International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.960

Hybrid time-series and time-frequency model for stock index prediction

Navin S Patel[1] and Y T Krishne Gowda[2]
[1]Research Scholar in Management Sciences, University of Mysore,
1017-A, Telecom Road, Roopanagar, Mysore – 570026
[2]Professor and Principal, MIT, Mysore

Received: 10 April Revised: 18 April Accepted: 26 April

Abstract

Prediction in financial markets is as old as the markets themselves. Short term speculators and long term investors are continuously on a lookout for techniques which can get them abnormal returns. There are a wide variety of techniques and methodologies which try to reduce the forecast errors. One of the oldest technique is the Auto-regressive Integrated Moving Average (ARIMA) model. These class of models assume that a stock index is basically a time-series and that an index is a linear stationary random processes. Based on this assumption, a model is developed by exploiting the temporal correlations within the time-series The fundamental aim of developing these models is to develop a measure of forecastability in the time-series. Over the years, there have been many attempts to improve the predictive performance of ARIMA models. In this paper, we use a hybrid model combining ARIMA model with Maximum Overlap Discrete Wavelet Transform (MODWT), a concept adopted from signal processing. We use a NIFTY 50 series to create a standalone ARIMA model and also create a hybrid MODWT-ARIMA model. We find that the hybrid model shows lower forecast errors compared to the standalone model

Key words: Stock index forecasting, ARIMA, MODWT, hybrid models.

1. Introduction

Investors and traders seek stock markets for investment opportunities. Irrespective of the investment horizon – short-term day trading or long term investments, the stock market offers something to everybody. It is not only the retail investors, but even corporate bodies partake in the stock market. But being involved in stock market is always a risky proposition as decision making is often based on emotions and tips from speculators. India has had a good capital markets ecosystem since long where stock market is a critical component. The 1990s reforms ensured increased participation in these markets by retail investors and institutional investors, both domestic and foreign. A key influencer of an investment decision is based on the anticipation of future returns. In this context, the forecasting methods play an important role

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.960
The early part of the twentieth century saw major developments in the theory of modeling stochastic processes. Fisher's seminal contribution maximum likelihood estimation (MLE) methods to estimation theory started a wave of similar research work. Yule and Walkerprocesses (**Walker, 1931; Yule, 1927**) made significant contributions in the study of time-series analysis, a field in statistical signal processing. The main goal of these studies was predict random phenomena. In the 1970s, there was an uptick in time-series modeling techniques using the autoregressive integrated moving average (ARIMA) model.**Box and Jenkins (1970)**created the Box-Jenkins methodology for a systematic analysis of time-series data and for predicting future values. While these models achieved significant success, efforts were always on to improve the prediction efficiencies. Apart from creating a new set of models, effort was also spent on how to improve the ARIMA models. One such approach is to create hybrid ARIMA models by using wavelets for decomposing original time-series. The central idea of wavelets analysis in time-frequency domain and to generate different frequencies of data embedded in the original signal and model each of these decomposed frequencies using ARIMA.

In this study, we select a NIFTY 50 index and:

1) Create an ARIMA model for the NIFTY index which is a standalone prediction model
2) Decompose the index into wavelet components using Maximum Overlap Discrete Wavelet Transform (MODWT), a type of wavelet transforms. Create ARIMA models for the decomposed components to create a hybrid prediction model
3) Predict the future values of the standalone and hybrid components. Compare the forecast accuracy of both the models with the actual values

The paper is organized as: In Section I we give a brief introduction of stock index prediction, Section II containsa discussion of previous work on ARIMA and wavelets, Section III provides theoretical background, Section IV contains data and methodology, Section V describes results and Section VI concludes with suggestions for future research

2. Related Work

Zhang et al. (2009) studied the Shanghai Composite Exchange and took 1-year closing day values. They created a ARIMA (1,1,0) model for this data and it performed reasonably well in terms of low prediction errors. **Adebiyi et al. (2014)** created an ARIMA model for the index data from NYSE and Nigeria Stock Exchange to forecast the future values. They concluded that their models had a high potential to forecast for short-term horizons compared to other methods. **Yermal and Balasubramanian (2017)** used ARIMA approach to model high frequency stock price for selected stocks in NSE. MAPE was used to measure the forecast errors. Except for 3 stocks, their models performed satisfactorily.

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.960

Many researchers have tried a hybrid models using wavelets and ARIMA. **Kriechbaumer et al. (2014)** used the Wavelet-ARIMA method to predict the monthly prices of Lead and Copper and Zinc. They used two varieties of wavelet transforms – Discrete Wavelet Transform (DWT) and Maximum Overlap Discrete Wavelet Transform (MODWT) to decompose the original time series into wavelet components then yused ARIMA model to forecast future monthly data. They used RMSE and MAE as the metrics for prediction accuracy. They showed their hybrid models performed better than the standalone ARIMA model. **Khandelwal et al. (2015)** used DWT to deconstruct four time-series data signals and used ARIMA and Artificial Neural networks to build forecast models. MAPE and MSE were the metrics they used to measure forecasting accuracy. The DWT based hybrid model performed better than the standalone models.

3. Theoretical Background

3.1. Time-series models

We present a short review of autoregressive integrated moving average models (ARIMA) for the linear random process. The stock indices are time-series signals considered to be emanating from linear random processes. The centralpremise in modeling linear random processes lie in the fact they can be thought of as comprising of a predictable deterministic portion and an unpredictable random component. From a Linear Time, Invariant (LTI) system perspective, a random process which is stationary can be shown to be outputs of LTI filters driven by random white noise. If $v[t]$ is the discrete time-series and $e[t]$ is the white noise, the random process can be represented as

$$v[t] = H(q^{-1})e[t]$$

The representation is also given by

$$v[t] = \sum_{n=-\infty}^{\infty} h_n \, e[t-n]$$

Where $e[t]$ is a white noise process with constant variance and zero mean.

Moving Average (MA) Models

An MA process is one whose current state is a combination of previous shock waves and white-noise. The MA process of order M has the following form

$$v[kt] = \sum_{n=-1}^{M} h_n \, e[t-n] + e[kt]$$

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.960

The Autocorrelation Function (ACF) of an MA(M) random process has a sharp cut-off after M lags which helps in determining the order of a pure MA process.

Auto-regressive (AR) Models

An AR process is a linear combination of its past data. The AR representation of order P has the can be shown to be: $v[t] = \sum_{j=-1}^{P} (-d_j)\, v[t-n] + e[t]$

Unlike an MA process, it is the partial autocorrelation function (PACF) and not ACF which determines the order of the process

Auto-regressive Moving Average (ARIMA) Models

In nature, one usually cannot find a process which is pure AR or MA type. Most process often are a combination of both which is the auto-regressive moving average models - ARMA (P, M), but many processes are non-stationary and they have to be differenced to make them stationary. So an ARMA model on a differenced series becomes a ARIMA (P, D, M) model where D represents the number of times the series is differenced.

3.2. Time-Frequency Analysis

Wavelet transforms are a type of time-frequency analysis which helps in decomposing a time-series into different frequencies at different time-scales. They transform the series into multiple components - trend and detail components. The trend components represent the low frequency part of the data whereas the detail components amplify the high frequency component of the data. The trend component can be decomposed iteratively into its own trend and detail parts and this process can be repeated to any level. The Discrete Wavelet Transform (DWT) is one of the most popular way to decomposed the signal. But the number of data points reduce by 50% for every level of decomposition. This is overcome by using Maximum Overlap Discrete Wavelet Transform (MODWT) which maintains the cardinality of the original data series irrespective of the number of levels of decomposition

4. Data and Methodology

NIFTY S&P 50 index was used as the index in this study. The time-series consisted of the daily closing value the index from Jan-2008 to Dec-2017, a total of 2651 data points. After the standalone and hybrid model configurations were finalized, they were used to forecast the future 15 values. The forecasted values were compared with the actual values to derive the forecast metrics.

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.960

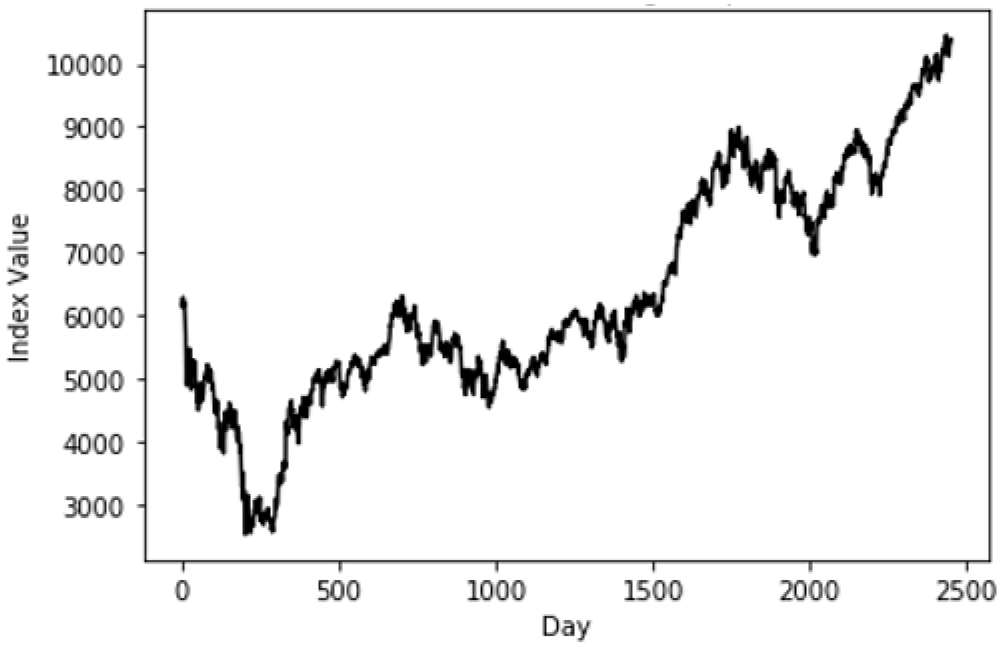

Figure 1: Time-series used in the models

Forecast Metrics:

RMSE: Root Mean Square Error (**RMSE**) measures a representation of the deviation between values predicted by a model with the values actually observed.:

$$\text{RMSE}= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(A_i - F_i)^2}$$

where A_i is actual value and F_i is the predicted value.

MAD: Mean Absolute Deviation measures the forecast error in terms of absolute deviations

$$\text{MAD} = \frac{1}{N}\sum_{i=1}^{N}|A_i - F_i|$$

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.960

MAPE: Mean Absolute Percentage Error (MAPE) measures the average absolute percent error for forecasted value over the actual value

$$\text{MAPE} = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{A_i - F_i}{A_i}\right|$$

In this study, results were captured for n=5,10,15

ARIMA method for forecasting:

1) Perform a visual inspection of the series to check for non-stationarity.
2) If this hints at non-stationarity, perform KPSS test to check for non-stationarity.
3) If the series is non-stationary, perform differencing operation iteratively till the series becomes stationary. KPSS test is used to statistically test if the resultant series is stationary.
4) Calculate ACF and PACF for the series to get an initial insight into the order of AR and MA processes. Iteratively, create models for different values for AR and MA combinations.
5) Use information theoretic measure like Akaike's Information Criterion (AIC) and the principle of parsimony to select the best fit model
6) Ensure the residuals of the fitted model don't exhibit auto-correlation
7) Forecast the 15 future values using the fitted values and capture the forecast metrics

MODWT-ARIMA method for forecasting:

1) Use Haar MODWT to decompose the time-series into trend and detail components. 4 levels of decomposition were experimented with
2) For each level of decomposition, one trend component and multiple detailed wavelets will be generated.
3) Fit an ARIMA model to each of these wavelet components and forecast the future valued
4) Perform inverse wavelet transform and aggregate the forecast values. Capture the forecast metrics.
5) Select the best fit Wavelet-ARIMA model based on forecast metrics

5. Results and Discussion

Standalone ARIMA model:Table 1 provides the details of the ARIMA model for the series and **Table 2**provides the details of the forecast metrics

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.960

Table 1 Standalone ARIMA Model

No of data points trained	2451
Fitted ARIMA model	ARIMA (0,1,1)

Table 2 ARIMA Standalone Forecast Metrics

Forecasting Window	RMSE	MAD	MAPE (%)
5- Day	148.11	103.77	0.99
10-Day	218.41	182.69	1.75
15-Day	199.72	170.67	1.63

Hybrid MODWT-ARIMA model:Table 3 provides the details of the ARIMA models for the decomposed wavelet components for the series and **Table 4**provides the details of the forecast metrics for the hybrid model

Table 3ARIMA Model for wavelet components

No of data points trained	2451
Fitted ARIMA model for Trend component	ARIMA (0,1,0)
Fitted ARIMA model for detailed component-1	ARIMA (1,1,1)
Fitted ARIMA model for detailed component-1	ARIMA (1,1,0)

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.960

Table 4 Forecast metrics for the MODWT-ARIMA model

Forecasting Window	RMSE	MAD	MAPE (%)
5- Day	130.75	87.69	0.86
10-Day	193.16	158.18	1.56
15-Day	171.68	140.45	1.38

MODWT-ARIMA method resulted in the lowest forecast errors across all the three metrics – RMSE, MAD and MAPE.

6. Conclusion

In this paper, we created two models for stock index forecasting. One was based on the traditional time-series method using the ARIMA framework. The second one was a hybrid model using time-frequency analysis for which MODWT based wavelet transform were used. The hybrid model showed superior forecasting accuracy. In this study, we used Haar family of wavelet for decomposition. There are multiple other wavelet families which can be explored to evaluate if they will further increase the forecasting power.

7. References

1. Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock Price Prediction Using the ARIMA Model. *AMSS 16th International Conference on Computer Modelling and Simulation*
2. Box G.E.P, G. Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, CA, 1970.
3. Khandelwal, I., Adhikari, R., and Verma, G. (2015). Time Series Forecasting using Hybrid ARIMA and ANN Models based on DWT Decomposition. *ICCC, Procedia Computer Science*, pp. 173 – 179
4. Kriechbaumer, T., Angus, A., Parsons, D., and Casado M.R. (2014). An improved wavelet–ARIMA approach for forecasting metal prices. *Resource Policy*. pp. 32 - 41
5. Walker G., On Periodicity in Series of Related Terms. Proceedings of the Royal Society of London, Ser. A 131, 518–532, 1931
6. Yule G , On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character 226, (1927): 267–298

7. Yermal, L., & Balasubramanian P. (2017). Application of Auto ARIMA Model for Forecasting Returns on Minute Wise Amalgamated Data in NSE", Computational Intelligence and Computing Research (ICCIC) 2017 IEEE International Conference on, pp. 1-5
8. Zhang, J., Shan, R., & Wenfang Su (2009). Applying Time Series Analysis Builds Stock Price Forecast Model, *Modern Applied Science*, Vol 3, No.5