CORNELL UNIVERSITY

College of Engineering

Operations Research and Information Engineering

EXPLORING RELATIONSHIPS AMONG INTERNATIONAL STOCK MARKETS

ORIE 5640

Professor David S. Matteson

Prepared by

Ying Guan (yg532)

Jingxuan Li (jl4267)

Shaohan Wang (sw2265)

May 14, 2021

Introduction

The stock market index is one of the most crucial metrics in reflecting a stock market's overall performance. In this report, we explore and analyze the performance of four international stock market indices based on their closing prices and net returns across time. The four selected international market indices are US S&P 500 (GSPC), Euro STOXX (STOXX), Hong Kong Hang Seng index (HSI) and Japan Nikkei 225 (N225). The selected time frame is between Jan 1, 2006 and Apr 30, 2021. The project is mainly divided into three parts. In the first part, we explore the appropriate multivariate distribution and copula to market indices' returns and test the goodness of fit. In the second part, we apply time series analysis and forecast the future trend and movements. In the third part, we relate to the theory of cointegration and explore possibilities of applying pairs trading methods to these stock indices.

Multivariate and Copula Analysis

The daily closing prices of the four selected international stock market indices, GSPC, STOXX, HSI, N225, are extracted from "Yahoo! Finance" by "quantmod" package in R. Then, we could transform the respective closing prices into simple returns and combine into one big matrix for further analysis.

We first check the normality of each individual return vector by applying Shapiro-Wilks normality test with the null hypothesis of sampling from normal distribution. The Shapiro-Wilks statistic is:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2},$$

The p-values for all these four return vectors are fairly close to 0 (details in Appendix 1) and we could believe they do not follow the univariate normal distribution. Then, we apply the Kolmogorov–Smirnov test, with the null hypothesis of sampling from the objective distribution, i.e. the t distribution. The Kolmogorov–Smirnov statistic is:

$$D_n = \sup_x |F_n(x) - F(x)|$$

We could find that the p-values for all these four return vectors are strictly greater than 0.05 (details in Appendix 2). Thus, we may conclude that each net return follows the univariate t distribution and we could obtain the fitted "std" t distribution parameters, i.e. mean μ, standard deviation σ and tail index ν for each individual return vector by applying "fitdist()" function.

Then, since each return vector might follow the univariate t distribution, we could take a step to try to fit the multivariate t distribution by the profile likelihood method. The Profile likelihood is often used when accurate point or interval estimates are difficult to obtain using standard methods, i.e. when the log-likelihood function has non-normal shape[1]. To determine the Maximum Likelihood Estimates of the three estimators in the multivariate t distribution, i.e. mean vector μ, scale matrix Λ and tail index ν, we repeatedly apply the "cov.trob()" function with respect to a series of potential tail index ν and calculate the corresponding profile log-likelihoods. Then, we locate the optimal value of the tail index which gives the maximum profile log-likelihood and reapply the "cov.trob()" function to determine the MLE of the other two candidates. Thus, we have successfully fitted the multivariate t distribution to our selected international stock market indices' net returns.

Then, we want to explore more about the possibility if the given multivariate distribution is a meta-t distribution. We need to fit the t-copula first. A copula C of the random vector Y is a function that maps the univariate marginal distributions' cumulative distribution function to the joint distribution F[2]. The copula of random vector Y is:

$$C_Y(u_1, \ldots, u_d) = P\{F_{Y_1}(Y_1) \leq u_1, \ldots, F_{Y_d}(Y_d) \leq u_d\}$$

To determine the input correlation matrix of the t-copula, we first find Kendall's tau correlation coefficients and transform them into the correlation matrix of our net return matrix. According to the lecture note, a copula is a multivariate distribution given the uniform marginal distributions, so we need to transform the net returns into a univariate uniform by plugging themselves into the corresponding estimated univariate std t-distribution cumulative distribution function. Then, we could apply the function "tCopula()" and "fitCopula()" to fit the t-copula to our net return matrix. Finally, we could define and fit a meta-t distribution to our net return matrix given the "std" marginal distribution by applying the function "mvdc()" and "fitMvdc()".

In conclusion, according to the estimated copula correlation coefficients (details in Appendix 3) in our fitted meta-t distribution, we could find that the correlation between GSPC and STOXX's net returns are higher than all the other estimated correlations. Besides, HSI and N225's net returns are also highly correlated. One of the reasons might be that those indices are in the same "big" market, i.e. HSI and N225 are both in Asian market, which could affect their performance and increase the possibility of "co-movement".

Time Series Analysis

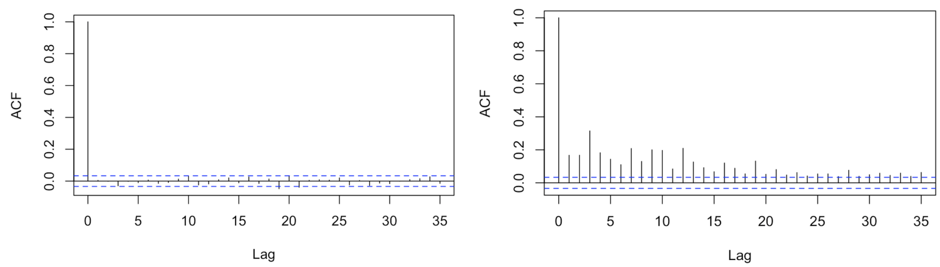
To determine the stationarity of the four stock indices, we plot the graph of daily stock prices and daily returns as well as its ACF graphs. From the GSPC Graph Set in the appendix 8, we can see that daily stock prices for GSPC are non-stationary since it does not show the mean reversal property, and the autocorrelation is high among different lags. However, the daily returns of GSPC do seem stationary as the mean of it is around 0 and the variance seems to be constant across time. Similar patterns are shown for all four stocks as well.

To further validate our initial observation from graphs, we perform the ADF test and the KPSS test on daily stock prices and daily returns for all four stocks. The table in appendix 7 in the appendix shows the p-value for each test. In the table, we can see that the p-values for the KPSS test using daily prices are all less than 0.05 while the p-value for the ADF test all higher than 0.05. On the other hand, the p-value for the KPSS test using daily returns are all higher than 0.05, while the p-value for the ADF test are all lower than 0.05. This further indicates that daily prices are non-stationary while daily returns are stationary.

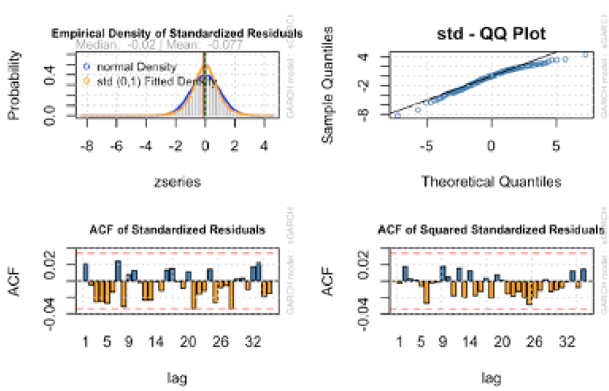
To model the data, we first apply the auto.arima() function in R on the daily prices of all four stocks separately, selecting the models of best fit using BIC. The best fit models are as follows:

-GSPC: ARIMA(2,1,0) -HSI: ARIMA(0,1,0)
-STOXX: ARIMA(0,1,0) -N225: ARIMA(0,1,0)

We can see that all stocks fit well with d = 1, which aligns with our analysis above. Except for GSPC where its differences are autocorrelated with each other, the others can be well modeled with ARIMA(0,1,0). However, by looking at the ACF graph of residuals of the fit (below is the ACF graphs for STOXX), we can see that some of the autocorrelations are outside the test bounds. Box-Pierce tests with lag equal to 25 are carried out and yet the outcomes indicate that the residuals of the ARIMA models are autocorrelated, indicating that they are not white noise. This shows that simple ARIMA models alone cannot model the indices well enough. By looking at the ACF graph of squared residuals, we can see that it is autocorrelated which indicates conditional heteroscedasticity between residuals.



To account for the heteroscedasticity, we then use the ARMA(1,0) + GARCH(1,1) model to model the daily returns, with an assumption that the residuals follow t-distribution. Graphs on the left describe the residuals of this model for GSPC. From the QQ Plot, we can see that most data points lie on the straight line, indicating that residuals do approximately follow the t distribution. From the ACF of standardized residuals and squared standardized residuals, we can see that under this model, the residuals are no longer autocorrelated. This is also supported by the weighted Box-Pierce test, with p-values greater than 0.05, there is no significant evidence to reject the null-hypothesis and we conclude that the residuals are not autocorrelated.

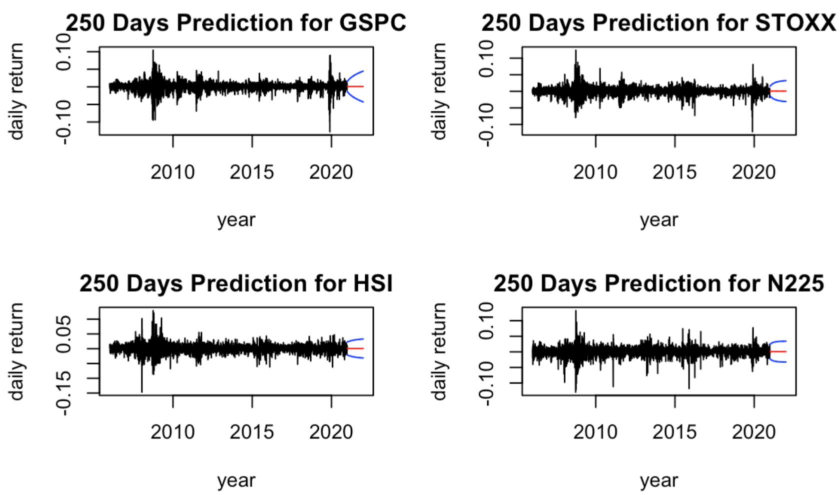


$$\text{ARMA(1,0) + GARCH(1,1) model: } Y = \mu + \phi(Y_{t-1} - \mu) + a_t$$

$$\text{where } a_t = \sigma_t \epsilon_t \text{ and } \sigma_t = \sqrt{\omega + \alpha a_{t-1}^2 + \beta \sigma_{t-1}^2}$$

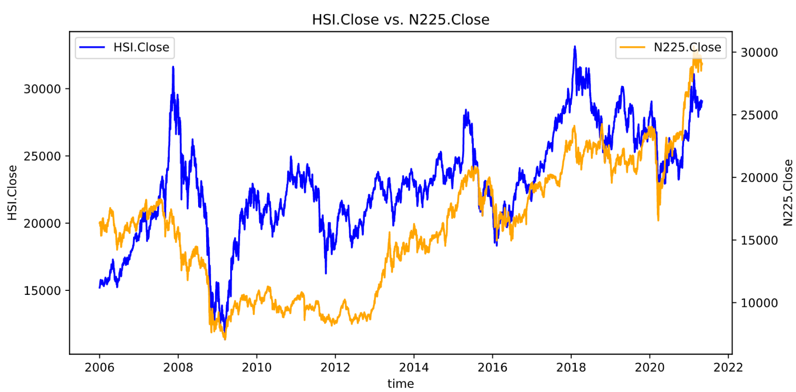
Looking at the parameters of the fitted models, all $\boldsymbol{\mu}$s estimated by the model are very close to zero. AR1 estimates are negative in all four models, indicating a trend of short-term reversal. The βestimates are around 0.9, with α estimates also being positive at around 0.1, this shows a high correlation between σ_t and σ_{t-1}. The graph below displays the predictions of future 250 days returns fitted with ARMA(1,0) + GARCH(1,1) model with a 95% confidence interval. From the graph, we can see that all future predictions are very close to zero. HSI index has the narrowest confidence interval, where the model expects the HSI index to be the most stable in the following 250 days. In the short run, GSPC and

STOXX indexes tend to be not as volatile as the N225 index. But as time passes, the uncertainty of GSPC increases dramatically.

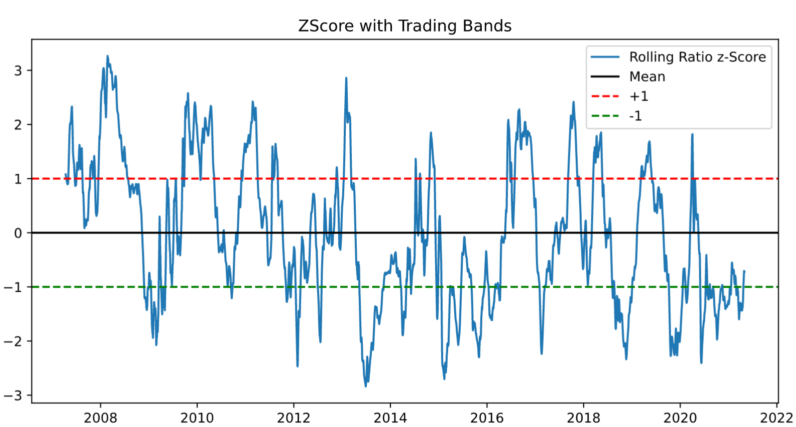


Cointegration and Pairs Trading

In statistical arbitrage, finding mean-reverting series is an important method in trading using technical analysis. Cointegration is a useful tool in creating mean-reverting series by a combination of two price series. With a stationary time series, a pairs trading strategy can be created such that when both securities' price series deviate much from its mean, one can long the overpriced security and short the underpriced security with dollar-neutral and market-neutral properties [3].

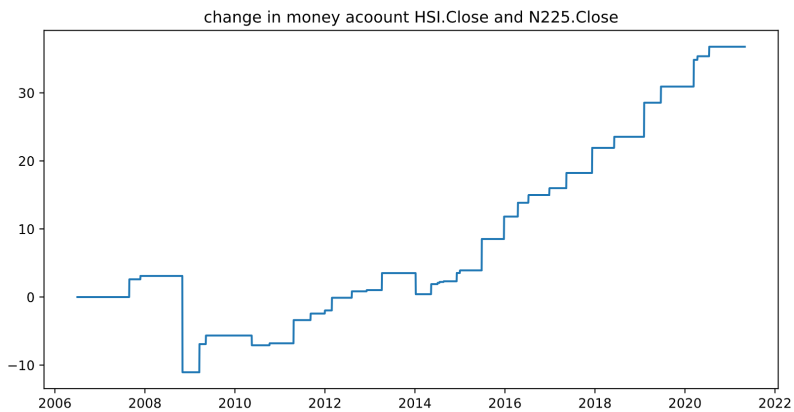


Using the augmented Engle-Granger two-step cointegration test, where the null hypothesis is that there is no cointegration, and the p-value for HSI and N225 is 0.0388 (details in Appendix 9), showing signs of cointegration. The graph above shows HSI and N225's historical price.



Following the result, we can design a pairs trading algorithm [4] that can exploit the cointegration property of these two assets. In particular, we use the 180-day moving average and volatility as the reference to the ratio between

HSI and N225, and then calculate the Z-Score. The graph above shows the evolution of the Z-Score across time, and when the Z-Score crosses either red and green dashed line from the middle, it opens a trading position. By central limit theorem and empirical analysis, the Z-Score follows a standard normal distribution. We then use the 1-standard-deviation rule, which means that we trigger a money neutral trade when Z-Score is above 1 standard deviation. As the two series are cointegrated, combining the two series should show mean-reverting property. Thus, the strategy would exploit opportunities when the pair deviates from its mean by much (1 standard deviation), and close out the position when Z-Score falls back to 0.5 standard deviation (mean reversion). Backtested with such a strategy from 2006, we can observe from the graph below, if we only allow $1 of long-short position every day, this pairs trading strategy could generate more than $30 profit. Since it is a dollar neutral strategy, there is no need for initial capital investment.



However, we can also see there is huge downside risk, if there is no loss limiting strategy. In addition, although two series might be cointegrated in the long run, it might not be cointegrated in the short run with hits such as the financial crisis, changes in fundamentals of a stock index, etc. As HSI and N225 are the two most influential indices in the Asian market, it is reasonable to infer that the two series are cointegrated. Thus, in the long run, the pairs trading strategy would be profitable between these two assets.

Conclusion

With the accelerated globalization process, booming international trades and the reduced barrier of global capital flow, the global equity market shows stronger connections. As it is observed in the multivariate and copula analysis, the US and the European markets show strong correlation and interdependence, as well as the Hong Kong and the Japanese markets. The level of interdependence could be explained by the interconnection of capital and international trades, and maybe the overlapping trading hours among different stock exchanges. From the time series analysis, we derive that the volatility of stock indices are strongly correlated with previous day's volatility, which suggests that in modeling price movements, it is not sufficient to simply model it as a geometric Brownian Motion with the independent increment assumption. In reality, the volatility could change across time with huge influence by previous volatilities. Comparing simple ARIMA model to GARCH model, the time-varying GARCH model fits financial volatility well significantly. In analyzing the cointegration property and its application in pairs trading, we observe how time series can be applied in quantitative trading. However, as statistical models are purely theoretical, it only infers that profit can be realized in the long term. In the short term, there are still huge downside risks if there is no rigorous loss limiting strategy in place.

References:

1. Cole, S. R., Chu, H., & Greenland, S. (2014). Maximum likelihood, profile likelihood, and penalized likelihood: a primer. *American journal of epidemiology*, *179*(2), 252–260. https://doi.org/10.1093/aje/kwt245

2. Andrew J. Patton. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, *vol. 110*, Pages 4-18. https://doi.org/10.1016/j.jmva.2012.02.021

3. Herlemont, D. H. (2004). *Pairs Trading, Convergence Trading, Cointegration*. http://docs.finance.free.fr/DOCS/Yats/cointegration-en%5B1%5D.pdf

4. *Cointegration and Pairs Trading*. (2020, July 12). Zhihu. https://zhuanlan.zhihu.com/p/158871695

Appendix:

1. Shapiro-Wilks normality test results

Indices	Test statistic	p-value
GSPC	0.862	< 2.2e-16
STOXX	0.8997	< 2.2e-16
HSI	0.919	< 2.2e-16
N225	0.9199	< 2.2e-16

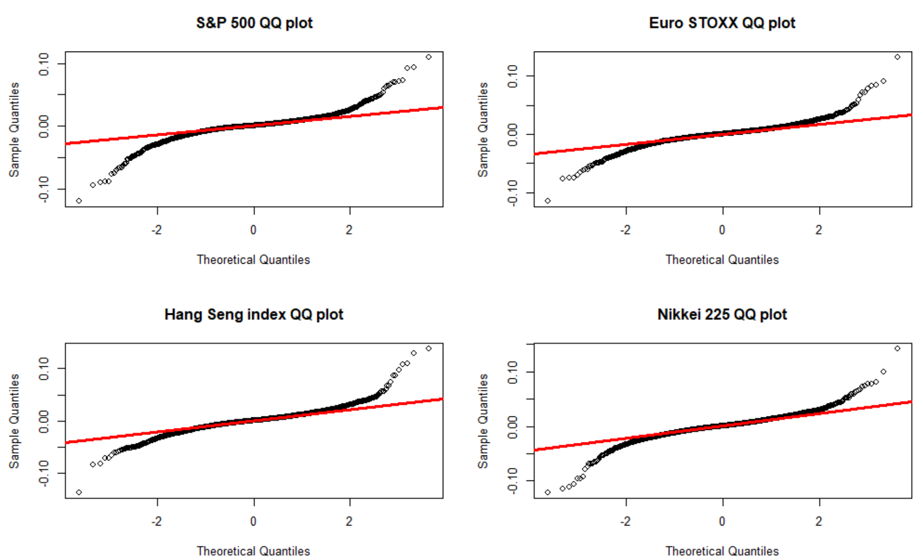
2. Kolmogorov–Smirnov t distribution test results

Indices	Test statistic	p-value
GSPC	0.0205	0.113
STOXX	0.0157	0.369
HSI	0.0172	0.267
N225	0.0147	0.451

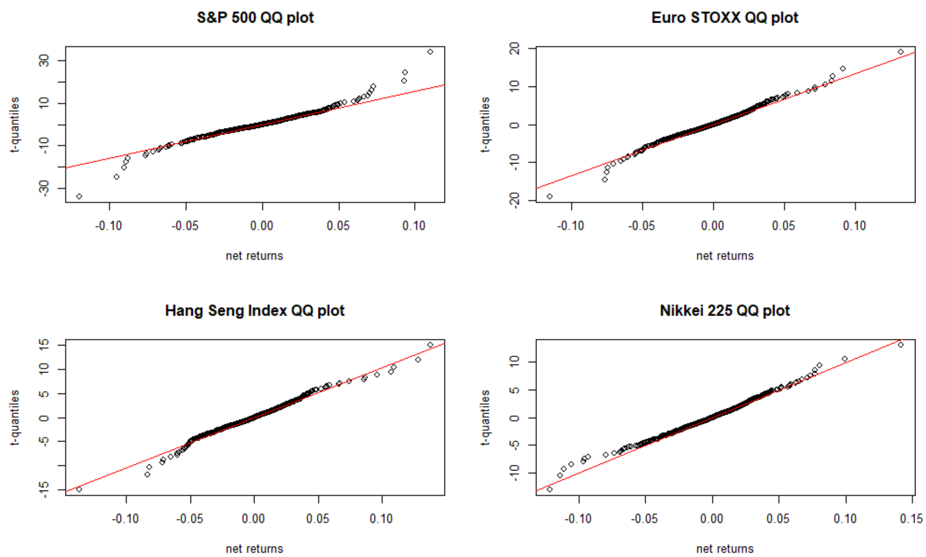
3. Estimated copula correlation for indices' net returns in meta-t distribution

Indices	Estimated correlation	Standard error
GSPC & STOXX	0.645	0.0116
GSPC & HSI	0.254	0.0182
GSPC & N225	0.183	0.0187
STOXX & HSI	0.439	0.0158
STOXX & N225	0.364	0.0168
HSI & N225	0.559	0.0134

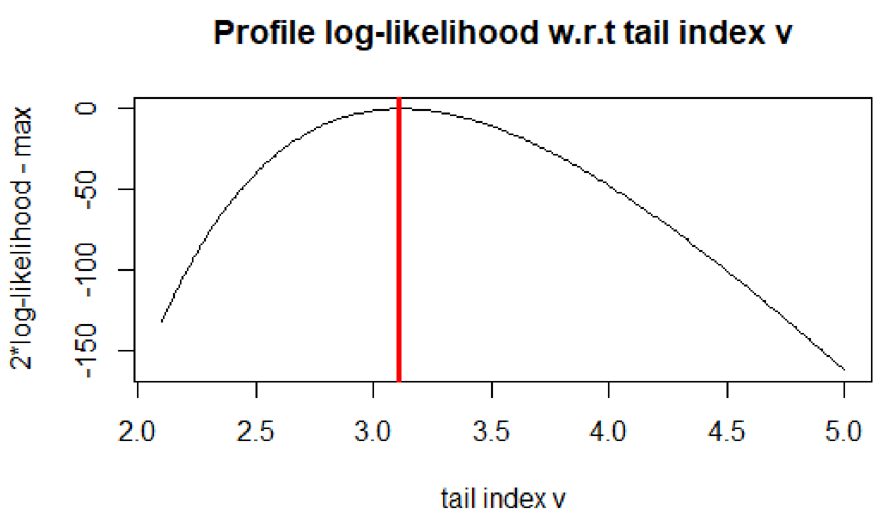
4. Normal QQ plot



5. t distribution QQ plot



6. Profile log-likelihood plot

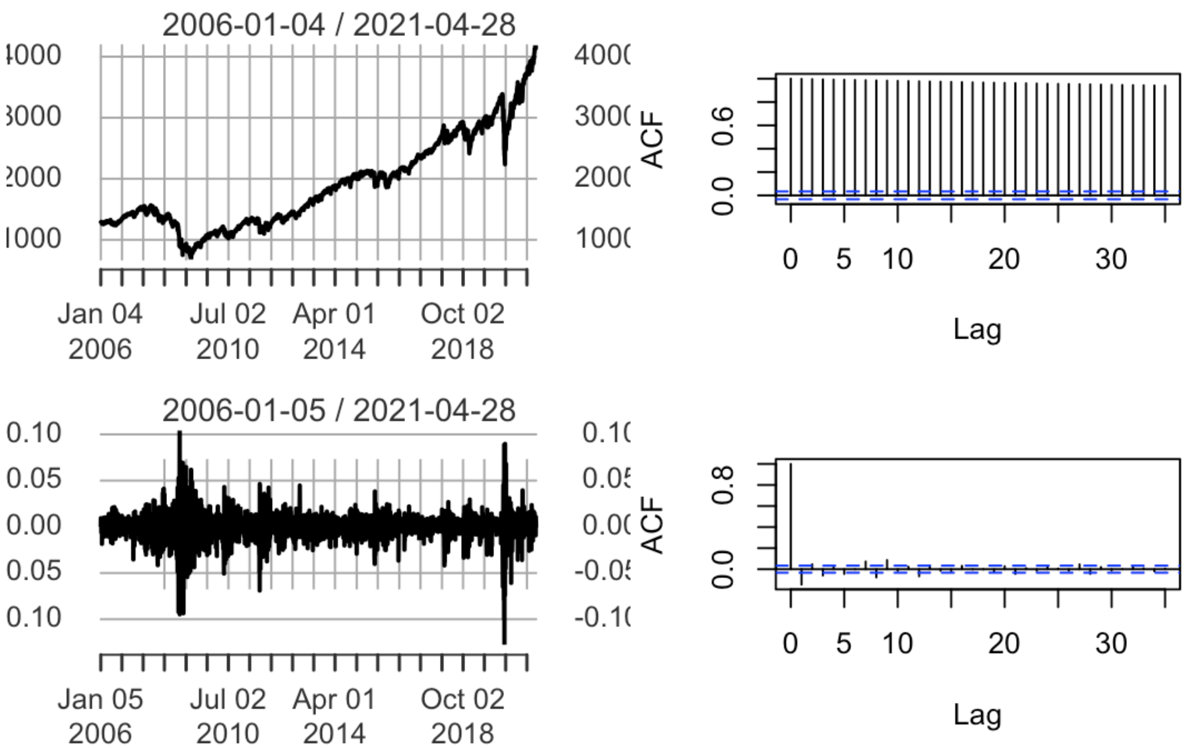


7. ADF-KPSS Test Result Table

	ADF p-value	KPSS p-value
S&P500 Price	0.9518	0.01
S&P500 Return	0.01	0.1

STOXX Price	0.4966	0.01
STOXX Return	0.01	0.1
Hang Seng Price	0.05597	0.01
Hang Seng Return	0.01	0.1
Nikkei225 Price	0.6636	0.01
Nikkei225 Return	0.01	0.1

8. GSPC Graph Set



9. Cointegration Test of Pairs

Pairs	EG Test p-value	Correlation
GSPC, STOXX	0.9615	0.7817
GSPC, HSI	0.8003	0.7461
GSPC, N225	0.5021	0.9137
STOXX, HSI	0.3996	0.7072

STOXX, N225	0.1733	0.9138
HSI, N225	0.0388	0.6915