Research Proposal

Application of Nonparametric Kernel Density Estimation in Colombo stock market (S&P SL 20)

Supervisor: Dr. M.K. Abeyratne - GROUP NO: 02

(G.H.V.C.Darshani – SC/2020/11781, Y.T.L.Peiries – SC/2020/11799, R.A.D.C.Didulan – SC/2020/11815, G.Jajeevan – SC/2020/11833)

Department of Mathematics, University of Ruhuna Matara
November 30, 2023

1. Introduction

When considering Sri Lanka's financial environment, Colombo Stock Market (CSM) is doing a major role as responsible institution of the country, which investors can trade shares of listed companies. The number of variables, including exchange rate fluctuations, economic conditions and political developments affect the CSM's behavior and performance. Investors, regulators, and policymakers can evaluate the stability and efficiency of the market as well as the risk and return of their investments by having a thorough understanding of the distribution as well as the fluctuations of the CSM returns. Using advanced statistical methods becomes necessary considering this in order to improve the reliability of predictions and risk assessments.

This research proposal evaluates the use of nonparametric kernel density estimation in an effort to fill in the gaps in the techniques currently used in the study of the CSM. Conventional approaches frequently include assumptions about certain distributions and parametric structures, which may cause them to ignore the variety of non-linear features that are present in stock movements. A more detailed comprehension of the basic trends and distributions is made possible by nonparametric techniques, especially KDE, which offer a flexible and data-driven substitute.

Traditional parametric approaches are still used extensively in financial market analysis, although it is still unclear how well they capture the real characteristics of the Colombo Stock Market. Assuming distributional forms and parametric structures could have drawbacks that make it difficult to accurately describe the dynamics of the underlying data. As such, it is imperative to investigate other approaches that are more suited to the special features of the Colombo Stock Market.

2. Literature Review

Classic parametric approaches provide solid basis for financial market analysis, But the problem of using parametric approaches is the underlying assumptions have sparked interest in nonparametric alternatives. When considering Kernel Density Estimation, the importance of this work, which attempts to provide fresh perspectives into the examination of stock prices in this financial context, is highlighted by the lack of research using these methodologies to the S&P SL20.

The work by (Wang, Y., & Wang, J. (2011).) reflects the application of nonparametric KDE in the Hong Kong stock market, laying a foundation for similar applications in other financial environments. The study emphasizes the flexibility and adaptability of KDE, enabling a detailed comprehension of stock price distributions without assuming rigid parametric structures.

Moreover, (Wijayasiri, M.P.A., & Abeyratne, M.K. (2015).) Trading in financial services are at the core of Sri Lankan market in Colombo Stock Exchange. This is measured using two indicator, the ASPI and the S&PSL-20. In as much as it is a serious concern, the availability of data about the distribution density of stock index fluctuations are nowadays pivotal aspect in financial markets. Distribution of Returns in a relation to Aspi Index and S & P SL 20 as indexes in a Colombo Stocks Markets.

This research endeavors to extend the principles applied in the Hong Kong stock market to the unique landscape of the CSM, aiming to enhance decision-making processes and refine financial evaluations.

3. Methodology

3.1 Research approach

In here the study we use Kernal density estimation for the given Colombo stock market data set applying nonparametric kernel methods in the context of the Colombo Stock Market involves leveraging these statistical techniques to analyze and model financial data.

3.2 Research design

In here the data set (Colombo stock market) that has more changing price. To predict a valuable Kernal density estimator we want get their return so respect to the last one year to make a best model we found the return of the stock price.

$$R_{(t+1)} = \log(\, p_{(t+1)}\,) + \log\,(\, P_{(t)}\,)$$

$R_{(t+1)}$ = Rate of return

$P_{(t+1)}$ = Current trading price (Closing price)

$p_{(t)}$ = Previous trading price (Closed price)

3.3 Research Question

This study proposal asks the following questions to address the previously described problems:

- In what ways might nonparametric kernel density estimation improve the Colombo Stock Market stock price analysis?
- What special patterns of the data from the Colombo Stock Market make nonparametric methods potentially useful?
- What is the difference between the outcomes of classic parametric approaches and nonparametric kernel density estimation?

3.4 Research Aim

This research aims to investigate the use of non-parametric KDE as a new analytical method in reference to the S&P SL20. By means of a comprehensive investigation of non-parametric techniques, the research aims to augment our comprehension of the fundamental patterns in stock prices, investigate the degree to which KDE can be tailored to the particularities of the S&P SL20 index prices, and combine the outcomes with non-parametric approach.

Test for normality

For this research model we expect to use the Jarque-Bera Test to check the normality.

H_0: The data is normally distributed.

H_1 : The data does not come from a normal distribution.

According to this we will check the normality in future research. After below we will be going to apply Kernal density estimator for Colombo stock market data.

Nonparametric kernel density estimation

Kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable. The basic idea behind KDE is to place a kernel (a smooth, symmetric function) at each data point and then sum up these kernels to obtain a smooth estimate of the underlying distribution.

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

n is the number of data points,

h is the bandwidth (a smoothing parameter that determines the width of the kernel),

x_i are the data points,

K(u) is the kernel function, which is a symmetric, non-negative function that integrates.

4. Kernel Density Estimation Process

Kernel Density Estimation (KDE) process is a non-parametric way to estimate the Probability Density Function (PDF) of a random variable. Let’s move to the process of the KDE.

1. *Selecting the Kernel function*

According to the smoothness of estimated density function, we can choose appropriate Kernel function among Gaussian kernel, Epanechnikov kernel, Biweight kernel, rectangular kernel, Triangular kernel etc. We choose the Gaussian kernel according to the density function we obtained. Here is the Gaussian kernel function K(t),

$$K(t) = \frac{1}{\sqrt{2\pi}} e^{-(\frac{1}{2})t^2}$$

2. <u>Choosing the optimal bandwidth</u>

   The smoothness of the density and the width of the kernel are controlled by bandwidth parameter. So it's important to choose the most appropriate bandwidth. We can use "Cross Validation method" to find the optimal bandwidth. Here are some notations regarding the Cross Validation method.

$$\begin{aligned} MISE(\hat{f}) &= \int (\hat{f}(x) - f(x))^2 dx \\ &= \int \hat{f}(x)^2 dx - 2\int \hat{f}(x) f(x) dx + \int f(x)^2 dx \end{aligned}$$

$$\widehat{MCV}(\hat{f}) = \frac{1}{n}\sum_{i=1}^{n} \int \hat{f}_i(x)^2 dx - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{-i}(x_i) \ ,$$

We can choose optimal bandwidth which minimizes the value of $\widehat{MCV}(\hat{f})$.

3. <u>Placing Kernels at data points</u>

For each data point we have, we need to place kernel function with the chosen bandwidth. The kernel function is centered at the data point.

4. <u>Sum up Kernel Contributions</u>

   To create the overall estimated density function, we must sum up the contributions of all the Kernel functions.

5. <u>Normalizing</u>

By dividing each point by the total number of data points and volume, we can normalize the resulting estimate to obtain a proper probability density function.

6. Plotting the Kernel Density Estimation

We can visualize the estimated density function by a plot. We can obtain 1one dimension for our univariate data set. The formula for the KDE at a point 'x' is given by,

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right)$$

For our non-parametric model, it can be derived as,

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi}nh}\sum_{i=1}^{n} \exp\left[-\frac{1}{2}\left(\frac{x - x_i}{h}\right)^2\right]$$

Where,

$\hat{f}$(x) = Estimated density at point 'x'

n = number of data points.

h = bandwidth.

k = Kernel function.

5. Data

5.1 Preparation for Analysis

Removing unwanted data indices and rows from dataset was an essential data cleaning step with several implications, such as data quality and accuracy, improving data understanding and simplifying and pretending cultural and term issues etc.

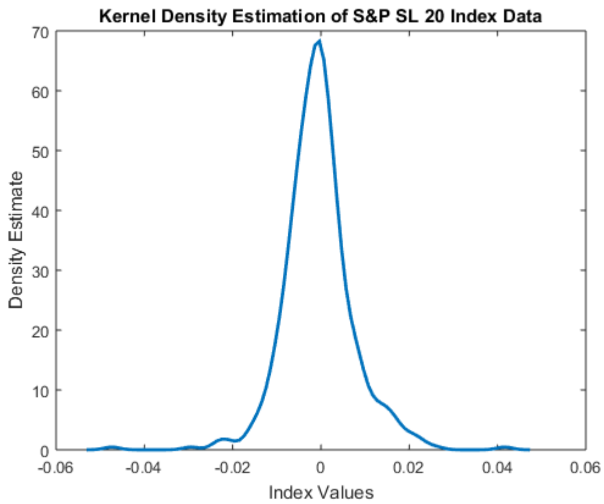
5.1.2 Dataset

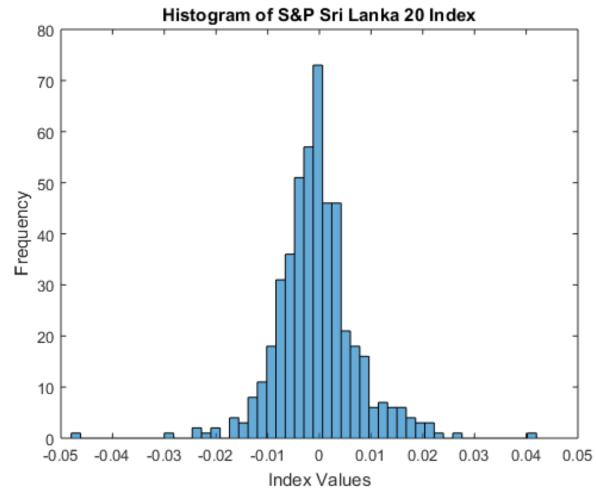
Data was collected from Colombo stock market…

5.1.3 Data dictionary

Variables	Variable type	Measurement Scale
Date	Temporal	Ordinal
Close price	Numerical	Ratio
Rate of Return	Numerical	Ratio

Kernel Density Estimation of S&P SL 20 Data





The data appears to be relatively symmetrically distributed around a mean close to zero, showing moderate variability. The JB value suggests that the data might conform reasonably well to a normal distribution.

References

- Wang, Y., & Wang, J. (2011). Application of Nonparametric Kernel Density Estimation in Hong Kong Stock Market. *Applied Mechanics and Materials*, Vols 55-57, pp. 209-214. Trans Tech Publications, Switzerland. DOI: 10.4028/www.scientific.net/AMM.55-57.209

- Wijayasiri, M.P.A., & Abeyratne, M.K. (2015). An Application of Nonparametric Kernel Density Estimators in Colombo Stock Market Indices. *Proceedings of 2nd Ruhuna International Science & Technology Conference*, University of Ruhuna, Matara, Sri Lanka, January 22-23, 2015.