S&P Global Quarterly Composite Index Modeling

**Introduction**

An important task of a financial analyst is to quantify costs associated with future cash flows. We consider here funds invested in a standard measure of overall market performance, the Standard and Poor’s (S&P) 500 Composite Index. The goal is to forecast the performance of the portfolio for discounting of cash flows. In particular, we examine the S&P Composite Quarterly Index for the years 1936 to 2020, inclusive.

The response variable chosen is Log returns of S&P Global composite index. Instead of forecasting the actual index, the research is targeted in forecasting the differenced logarithmic series. The reason is that the actual series has a unit root. It can be observed using Augmented Dicky Fuller test that after first order differencing the log returns the series becomes stationary. Moreover, the difference of logarithm can be interpreted as proportional changes. The independent variables for the research question are the first lag of the differenced logarithmic series. The independent variables are the first lag and the dummy variable after2008 (1 = during or after 2008 or 0 = before 2008)

**Data Description**

The below table gives a detailed description of the dataset.

|  |  |  |
| --- | --- | --- |
| File Name: Number of Number of SP500Quarterly obs: 284 variables: 5 | Number of SP500Quarterly obs: 339 | Number of  variables: 5 |
| Variable | Num of Obs Missing | Description |
| YEAR  SPINDEX  DIFFINDEX  LNSPINDEX  DIFFLNSP |  | Year  The Standard and Poor’s (S&P) 500 Composite Index  The difference of the SPINDEX between this year and last year  The natural logarithm of SPINDEX  The difference of LNSPINDEX between this year and last year |

Source: Center for Research on Security Prices, University of Chicago and Yahoo Finance

**Exploratory Data Analysis:**

*Comparative plot for 4 variables in data*

Chart

Description automatically generated

Figure : Comparative plot

From the original index values in the upper- left-hand panel, we see that the mean level and variability increase with time. This pattern clearly indicates that the series is nonstationary.

The time series plot of the differences, in upper-right-hand panel, still indicates a pattern of variability increasing with time.

An alternative transformation is to consider logarithmic values of the series. The time series plot of logged values, presented in lower-left-hand panel of Figure 1, indicates the mean level of the series increases over time and is not level. Thus, the logarithmic index is not stationary.

Another approach is to take the difference of log of index. This is especially desirable when looking at indices, or “breadbaskets”, because the difference of logarithms can be interpreted as proportional changes. From the final time series plot, in the lower-right-hand panel of Figure 1, we see that there are fewer discernible patterns in the transformed series, the difference of logs. This transformed series seems to be stationary. Hence, we consider DIFFLNSP as the response variable.

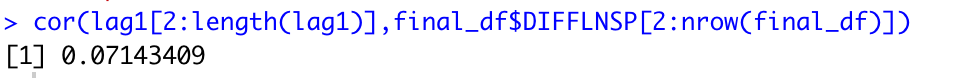
*Exploring the relationship between first lag of DIFFLNSP and DIFFLNSP*

Chart, scatter chart

Description automatically generated

Figure : Relationship between first lag of DIFFLNSP and DIFFLNSP

From Figure 2 we can see that there is weak linear relationship between the first lag and DIFFLNSP. The correlation coefficient between lag1 and DIFFLNSP is 0.07 which confirms the aforementioned statement.



Summary statistics:

A picture containing text

Description automatically generated

As we see the mean becomes close to zero as from SPINDEX to DIFFLNSP confirming with the observation from the graphical analysis that the process is close to white noise.