European Journal of Business and Management
ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online)
Vol.14, No.12, 2022

Designing a Forecasting Model for Stock Market under Non- Normality: Case Study of Chinese Stock Exchange

Mona Ebrahimi[1] Alireza Movassagh[2*] Mahshid Ebrahimi[3]
1. Department of Industrial Engineering, Tsinghua University, Shunde Building, Beijing 100084, China
2. Department of Industrial Engineering, Iran University of Science and Technology, PO Box 16846-13114, University Ave, Tehran, I.R of Iran
3. Researcher
* E-mail of the corresponding author: alireza.movassagh@gmail.com

Abstract
In recent decades, the stock market and its growth have attracted investors. One of the biggest challenges investors have always faced is the high volatility of stock prices. There are several studies on the prediction of stock prices, but most of them are in consistent with the fact that data must be fitted in the normal distribution. In other words, conventional prediction of stock prices beyond the normal distribution is limited, leaving a gap due to the non-normality of data. For this reason, this research has focused on data behavior and its effect on the forecasting accuracy. Therefore, we reviewed 72 recently published papers. This review identified three forecasting models and two bootstrapping methods which have been combined into a new alternative model to develop and present an uncertainty model with deeper insight into the data. In this line, we developed a five-stage model to analyze and select the best combination for prediction. This model was coded with Python and 10 selected stocks from Chinese stock exchanges were used as input to ensure the robustness of the model. The model has 20 outputs per share. Holt's winters' models accurately reproduced the trend. And "BS1-Holt's additive damped trend", "BS1- HW" were the most accurate models. In conclusion, investors could benefit from this data-based uncertainty model to improve their forecasts and profits.
Keywords: Stock Price prediction, Bootstrap, Combination model, Data Science, Non-normality, Uncertainty model
DOI: 10.7176/EJBM/14-12-03
Publication date: June 30th 2022

1. Introduction

Stock market performance is of paramount importance in planning of business activities and individual investment worldwide. Nevertheless, investors should tackle many challenges in order to take advantage of investment. Not only is the prediction of future output critical for stock market but also it is vital for every industry involved with control and risk management (Brockwell PJ, Davis RA., 2016). A lot of research has been investigating the investors' problems, for instance, Widodo Budiharto, 2021 found the trend forecasting supported by a prediction software necessary for decision making. There are also some internal and external variables that could lead to movement of the price in stock market. General technical analysts have developed many indices and sequential analytical methods which may reflect the trends in the movements of the stock price (Zhang et al., 2018). Budiharto, 2021 proposed a method using Long, Short-Term Memory (LSTM) as a predictor in short term with accuracy 94.57% with high epoch in the training phase rather than using three years training data. Some papers have also studied the influence of either external or policy factors such as Xiong et al., 2018 found Chinese stock market more sensitive to internal crisis and crashes rather than global crisis for instance in 2015, the dynamic conditional correlation reached the lowest value while it had a sharp soar in the time of global financial crisis in 2008. In terms of policies, the limit- up and down pricing system in China has been %10 from 16 December 1996 and it would have a significant role in forecasting. Nonetheless, investors are still struggling with some challenges.

In this research, three challenges have been studied. Firstly, investors are seeking models which are able to face uncertainty problems and could make both daily and trend forecasting. Uncertainty has become a topic of importance to almost anyone who involved with business and has been one of the most challenging areas of research in data science and forecasting. Financial markets throughout the world have faced higher volatility in recent periods and investors are concerned about their investment in these volatile and unstable markets. In this regard, an uncertainty Model has been proposed with estimations of both daily price & trend to strengthen the investors judgement with the support of more than one combination model.

Secondly, using the historical observations is argued to be playing a significant role in stock price forecast and investors are interested in models utilizing available data because they prefer to ignore external effectors such as global financial crisis or etc. Therefore, the targeted data in this research is daily stock price from China's stock exchange and no external factors such as oil price and policies are considered.The third problem is that the current models' forecasting could be inaccurate due to ignoring non-normality of data. Investors have been seeking models utilizing available data with methods to avoid normality while the published models have ignored non-normality of data. To investigate taking the non-normality of data into account two bootstrap methods and the adjustment strategies have been proposed. Bootstrap (BS) (Efron, 1979; Efron & Tibshirani, 1993) is a data driven statistical tool based on re-sampling with replacement that could be used to quantify the uncertainty associated with a given estimator or statistical learning method. In non-normal cases, it could specifically be used for data preparation along with controlling the uncertainty. BS estimates population statistics by the use of sample statistics available. It also generates multiple samples from existing samples to more accurately estimate target statistics by reducing errors. BS could be parametric or non-parametric and calculated based on residuals or could involve Bootstrap pairs.

The aim of this study is to examine the use of bootstrap and adjustments strategies along with combined forecasting models for studying of uncertainty with proposing an uncertainty model. The study detects the best fit exponential family models on the bootstrapped data. Hence, the objectives of this study are developing a data review panel, developing two adjustment strategies, developing a combining homogenous model, fitting forecasting models on the bootstrapped data and finally computing the uncertainty of each model with the error measurements MAE, MSE and RMSE by running from each bootstrap method.

We intend to provide a robust and efficient uncertainty model based on conventional forecasting models, bootstrap and data science. Employing bootstrap and integrating numerous models has become a universal technique to enhance forecasting accuracy and the literature on this matter has extended noticeably.

2. Methodology

This study employs empirical research in uncertain and non-normal situation, and it is much more focused on the collection and analysis of data with purpose of improving the accuracy of ongoing predictions. There were three ideas to come up with at the beginning, however, the third one was selected due to coding limitations. Although the python code has the capacity to forecast for one month, in this study the forecasting would be for the next four days to make the analysis possible.

I. Idea 1

The first idea was a four-dimensional model in which each forecasting parameter would be one aspect of the model. This proposed model (figure 1) is a hybrid model which parallelly links to high volatility parameter and short-term parameter. The pool models out of which the aspects have been selected were MLP, GARCH, ARIMA, Holt-Winters, Smoothing or SVM, and interval forecasting. Therefore, one aspect of the model would be ARIMA which is a classic linear model based on historical data and MLP which is an intelligent nonlinear forecasting model as a neural network and one of the deep learning techniques for pattern recognition and movement direction of data would be another aspect. The third aspect would be Exponential Smoothing which is for short-term forecasting and the last aspect would be bootstrap to fulfill the high level of volatility.

Figure 1. The First Idea

II. Idea 2

The second idea was treating the uncertainty as robust optimization (RO) model. In decision making, uncertain parameters can affect our objective function and constraints due to unknown future developments, so we can use RO model as follows: (i) assuming the stock market return as an objective function and uncertainty as an effector on the objective function. (ii) Investigating the behavior of objective function by variation of the uncertainty.

III. Idea 3

The above-mentioned ideas are time-consuming and complicated to code. The third idea was a bootstrap-based uncertainty model with adjustment, which is a simplified state of “idea 1”. The forecasting parameters considered in this model are volatility, short term, trend and recent data. Due to the fact that the experiment would be done in a non-normal situation, non-normality and volatility are taken into consideration and consequently some volatility distributions are empirically generated on a large scale. Meanwhile, a bootstrap-based model was used in order to seek a model assumption for uncertainty while avoiding normality. At last, the result was compared with actual return distributions.

2.1 Model building strategy for uncertainty model

The model building strategies developed in this study have used the techniques by Kim & Kim (2018), Fresoli et al. (2015) and Flores-Agreda & Cantoni (2018). As a result of combination of these techniques, a five-step procedure as below is proposed to analyze data and come up with the best combination models fitted on data. Figures 2, 3 and 4 illustrate the procedure in terms of data review, schematic and mathematic form of our Uncertainty Model.

I. Stage 1: Data Review Panel

Since data and data science play significant roles in the study, a review panel has been created to accurately data and data behavior. Figure 2 illustrates procedures in terms of data in which ten shares were picked as the dataset with daily prices during the period 2012 to 2019/03/29. Descriptive statistics, normality tests and stationary check on data could also help identify and analyze the nature and behavior of data. For instance, the outliers are important to be identified in order to give some hint for choosing a proper forecasting method.

II. Stage 2: Bootstrap Methods

Given the complex expressions of the estimators of the uncertainty for point predictors and their corrections, a typical alternative way of dealing with this problem is via resampling methods (Flores and Cantoni, 2018). The behavior of bootstrapped data is considerably different from that of the volatility of the original data.

III. Stage 3: Forecasting Models Identification

In the first step of identification, the data was plotted. If it was non-stationary, the trend would be identified and removed. If the series was seasonal, it would require to be specifically treated. After that, ACF and PACF were computed for better insight into data.

To consider whole data while weighing the data points differently, it may be sensible to assign larger weights to the more recent observations than the distant past ones. In this case, SES forecasting model would be used. Forecasts were calculated using weighted averages where the weights decrease exponentially as the smallest weights are associated with the oldest observations. The value of $0 \leq \alpha \leq 1$ is the smoothing parameter. To map the trend accurately without any assumptions, we would use Holt's smoothing models. Any data set which follows a trend can easily use Holt's linear trend while Holt's Exponential trend extends SES to allow forecasting of both data and trend. Additive damped trend is usually used when the trend increases or decreases linearly. For more complicated cases, Holt's winters models could be used in which:

alpha = Value of smoothing parameter for the base level.
beta = Value of smoothing parameter for the trend.
gamma = Value of smoothing parameter for the seasonal component.

IV. Stage 4: Adjustment Strategies

The estimations could also be corrected and steered with the adjustment strategies according to the properties and behavior of data. The adjustment strategies work as a valve which will be opened in case of necessity.

V. Stage 5: Uncertainty Measurement

There are three forecasting models and two bootstrapping methods multiplying we have at least six different estimations. At the end, the forecasting will be compared with the actual and the error comes out.

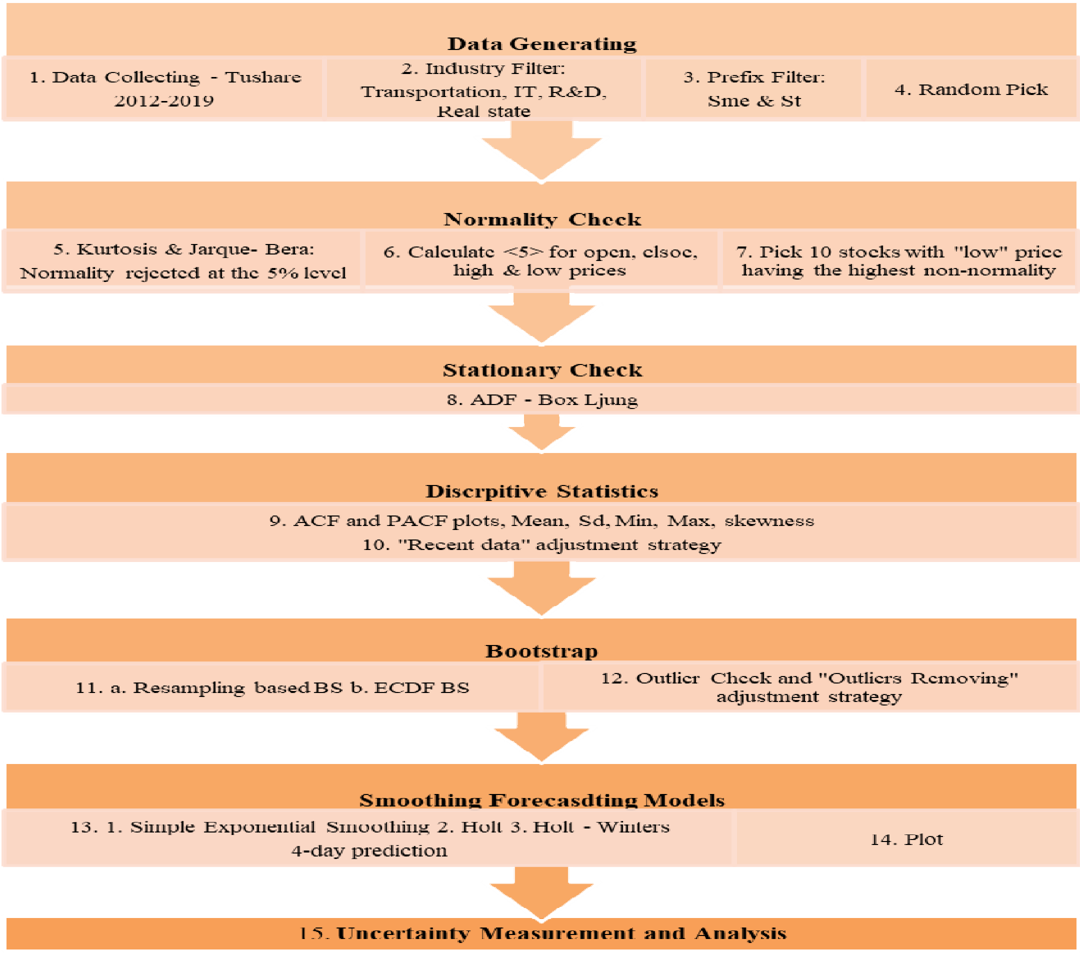

Figure 2. Look into the model in terms of data - Data Review Panel

European Journal of Business and Management
ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online)
Vol.14, No.12, 2022

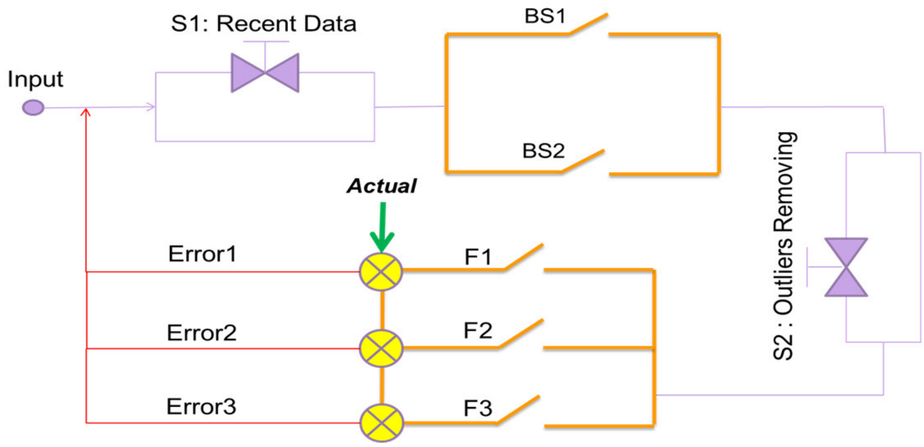

Figure 3. The uncertainty Model

$$cS_1 + BS_i + dS_2 + F_j \pm \varepsilon_{ij} = Actual$$

Adjustment Strategies

S_1= Strategy number one named as "Focus on recent data" strategy with finding the critical points.
Can be used for both data classifications," With outliers" & "With no outliers" and in case of sharp changes in the trend

S_2= Strategy number two named as "Outliers removing". Removing Percentile can be changed. Mostly used for classification, "With outliers" and stocks with high fluctuations

BS = Bootstrapping methods
i = 1, 2
F = Forecasting Models
j = 1, 2, 3
c, d = binary
ε = error = Uncertainty

Figure 4. Mathematic Form of the Uncertainty Model

3. Bootstrap Procedures

In high volatility, a bootstrap-based measure is used as an alternative for uncertainty estimation. Two bootstrap algorithms are used here;

3.1. Algorithm 1 (BS1)

The first Algorithm (BS1)) re-samples an array-like data and has two parameters, X as the array-like data to re-sample, n as the length parameter or the length of re-sampled array. The parameter "n" is the parameter which we can steer the uncertainty on the fact that it gives different results. The first step in BS1 is the array creation represented with "X". The second one is bootstrap resampling process which is being applied to the array-like data, "X", with the output, X-Resampled as figure below.

The length of re-sampled array equals "n". In the bootstrap re-sampling step, both shape and mean of resampled data will be checked. In terms of shape, the resampling is checked in the following two situations; Without "Re-sampling length" parameter and with "Re-sampling length" parameter. In the first situation, the re-sampled length should be length(X). Moreover, the re-sampled data will be checked to see if the "mean" of data and re-sampled data are approximately equal. The re-sampled data also will be tested if means are close for data frame with unusual index.

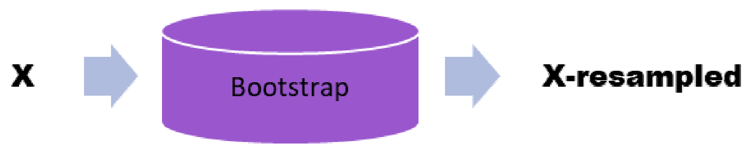

Figure 5.The Bootstrap Concept

3.2. Algorithm 2 (BS2)

The second algorithm (BS2) uses empirical cumulative distribution function (ECDF). This function, first, takes a one-dimensional array of data as input then returns the x and y values of the ECDF. Hence, BS2 is dependent on an attribute named range which is set on 50 & 100. After that, the forecast is made the results of which are different based on different ranges (Adhikari and DeNero, 2016). Following this further, the forecasting results of the two ranges are averaged illustrated with share "600206" in figures 6 - 8. The forecasting results vary depending on different ranges in BS2. Meanwhile, the average range has more accurate result than 50 or 100. For share "600206", neither the "Outliers removing" nor "Recent data" strategies were applied to data. If the "recent data" strategy needs to be applied, it would be before bootstrapping.

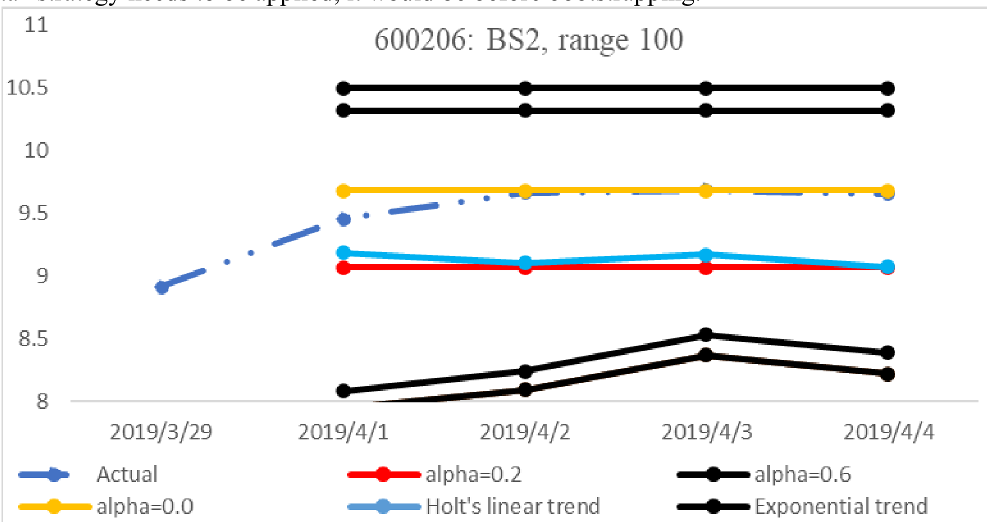

Figure 6. Forecast via BS2 in range of 100, 600206

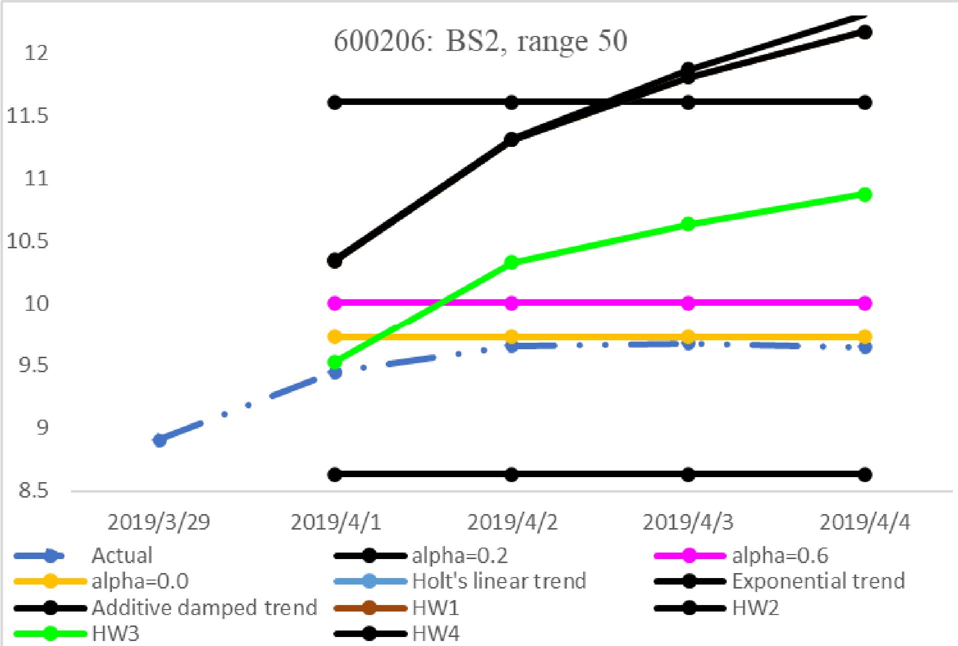

Figure 7. Forecast via BS2 in range of 50, 600206

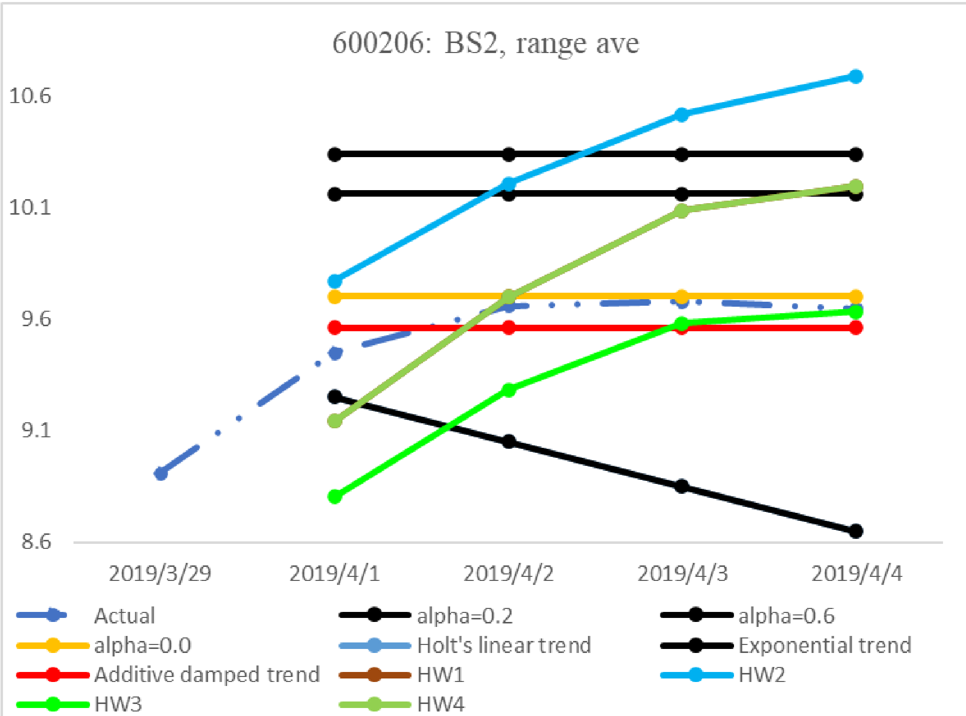

Figure.8 Forecast via BS2, Average Range, 600206

4. Forecasting Methods

Three forecasting methods were chosen based on the needed forecasting parameters which are volatility, short term investment, trend and recent data.

4.1 The Simple Exponential Smoothing

This method employs an exponential-decay weighting system. The averaging out of calculating a weighted average of the past data can be implied as Smoothing. The one-period-ahead forecast of historical data formula is given by

$$Y_t(1) = Y_{t-1}(1) + \alpha \, [Y_t - Y_{t-1}(1)] \tag{1}$$

Where $Y_t(1)$ is the smoothed value at time t and the current period's forecast $Y_{t-1}(1)$ adjusted by a proportion of the current period's forecast error (Levenbach, 2017)

4.2 Holt

Trends and Residuals are components of time series dataset which could help us to decode the data. To take into account the trend, we would use Holt method. To map the trend with no assumptions, Holt's linear trend is used. Holt's Exponential trend is extension of SES to do both trend and one-step-ahead forecasting as level equation shows a weighted average of observations and one step ahead forecast, trend equation indicates a weighted average of the estimated trend.

Forecast equation: $\hat{Y}_{t+h|t} = l_t + h\, b_t$ (2)

Level equation: $l_t = \alpha Y_t + (1-\alpha)(l_{t-1} - b_{t-1})$ (3)

Trend equation: $b_t = \beta(l_t - l_{t-1}) + (1-\beta)\, b_{t-1}$ (4)

When the trend increases or decreases linearly, additive equation is used.

4.3 Holt-winters

Holt-winters model considers seasonality and it is a model that consider both trend and seasonality to forecast the future prices. There are two methods to implement one is additive method as the seasonal variations are constant through the series and multiplicative method as variations are changing proportional to the level of the series.

Forecast equation: $F_{t+k} = L_t + k\, b_t + S_{t+k-s}$ (5)

Level equation: $L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1})$ (6)

Trend equation: $b_t = \beta(L_t - L_{t-1}) + (1-\beta)\, b_{t-1}$ (7)

Seasonal equation: $S_t = \gamma(y_t - L_t) + (1-\gamma)\, S_{t-s}$ (8)

5. Adjustment Strategies for more accuracy

Using and focusing on the whole data history would sometimes be misleading in these conditions:

1. When there is one critical point or outlier in data.
2. When there are a small number of critical points or outliers in data.
3. When there are a couple of critical points or outliers in data.
4. When both the whole trend and the recent trend should simultaneously be considered.

So, in case of necessity and based on the flowchart below, the adjustment strategies would be used to correct the forecasted price.

Challenges and issues in data-based forecasting so as to increase the accuracy are:

1. Specifying critical points and trend in data.
2. Specifying the recent data based on critical points.
3. Specifying the recent data based on both Critical points and the whole trend.

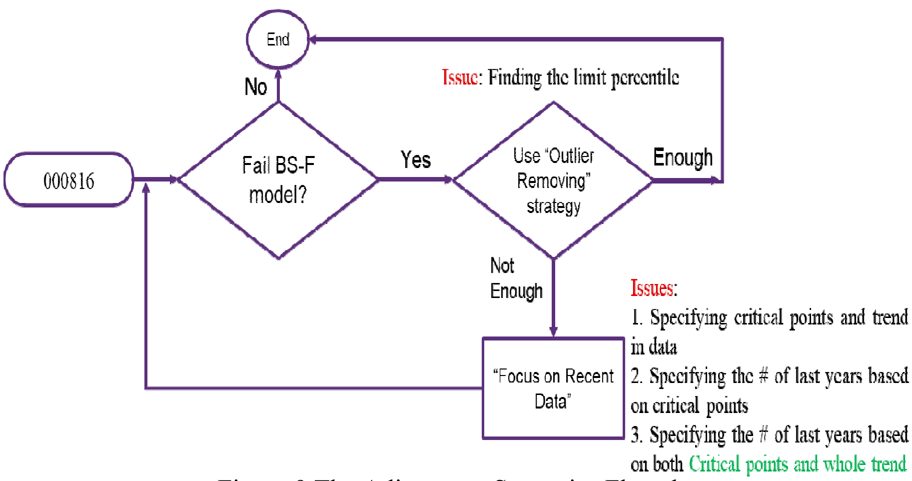

Figure.9 The Adjustment Strategies Flowchart

6. Data Review Panel

Considering the fact that data points taken over time may have an internal structure (such as autocorrelation, trend or seasonal variation, it is of great importance to identify the data nature and behavior and do the time series analysis. Stockbrokers stock dealers who need to predict the stock price usually use technical and fundamental or time series analysis in order to advise client ((Widodo Budiharto, 2021). Some of statistics methods in this study are as follows:

6.1. Stationary check with Augmented Dickey-Fuller test

In order to conduct the stationary check for the shares, Augmented Dickey-Fuller (ADF) test could be used which is one of the most widely used unit root tests and shows how strongly a time series is defined by a trend (Kim& Kim, 2018). The hypothesis test is as below:
H0: it is non-stationary and having some time dependent structure.
H1: it is stationary. It does not have time-dependent structure.

6.2. Test of Normality

To check the non-normality of each share, test of normality would be done.

6.3. Skewness and Kurtosis

The skewness describes the asymmetry of a distribution. A right tail has a positive skewness value while a left tail indicates a negative value of skewness. The skewness of a random variable (X) is:

$$S = \frac{E(X-\mu)^3}{\sigma^3} \quad (9)$$

The Kurtosis which is the fourth power of standardized data is a measure of flatness of a distribution. Distribution with higher than three Kurtosis peaks close to its mean.

$$K = \frac{E(X-\mu)^4}{\sigma^4} \quad (10)$$

6.4. Jarque-Bera (JB) test

The normality of the data is checked with Jarque- Bera test having asymptotic $\chi 2$ distribution with 2 degrees of freedom. JB incorporates both skewness and kurtosis.

$$JB = \frac{n}{6}\{S^2 + \frac{(K-3)^2}{4}\} \quad (11)$$

H_0: The data are normally distributed
H_1: The data are not normally distributed
If P-value is less than 5.99, normality will be rejected at the 5% level of significance.

6.5. Histogram and Kernel Density estimation

Histogram is a graphical representation of the frequency providing insight into both skewness and kurtosis, the behavior of tails and outliers. Although histogram is a good beginning for the location and shape of the data, its non-smooth property remains unsatisfactory. The Kernel density estimation is a nonparametric estimator of densities alleviating non-smooth problem. The Kernel estimator of "f" at the point of x is:

$$f_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\frac{(x-X_i)}{n} \quad (12)$$

Where h is the smoothing parameter known as the bandwidth

6.6. Autocorrelation and Partial Autocorrelation function

The autocorrelation function (ACF) gives an idea of the degree of dependence between the values of a time series. The visualization of the ACF or of the partial autocorrelation function (PACF) helps to identify the suitable models to explain the past observations and to do predictions. The theory shows that the PACF function of an AR (p), an autoregressive process of order p, is zero for lags greater than p.

6.7. Outlier Detection with the interquartile range

A robust method for labeling outliers is the IQR (interquartile range) method of outlier detection developed by John Tukey. A box-and-whisker plot uses the quartile (points that divide the data into four groups of equal size) to plot the shape of the data. The box represents the 1st and 3rd quartile, which are equal to the 25th and 75th percentile. The line inside the box represents the 2nd quartile, which is the median (Colin Gorrie, 2016). The interquartile range is between:
Lower Limit = 25th percentile -1.5 IQR
Upper Limit = 75th percentile +1.5 IQR

6.8. Forecasting and Accuracy Measures

Evaluating the performance of the model plays a crucial role in making a good decision. The most widely used evaluation measures are Mean Absolute Error (MAE), MSE and Root Mean Square Error (RMSE). MAE is less sensitive to large deviation than the usual square loss.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|A_i - F_i| \quad (13)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2 \quad (14)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2} \quad (15)$$

7. Implementation

We used the Python environment for implementation since it is easily adaptable and compatible to the modeling environment (Ghadipash, et al., 2017). The coding process is depicted in Figures 10. The package Tushare was used to fetch the data.

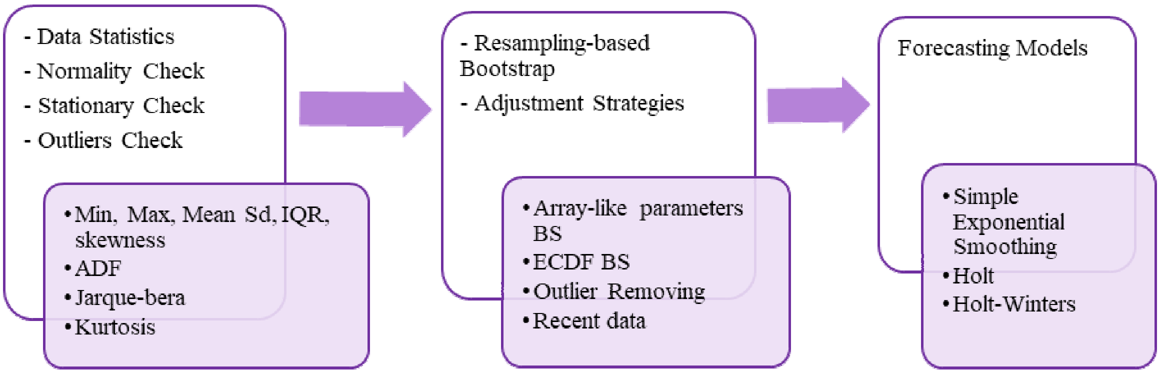

Figure.10 Python Coding Process

8. Research Findings

The data of 10 shares in Chinese stock exchanges indicate the characteristics as high volatility and occasional movements. Like Makhwiting (2014), the conclusion that daily prices of these shares have non-normality is reasonable as it is common in dataset of emerging markets. Overall results provide evidence to show platykurtic distribution and leverage effects for these shares. The high value of kurtosis for “002195” suggest that extreme price changes occurred more frequently during the sample period, 2012 to 29 March, 2019. The findings of this study also show that HW models mapped the trend accurately. And “BS1-Holt’s additive damped trend”, “BS1-HW” were the most accurate models. Moreover, the following results need to be augmented; Firstly, data is not always well behaved with normality assumption and in case of high volatility, forecasting would not be accurate in absence of non-normality. In this regard, two bootstrap methods have been used to evaluate the error distribution properties and then models are fit on the data available. Secondly, we planned to make a robust model to reinforce the judgment of investor’s reliance on more than one method. That is why, our uncertainty Model could make 20 estimations of the future price. Adjustment Strategies have brought flexibility to the forecasting model in case of asymptotic behavior of unusual or extreme observations. Third, examining different turning points in dataset to find the best turning point while observing the entire trend could improve accuracy. Fourth, availing ourselves of outliers removing strategy in the right time & with proper percentile could be efficient.

Study of forecasting and uncertainty is a practical and customer-oriented issue on account of affecting the trader’s decisions on what to buy and when to sell a share. We are not able to eliminate uncertainty but we could minimize it by gaining insight into the accuracy of the models and forecasts. This combined-multi-method model can create a forecasting panel consisting of several prediction by finding the behavior and pattern of data, utilizing bootstrap and adjustment strategies along with traditional forecasting models. It is to assure that investors’ judgment is not only dependent on one particular forecasting model.

An argument to be raised here is why bootstrapping should be coupled with forecasting models. There are two main reasons for this; 1) Some forecasting models need structured data and without giving structure to the data they do not properly work such as Holt-winters’ models. 2) The forecasting results vary depending on each bootstrap sample and it could help to decrease uncertainty.

Another argument is whether or not applying the adjusting strategies is scientifically, correct. Imbalanced datasets of stock market forced us to utilize the adjustment strategies in contrast to other studies mentioned in the literature review section. To make an accurate forecast, using the following methods are inevitable: “Bootstrap” method, “Outliers removing” and “Focusing on recent data” strategies. Flores and Contoni, 2019, Zhang, 2018, Freesoli et al., 2015, Ahmadi et al., 2017 and Shimizu & Kano, 2008 are in favor of our result. To reduce the bias and error, we could track and investigate the pattern of error and investigate the mentioned strategies and components to devise an adjustment. That is why, this combined model has the capacity of prediction in imbalanced and non-normal data. Our model shows that with a comprehensive look into data and data manipulation along with the traditional forecasting models, the quality of decision-making will rise because investors decide based on more than one factor. The results in this research strengthen this finding.

The Root Mean Square Error (RMSE), Mean Square Error (MSE) and Mean Absolute Error (MAE) are used to rank the models based on their accuracy. The negative percentages are over forecasting where we must make a negative adjustment to reach the actual price while the positive ones are under forecasting. To reduce the bias and error, we can either investigate the pattern of error or investigate the mentioned strategies and components to devise a bias adjustment. The ACF and PACF are used to find possible models.

RMSE and MSE statistics per each bootstrap method are plotted (scatter plots) for each share in figure 11. Comparison among RMSE-BS1, MSE-BS1, RMSE-BS2 and MSE-BS2 indicates that BS2-HWs and BS2-Holt's models provide the most accurate forecasts. The RMSE indicates that "BS1- Holt's additive damped trend" and "BS1-Holt's winters" are the most accurate models so that BS1 is more accurate than BS2. The MSE statistics indicate that BS1 is more accurate than BS2 so "BS1- Holt's additive damped trend", "BS1-Holt'linear trend" and "BS1-Holt's winters" are the most accurate models. Furthermore, the scatter plots' comparison in figure 11 show that BS1 is more densely populated than BS2. Besides, the uncertainty or the error value reached a peak of 4.54 for "MSE- BS1- SES with alpha as 0.2" in share "002184" and a low of 0.0 in a minimum of several combinations (higher frequency for HW models).

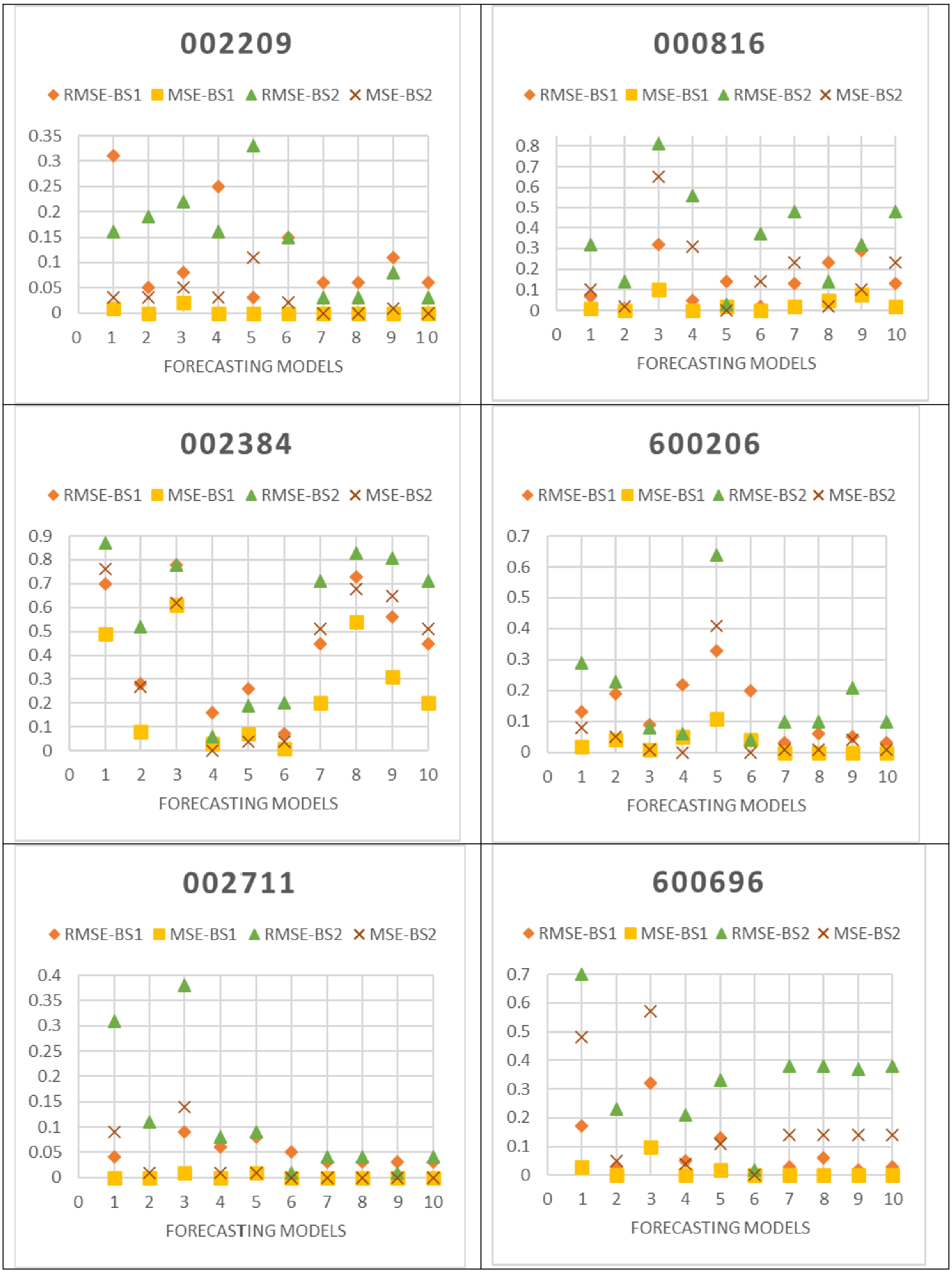

Figure.11 Forecasting accuracy comparison by RMSE & MSE

9. Research Contribution

Previous research has examined the forecasting models with normality assumption. This research has typically shown that shareholders purchase their stocks more favorably when they have a robust tool with available data to enforce their decision making. In contrast, the current study is the first attempt to study and code a combination of forecasting models, bootstrap, and adjustment strategies in Chinese Stock exchanges and forecast both trend and one-step ahead forecasting. We find that different bootstraps can affect the uncertainty as different forecasting models can. Furthermore, we showed that how finding the turning points are crucial in prediction.

10. Research Limitation

The uncertainty model has numerous advantages over forecasting models (Kim & Kim (2018), Fresoli et al. (2015) and Flores-Agreda & Cantoni (2018). Some of the advantages are as follows: (1) It has a Data Panel testing data's property and behavior; (2) The uncertainty model provides more informative data as you are able to add large amounts of data points in the analysis, and also to increase the degree of freedom, thereby increasing the efficiency of the estimates; (3) when using longitudinal data, it is possible to analyze effects that is not detectable when using a pure time series data set; (4) The uncertainty Model via bootstrap and adjustment strategies makes it possible to conduct more complex behavioral models than when using time series data; (5) The uncertainty model provides a more accurate prediction compared only relying on conventional forecasting models and normality. There are, however, some limitations when using the uncertainty model: (1) the main problem with the uncertainty model is the design and coding of the model; (2) In coding of the adjustment strategies, finding the turning points in data is not easy to code. (3) The order of the adjustment strategies is disputable.

11. Conclusion

This paper proposed a flexible model to facilitate and accurate stock price prediction. To answer the question "Whether uncertainty model with focusing on data and modern computing power is able to improve forecasting in stock market?", we have made 20 estimations of the future price with our uncertainty model.

In the first set of methodology, bootstrap and adjustment strategies regard non-normality as a nuisance and circumvent the assumption of ϵ~MVN(0,σ2IN)ϵ~MVN(0,σ2IN) with statistical theory or changing the estimator so as to relax assumptions about the form of the errors, ϵ (e.g., bootstrap) (Pek et al, 2018). King and Roberts (2014) proposed that heteroscedasticity-corrected covariance matrices are useful for detecting model misspecification. However, in this research we used bootstrap like Fresoli et al., 2015. In general, our estimations assume that all forecasted prices are valid and it implied that the model has less diversion between forecasted and observed prices.

In the second set of robust approaches, non-normality due to outliers is regarded as indicative of data imbalance. Such imbalance data is addressed by modifying or discarding extreme data points. A lot of information about the nature of the phenomenon in the study, and the characteristics of valid data is required to enable us confidently identify and justify the removing outliers. Robust approaches imply that data comes from multiple populations.

We anticipate that our review, classification, and examples could open up research opportunities for papers on extending approaches addressing uncertainty in terms of non-normality. Future research would model uncertainty to investigate when and how much non-normality becomes important with focus on data around a business cycle peak or through an important "China event" to find some definitive differences in behavior that would be of interest to investors to change their strategies. For the forecasting models, the reproducible models such as VARMAX family would be recommended. Uncertainty distribution such as Tukey g & h which is location-scale invariant would be recommended. MAPE & MDAPE are also suggested to be used instead of current performance measurements. Future research should look at uncertainty modeling examining the sensitivity of data to the lower and upper limits in the outlier's percentile. Finding the turning points, the number of re-sampling in bootstrap and aggregating other forecasting models need to be investigated. It would be recommended to conduct this experiment on more shares and worldwide.

12. Acknowledgements

The authors would like to express deep and sincere gratitude to Dr Tianhu Deng, Tsinghua University for providing invaluable guidance and extensive support and encouragement.

The authors would like to appreciate Dr Hans Levenbach, CPDF for support, inspirational comments and precious hints on an earlier stage of the project.

References

Adhikari, A., DeNero, J (2016), "The Foundations of Data Science, Online Textbook, In partnership with the Berkeley Division of Data Sciences". Data-8.github.io, https://www.inferentialthinking.com/chapters/13/2/Bootstrap.html

Ahmadi Mobarakeh, N.A., Shahzad, M.K., Baboli, A., Tondare, R. (2017), "Improved Forecasts for uncertain and unpredictable Spare Parts Demand in Business Aircraft's with Bootstrap Method". Elsevier, IFAC Papers OnLine 50-1 (2017) 15241–15246. DOI: 10.1016/j.ifacol.2017.08.2379

Brockwell PJ, Davis RA (2016). “Introduction to time series and forecasting (Springer text in statistics)”. Cham: Springer International , Publishers; 2016.2. Introduction to Capital market and problem in Indonesia.

Budiharto, W (2021), “Data science approach to stock prices forecasting in Indonesia during Covid-19 using Long Short-Term Memory (LSTM)”, Budiharto J Big Data 8:47, https://doi.org/10.1186/s40537-021-00430-0, (2021)

Flores-Agreda, D., Cantoni, E. (2019), “Bootstrap estimation of uncertainty in prediction for generalized linear mixed models”, *Computational Statistics & Data* **130**, 1-17. DOI: 10.1016/j.csda.2018.08.006.

Fresoli, D., Ruiz, E., Pascual., L (2015), “Bootstrap multi-step forecasts of non-Gaussian VAR models”.International Journal of Forecasting 31 (3), 834–848. DOI: 10.1016/j.ijforecast.2014.04.001.

Gorrie, C (2016), “Three ways to detect outliers, Colin Gorrie's Data Story”. Unpublished data, http://colingorrie.github.io/outlier-detection.html

Kim, H., Kim, Jintae (2018), “London calling: Nonlinear mean reversion across national stock markets” North American Journal of Economics and Finance 44, 265–277. DOI: 10.1016/j.najef.2018.01.008.

King, G., Pan, J., Roberts, M. E (2014), “Replication Data for: Reverse Engineering Chinese Censorship”. Randomized Experimentation and Participant Observation.

Levenbach, H (2017), “Change & Chance Embraced – Achieving Agility with smarter Forecasting in the Supply Chain”. Delphus Publishing Inc. ISBN-10: 0692945989, ISBN-13: 978-0692945988

Makhwiting, M .R (2014), “Modeling Volatility and Financial Market Risks of shares on the Johannesburg stock exchange”. Master of Science Dissertation, Faculty of Science and Agriculture (School of Mathematical and Computer Sciences), University of Limpopo.

Pek, J., Wong, O., Wong, A C. M. (2018), “How to Address Non-normality: A Taxonomy of Approaches, Reviewed, and Illustrated”. doi.org/10.3389/fpsyg.2018.02104

Sedighi, M., Jahangirnia, H., Gharakhani, M., Farahani Far, S (2019), “A Novel Hybrid Model for Stock Price Forecasting Based on Metaheuristics and Support Vector Machine”. Data 2019, 4, 75; doi:10.3390/data4020075

Shimizu, S., Kano, Y (2008), “Use of non-normality in structural equation modeling: Application to direction of causation “. Journal of Statistical Planning and Inference, 138 (11), 3483–3491. DOI: 10.1016/j.jspi.2006.01.017.

Yong L., Wei-Ping H., Jie Z (2013), “Forecasting volatility in the Chinese stock market under model uncertainty”. Economic Modelling,. Elsevier, vol. 35 ©, 231-234. DOI: 10.1016/j.econmod.2013.07.006

Zhang, X-L (2018), “Multilayer bootstrap networks. In Neural networks”. The official journal of the International Neural Network Society 103, 29–43. DOI: 10.1016/j.neunet.2018.03.005.

Xiong, X., Yuxiang, B., Dehua S (2018), “The time-varying correlation between policy uncertainty and stock returns: Evidence from China”. In Physica A: Statistical Mechanics and its Applications 499,. 413–419. DOI: 10.1016/j.physa.2018.02.034.