Determining the Optimal Forecasting Model for the Dow Jones Industrial Average (DJI)

Alisa Dmitrieva, Koki Yamanaka, Kumari Herath

Abstract - This paper presents a comprehensive analysis of the forecasting techniques applied to the DOW Jones Industrial Average (DJI), a leading equity index comprising 30 prominent companies in the United States[1]. The study focuses on utilizing various methods, including benchmark, ARIMA, neural network, EWMA, Prophet, regression, and dynamic regression, to predict the future movements of the DJI based on a three-year daily closing price dataset. In addition, external factors such as the company that constitutes for DJI and other big markets indexes are used as predictors and tested on regression models. The investigation delves into the realm of time series forecasting, evaluating and comparing these diverse techniques with real-world data. Our primary objective is to identify the most accurate forecasting method for the DJI. Specifically, three models—the naive model, a dynamic regression model, and an EWMA model were selected for an in-depth comparison, aiming to determine the technique with the lowest forecasting error in our scenario. Initially, it seemed like the Prophet model was the most effective in predicting the fluctuations of the DJI. To conclude, we analyzed more on RMSE and visualization for NAÏVE, Dynamic and Prophet models and decided that the NAÏVE method is the best and most reasonable forecast method for the DJI.

1.0 Introduction

It is notoriously difficult to predict stock market data - there have been many methods and models devised for this purpose. The DOW Jones Industrial Average (DJI) is a prominent equity index which was chosen for forecasting purposes because it is often seen as a benchmark to gauge the overall performance of the U.S stock market[1].

The purpose of this project is to determine the optimal method for forecasting the DJI by comparing different forecasting models and selecting the model(s) with the best predictive accuracy. If a sufficiently accurate forecasting model for the DJI could be identified, then it would be theoretically possible to predict the performance of the entire U.S stock market. An accurate forecasting model of the DJI would have numerous applications in various industries and would thus be a powerful tool.

To achieve this, the benchmark models (naïve, mean, and drift) were fitted and analyzed in addition to more complex models including ARIMA/SARIMA, neutral network, dynamic regression, linear regression, Prophet, and EWMA. Residual diagnostics and accuracy metrics were carefully performed on each model.

Finally, forecasts produced by each type of model were carefully analyzed on a variety of metrics in order to select the best model.

2.0 Data

The data used in the project was the daily closing price of the Dow Jones Industrial Average (^DJI) which was extracted using the Yahoo Finance library through the “quantmod” package. The training period was from '2020-10-01 ' to '2023-09-30', and the test period was from '2023-10-01' to '2023-11-28’. There were 794 observations overall in the training and test data sets combined. There were 754 observations in the training data set and 40 observations in the test data set (representing 2 months of stock market trading data). The sign correlation calculation indicated that the overall data set followed a normal distribution. The mean was 33231 and the standard deviation was 1950.978. An upward trend and some remainder of white noise were detected through an STL decomposition model.

3.0 Method

3.1 Data Preparation

The raw data was transformed into a Tsibble, and the missing values were addressed in order to maintain data integrity. Additionally, the data was transformed to make it stationary for the requisite of some methods we used for reliable analysis.

Three different studies were performed to compare different types of models. The method of each of the three studies is summarized below in 3.2-3.4.

3.2 Dynamic Regression & Time Series Linear Model

Here we employ 2 different groups of experiments using dynamic (DR) and linear regression(TSLM) with the aid of the “fable” package in R. Using identical train/test time frames for predictors, our experiment divides into two groups. Group 1 employs 1TSLM and 1DR models, centered on Visa (^V), a 4% DJI constituent. Group 2 tests various 3DR models with predictors from big market indices (FTSE, N225, FTSE+N225), utilizing an automatic parameter chooser. Before fitting, both groups underwent correlation analysis and normalization (min-max and z-score). Group 1 used linear regression, while Group 2 examined correlation through a matrix and scatter plot and a decomposition model has been fitted. Model diagnostics included analysis of residuals, distribution, ACF, and PACF. Please refer to Hyperparameters from Table 1.0. This architecture is effective due to the substantial influence of all four predictors on DJI. Before fitting, an upward trend in the long term and random walk patterns in the short term are observed, thus TSLM and DR are best suited to capture these.

3.3 Benchmark Models (Naive, Mean, Drift), ARIMA/SARIMA, and Neural Network

The initial phase of modeling involved the implementation of a benchmark model, which included naive, mean, and drift methods. We used this model as our base and following this model, we used ARIMA and Neural networks models to explore more on the dataset.

When the ARIMA model was applied, with an initial identification of ARIMA(0,1,0) as the best fit. To ensure the robustness of this choice, we did an exploration of alternative ARIMA configurations and came to a conclusion that ARIMA(0,1,0) is the best fit ARIMA model. Since the best fit ARIMA model has no average moving or mean average components we did not use this method for further investigation.

Next, we studied the Neural Network (NN) model to capture intricate relationships within the DJI data. We analyzed the residuals of the NN model to confirm their adherence to white noise and a normal distribution, validating the model's appropriateness for the dataset.

Furthermore, we used accuracy metrics to compare these forecasting models based on the Root Mean Squared Error (RMSE). Notably, the MEAN model exhibited the smallest RMSE among the considered methods. Additionally, we used all the methods for forecasting to visualize any reduced uncertainty or any unusual patterns, to compare with the MEAN model. However, the forecasting phase revealed that, despite its lower RMSE, the NAIVE model demonstrated a more favorable predictive pattern with reduced uncertainty compared to the MEAN model.

Ultimately, based on a comprehensive evaluation, the Neural Network model emerged as the most suitable and accurate method for forecasting DOW Jones Industrial Average movements, outperforming other models in the study

3.4 Prophet and EWMA

The default Prophet model from Facebook was fitted to the daily closing price of the DJI, using the “prophet” package. The Prophet model appeared to be a good fit for the data from a visual inspection. Next, the forecasts from the Prophet model were plotted and they appeared to capture the trend well when compared to the test data, which also showed a moderate positive trend. A decomposition was performed on the Prophet model to examine its components in further detail.

After this, the sign correlation value was calculated for the daily closing price of the DJI. The sign correlation value indicated that the daily closing price of the DJI for the specified dates followed a normal distribution.

Next, an EWMA model was fitted to the daily closing price of the DJI using the ets() function from the “forecast” package. The forecasts generated by the EWMA model showed no trend. They also showed greater variance when compared to the forecasts produced by the Prophet model.

The Prophet and EWMA models were fitted again to the data using the “fabletools” and “fableprophet” packages. Residual analysis was performed on both models, and it was determined that the Prophet model residuals were strongly autocorrelated and were thus not white noise. However, the EWMA model performed very well in the residual analysis.

Finally, various accuracy metrics were calculated for both models, using both the training and test data. The EWMA model was the clear winner in all accuracy metrics, including the Root Mean Squared Error (RMSE).

Considering all of the above, it was decided that the EWMA model performed much better than the Prophet model in forecasting the daily closing price of the DJI.

Link to project repository.

4.0 Results

4.1 DR & TSLM :HSI predictor is omitted in our test due to its -0.03 correlation with DJI. 2 Key observations are revealed. [1] DJI is more influenced by N225 than FTSE. [2] ARIMA(0,1,0) exhibits the least RMSE for both train and test, suggesting that short-term movements in DJI are best captured by immediate past data; higher-order lag values not significantly improving the model. This is verified by table y. (a) ARIMA(0,1,0), random walk performs better than models that consider larger lag values ARIMA(3,1,3) and ARIMA(0,1,3) (b) Most ACF are bounded to 0, indicating the DJI series is random. In the short term, DJI itself shows no distinct pattern, making it challenging to predict, while a long-term upward trend is captured effectively. Overall, the approach is effective, as DR was able to capture the long term upward trend with the aid of N225 and its ARIMA error captured the random dynamics in the short term. Note, cross validation was not able to perform due to time constraint.

Model no.	Model/ metrics	Predictors	Train RMSE	Test RMSE	ACF	PACF
1	MEAN	-	-	216	-	-
2	NAIVE	-	-	361	-	-
3	DRIFT	-	-	372	-	-
4	NNEAR(1,1)	-	-	714	-	-
5	TSLM	Visa Close	1493.8102	1493.8102	Geom decay	Lag 1 spikes
6	DR ARIMA(0,1,0)	Visa Close	227.1161	467.969	B0	Similar to ACF
7	DR - ARIMA(3,1,3)	FTSE Close	272.4554	1099.418	B0	
8	DR - ARIMA(0,1,1)	N225 Close	303.3204	731.225	B0, Differ to model 7,9	
9	DR - ARIMA(0,1,3)	FTSE + N225 Close	271.5634	1006.004	B0	

10	**EWMA**	**-**	**317**	**875**	**-**	**-**
11	**Prophet**	**-**	**780**	**1190**	**-**	**-**

Table 1.0

Note: B0: ACF bounded to 0.

5.0 Discussion/Conclusions/Limitations

Model 1-4 : The study concludes that the NAÏVE model outperforms other models, including ARIMA(0,1,0) and Neural Network (1,1), in forecasting the DOW Jones. The simplicity of the Naive method proves effective in capturing stock market dynamics, emphasizing the need for tailored models. **Model 5 - 9** : Dynamic regression with the inclusion of external factors like N225, forecasts were inaccurate. The best model from the Model 5 - 9 experiment is ARIMA(0,1,0), random walk. Thus, it suggests DJI itself does not exhibit any certain pattern. **Model 10 - 11**: The Prophet model's visual appeal and apparent accuracy in forecasting may lead to a misconception, as its effectiveness is compromised by pronounced autocorrelation in residuals. The decomposition of the Prophet model reveals a misleading identification of a yearly trend, rendering it inappropriate for reliable predictions.

The general conclusion is that the benchmark models achieve better results than the more complex models, suggesting the DJI itself is a random walk in the short term. From table 1.0, the benchmark models have smaller RMSE when compared to other, more complex, models.

Overall, the naive method was identified as the best method for forecasting the price of the DJI. This is often the case with stock market data because the price of a stock at any given point in time is meant to reflect the optimal value due to supply and demand forces in the market. Thus, the most recent closing price of a given stock usually produces decent forecasts.

Although the mean model produced forecasts with the lowest RMSE overall, these forecasts exhibited very high variance. The naive model forecasts had the second-lowest RMSE value. However, the naive model produced forecasts with much less variance than the mean model.

There were three key limitations in this project. First of all, the training set of 3 years was possibly insufficient, which impacted the forecast accuracy of each model. Thus, selecting a training set with a longer time period could improve forecasts. Secondly, most tested models failed to capture the dynamic complexity, and thus deeper neural networks and combinations of different models may be able to better incorporate that. Lastly, more reliable tests for our models are required as some models did not undergo cross validation.

References

1. Kenton W. S&P 500 Index: What It's for and Why It's Important in Investing.

2. Forecasting: Principles and Practice (3rd ed) Rob J Hyndman and George Athanasopoulos
Monash University, Australia