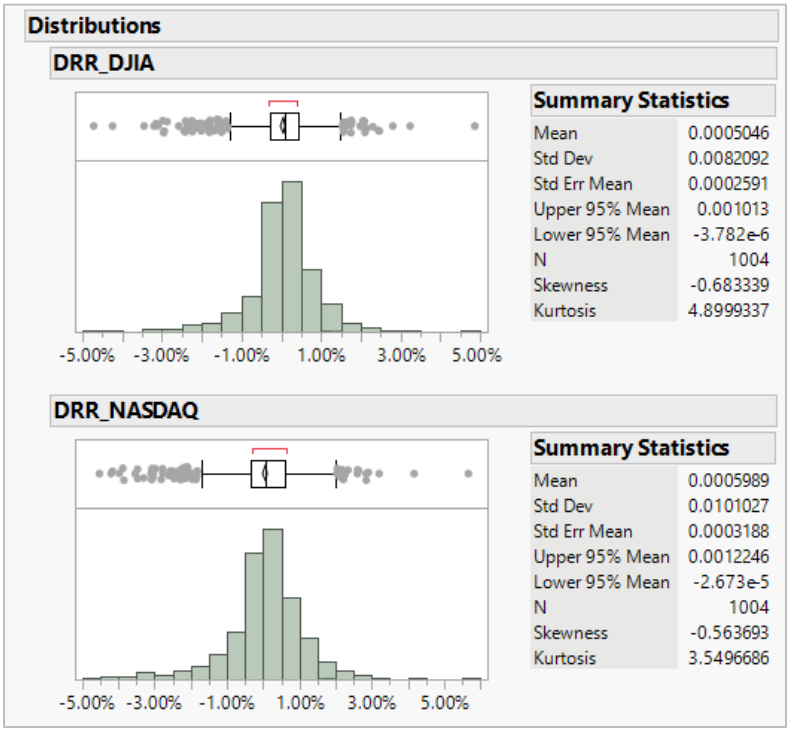
Average Returns, Risk and Other Characteristics

The basic characteristics of rates of return are analysed using the average daily returns over a period, risk measured by standard deviation, skewness and kurtosis of returns.

Analyze > Distribution > Y = DRR_DJIA and DRR_NASDAQ > ok

The characteristics of returns on DJIA and NASDQ indices are shown in Exhibit 2 as part of Summary statistics, box plot and histogram. Under the summary statistics drop down, select customize summary statistics option and select skewness and kurtosis.

Exhibit 2 Summary Statistics of Indices



The daily rates of return on financial assets are expected to exhibit zero mean and non-zero standard deviation. Exhibit 2 shows that the mean of daily rates of returns is very close to zero for both the indices. The risk measured as standard deviation is 0.0082 (0.82%) for DJIA and 0.0101 (1.01%) for NASDAQ. The returns on NASDQ are more volatile compared to DJIA. Skewness and kurtosis are measured to understand the nature the distribution of returns. Most of the financial return series are usually negatively skewed. Same is the case with DJIA and NASDQ with DJIA showing higher degree of skewness. The financial return series are found to have sharp peaks with fatter tails. Such distributions with a measure of kurtosis above 3 are called leptokurtic. The daily returns on DJIA and NASDAQ are leptokurtic.

Year wise Rates of Returns and Risk

The return series of four years is divided into individual years and the characteristics of the data series is analysed for each year. Rate of return over one-year period is an additional measure used here. The rate of return during a period is defined as:

$$r_n = \sum_{t=1}^{n} \left[[l_n(P_{t+1})] - [l_n(P_t)] \right]$$

Thus, the rate of return over a period of ‘n’ days is the summation of daily rates of returns over ‘n’ days.

Exhibit 3 Annual Rates of Return

Tabulate

	DRR_DJIA			DRR_NASDAQ		
Year	Mean	Std Dev	Sum	Mean	Std Dev	Sum
2016	0.001	0.008	0.142	0.000	0.010	0.093
2017	0.001	0.004	0.224	0.001	0.006	0.249
2018	-0.000	0.011	-0.058	-0.000	0.013	-0.040
2019	0.001	0.008	0.199	0.001	0.010	0.299

Year wise summary statistics for DJIA and NASDAQ are created using Tabulation. Analyze > Tabulate > Years as Dropdown for Rows, DRR_NASDAQ and DRR_DJIA as Resulting Cells, and select Mean, Standard Deviation and Sum from the statistical measures list to arrive at the table describing the annual returns of both the markets as shown in Exhibit 3.

From Exhibit 3, it is observed that, the mean daily returns are close to zero for both the markets in each year. The volatility of markets, captured by standard deviation, is moderately higher for NASDAQ compared to DJIA in all the four years.

The annual rates of returns are measured as the total of daily log returns. DJIA performed better in 2016, while NASDAQ recorded higher rates of returns in 2017 and 2019. 2018 was a year of loss for both the markets, with DJIA reporting greater loss than NASDAQ. When the markets recovered in 2019, NASDAQ performed much better than DJIA.

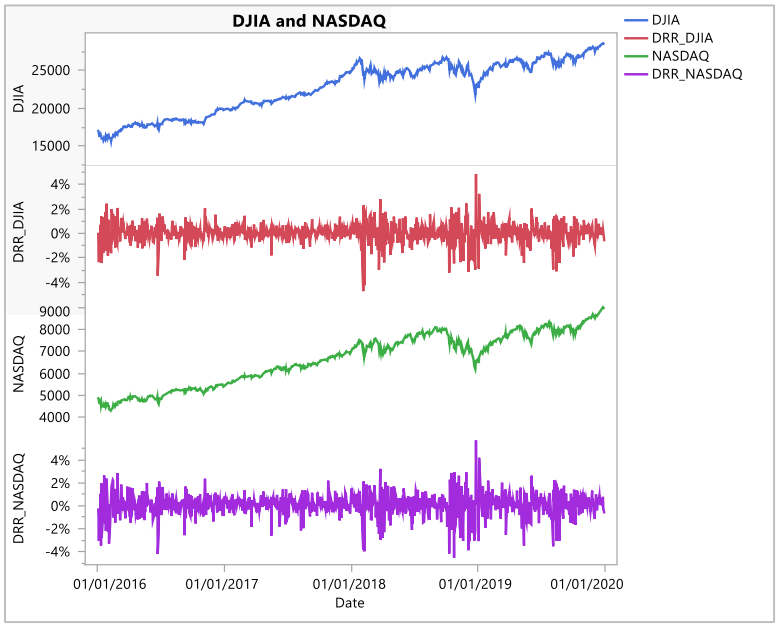
Stationarity of Data Series

Stationarity is an important concept in time series data analysis. A stationary process is one where the mean, variance, and autocorrelation structure does not change over time. (Autocorrelation measures the relationship between a variable's current value and its past values). In case of a non-stationary series, mean and variance depend on the time at which they are measured. A non-stationary series is said to exhibit unit root property. Most of the advanced time series analytical tools are applied on stationary data. Hence, testing for stationarity of data series is the primary step in time series modelling.

If a data series is non-stationary, it must be converted to a stationary series before applying time series modelling tools. Differencing the data series is one of the ways of transforming a non-stationary series to stationary. Some data series turn stationary on first differencing, while some others require second or third differencing. Most of the financial price series are non-stationary, whereas the rates of returns are stationary. So, the first difference of log of price series, which is also the rate of return, usually does not have unit root. Stationarity of data series can be examined by observing the behaviour of the series over time through graphical representation. However, a more accurate method is to apply statistical tests for unit root.

Let us explore the indices and daily rates of returns using graph builder.

Exhibit 4 Line Graphs of Indices and Daily Rates of Return



Graph Builder > Drop Date to X, Drop DJIA to Y. Further keep dropping DRR_DJIA, NASDAQ and DRR_NASDAQ one by one to the bottom of the Y axis. Select Line graph.

Close observation of Exhibit 4 shows that both the indices are non-stationary as they exhibit trending behaviour over time. However, the daily rates of returns measured as the first difference of log of index value are stationary for both the indices. Further analysis for confirming this is carried out by applying statistical tests.

Augmented Dickey Fuller (ADF) Test

Augmented Dickey Fuller (ADF) Test is a popular test used to analyse the unit root property of time series data. The null hypothesis for ADF test is that the series has unit root, or the series is non-stationary.

ADF test returns three test statistics related to tests of stationarity:

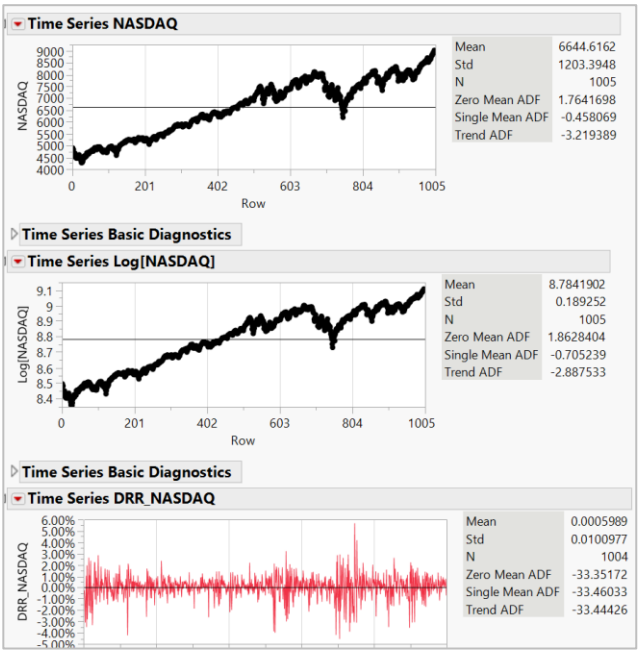
- Zero Mean ADF - A test against a random walk with zero mean
- Single Mean ADF - A test against a random walk with a non-zero mean
- Trend ADF - A test against a random walk with a non-zero mean and a linear trend

The test statistic is expected to be negative; therefore, it must be more negative (less) than the critical value for the hypothesis to be rejected.

The values generated for the Zero Mean, Single Mean and Trend ADF in JMP are the Tau statistics associated with the Dickey-Fuller test. The critical values for the ADF test at 5% level are -2.86 without trend and -3.41 with trend for large samples.

Let us apply these tests for DJIA and NASDAQ series.

Exhibit 5 ADF Test Results of DJIA

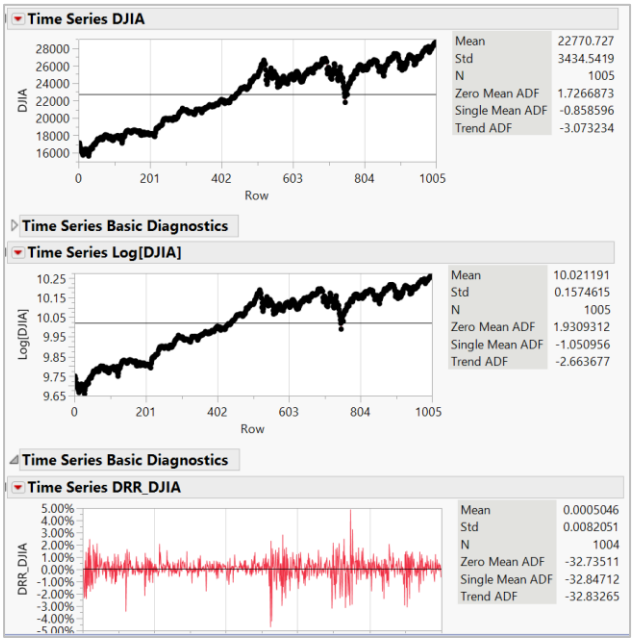


Analyze > specialized Modeling > Time Series > Drop DJIA, Log DJIA and DRR_DJIA as Y > ok. Since the data is time ordered, you can ignore the X, Time ID. By default, JMP considers it as Time series data.

The results are shown in Exhibit 5.

Follow similar steps to derive the values for NASDAQ.

Exhibit 6 ADF Test Results of NASDAQ



Since most of the daily rates of return series in financial markets have zero mean, Zero Mean ADF is considered for analysis.

Exhibit 5 and 6 show that the ADF test statistic for DJIA and Log of DJIA is 1.72 and 1.93 respectively. These values are greater than the critical value. Thus, null hypothesis that the series has a unit root is accepted and DJIA and Log of DJIA are found to be Non-stationary. However, the test statistic of daily returns of DJIA (DRR_DJIA) is -32.74, which is less than the critical value. Thus the null hypothesis that the series has a unit root is rejected. So, the daily rates of returns of DJIA is found to be stationary.

Similarly, NASDAQ and Log of NASDAQ series are found to be non-stationary, whereas Daily returns of NASDAQ, i.e. DRR_NASDAQ is stationary.

Summary

Statistical Insights

The case described the basic analysis of financial time series using JMP by taking example of two US indices, DJIA and NASDAQ for a period of four years. The scheme of analysis was

- Transformation of indices – Log and First Difference
- Analysing the summary statistics like mean, standard deviation, skewness, and kurtosis for the entire period.
- Tabulating the summary statistics year wise
- Visually exploring the data for stationarity using graph builder.
- Testing the unit root property using ADF test.

Implications

James can draw the following conclusions from the analysis:

- Both the indices (DJIA and NASDAQ) have shown upward trend during the four-year period (2016-19).
- The daily rates of returns are close to zero.
- Volatility of NASDAQ is marginally higher than that of DJIA.
- The daily rates of returns are negatively skewed and are leptokurtic.
- Both the markets reported positive annual rates of returns, except in 2018.
- The indices are non-stationary, whereas the daily rates of returns are stationary.

Based on the conclusions of James, Sarah can use the daily data of the US markets to carry out further analysis leading to forecasting.

JMP® Features and Hints

This case used the Distribution platform to display histograms and summary statistics and Graph builder to visualize the data in a time series manner. Time series analysis which is under Specialized Modeling Platform was used to conduct Augmented Dickey Fuller Test for stationarity.

Transformations are applied to create new columns followed by checking for the stationarity of the data series.

Exercise

Perform the time series analysis as described in the case for S&P 500 and NYSE series.

- Perform log transformation for all indices.
- Compute the daily rates of returns for all indices.
- Compare the characteristics of values of indices and daily rates of returns.
- Generate the summary statistics for each year for all the four indices.
- Compute the annual rates of returns for each year for all the indices and compare.
- Comment on stationarity of indices and daily rates of returns using graph.
- Analyse the unit root property using ADF Test and comment on the results.