Predicting Future Index funds using Historical Data

**Introduction**:

Index funds' low costs and passive investment strategy have made them more and more popular among investors. These funds aim to duplicate the returns of a certain market index, such as the S&P 500 or the Dow Jones Industrial Average, by tracking its performance. The objective is to give investors a wide range of market exposure while lowering the chance of performing poorly compared to the market. This project uses time series analysis to forecast the future performance of three index funds, iShares Core S&P Small-Cap ETF (IJR) and Schwab International Index Fund (SWISX). A variety of models, including ARIMA and linear regression, will be used to anticipate future performance once we have analyzed historical pricing data and other pertinent characteristics. In order to provide investors with a reliable and accurate technique for maximizing their benefits while minimizing risk, this research aims to determine the most effective strategy for predicting the future performance of index funds. Individual investors, financial advisors, mutual fund firms, ETF providers, institutional investors, researchers, and financial regulators who are interested in improving investing decision-making and optimizing returns while minimizing risk would find the research's conclusions useful.

**Background research:**

The interest in index funds is expanding for a variety of reasons. According to studies (Fama & French, 2010; Malkiel, 1995), passively managed index funds typically beat actively managed funds over the long term on a net-of-fees basis. Investors are also getting increasingly concerned about the taxes and costs related to actively managed funds, which can reduce returns over time (Carhart, 1997). Index funds are a desirable choice for investors who favour a hands-off approach to portfolio management due to their passive nature. Based on previous data, time series analysis is a potent tool for predicting the performance of financial assets in the future. Stock prices, currency rates, and other financial variables have been successfully predicted using methods like linear regression and autoregressive integrated moving average (ARIMA) models (Box & Jenkins, 1976; Engle & Granger, 1987). This research aims to find the best technique for predicting the future performance of index funds by studying historical price data, trading volume, liquidity, and the opening and closing prices of index funds. However, there remains significant room for improvement in the accuracy and reliability of these models, particularly when applied to index funds.

**Problem Presentation:**

Due to their low fees, passive management strategy, and wide market exposure, index funds are becoming increasingly popular among investors, increasing the demand for precise performance predictions. This project seeks to address this issue by using time series analysis methods to forecast the future performance of three index funds: iShares Core S&P Small-Cap ETF (IJR) and Schwab U.S. Aggregate Bond Index Fund (SWISX). The objective is to create a trustworthy forecasting technique based on historical price information, trading volume, liquidity, and the opening and closing prices of the index funds, which will empower investors to make wise decisions, maximize returns, and reduce risk.

This study aims to contribute to the development of a reliable and trustworthy approach for forecasting index fund performance by improving and expanding prior research. The successful conclusion of this project will benefit a wide range of stakeholders, including financial advisors, mutual fund providers, ETF providers, institutional investors, academic researchers, and financial regulators, by giving investors an invaluable tool for forecasting the future performance of index funds. The ultimate goal of this research is to support the creation of investment portfolios that are more reliable and efficient.

**Specification and Design:**

iShares Core S&P Small-Cap ETF (IJR) and Schwab U.S. Aggregate Bond Index Fund (SWISX). are the index funds that this project is primarily focused on. This project’s scope will focus on 3 years of historical data of the 3 index funds. With predictor variables such as historical price data, trading volume, liquidity, opening price, closing price, and past performance of comparable index funds, the approach incorporates time series analysis, linear regression, and ARIMA. The ultimate objective is to give investors precise forecasts so they may make informed decisions and reduce risk.

Descriptive: Conducting exploratory data analysis on historical price data, trading volume, liquidity, opening price, closing price, and the performance of comparable index funds are all part of the project's descriptive component. This will make it easier to spot trends, patterns, and connections between various variables.

Predictive: To predict the future performance of the iShares Core S&P Small-Cap ETF (IJR) and Schwab U.S. Aggregate Bond Index Fund (SWISX)., the predictive component develops and trains models using time series analysis, linear regression, and ARIMA. To gauge the accuracy and dependability of these models, they will be tested on a different dataset and verified using methods like cross-validation. The target index funds' upcoming performance will be forecast using the model that has performed the best.

Prescriptive: Based on the findings of the predictive analysis, the project's prescriptive component intends to offer actionable insights and suggestions for investors, financial advisors, mutual fund firms, ETF providers, and other stakeholders.

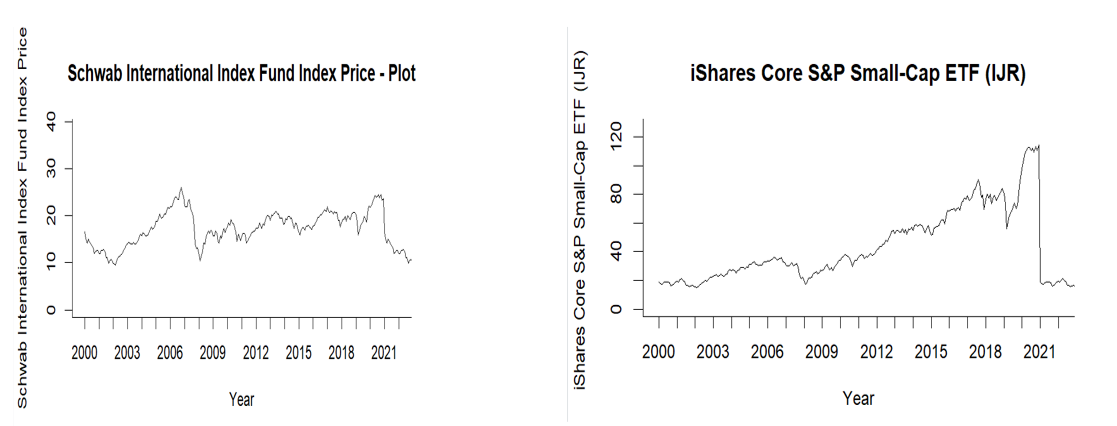
**Data Acquisition:**

For the project, historical price information will be gathered from Yahoo Finance and Google Finance for the iShares Core S&P Small-Cap ETF (IJR) and Schwab U.S. Aggregate Bond Index Fund (SWISX). In addition, financial databases like Morningstar or other trustworthy sources will be used to acquire trading volume, liquidity data, and historical performance information for index funds that are similar to them. Understanding the market's behaviour and the performance of the target index funds will be based on this data.

* <https://finance.yahoo.com/quote/IJR?p=IJR&.tsrc=fin-srch>
* <https://finance.yahoo.com/quote/SWISX?p=SWISX&.tsrc=fin-srch>
* <https://finance.yahoo.com/>

**Data Exploration:**

The gathered data will be processed and visualized to find trends, patterns, and correlations between variables during the data exploration phase. This involves producing summary statistics, making different graphs, looking at correlations, and performing seasonality decomposition, among other things. The project will create a strong foundation for creating, training and validating predictive models to anticipate the future performance of the target index funds by performing an extensive data investigation. Below are the price plots for both index funds from Jan 2000 to Dec 2021.

****

**Models used for Index Funds:**

SWISX & IJR Index Funds

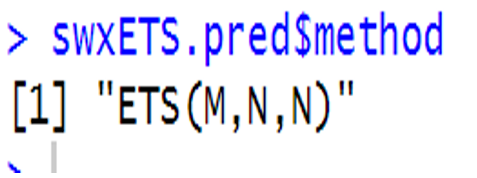
→ Exponential Smoothing

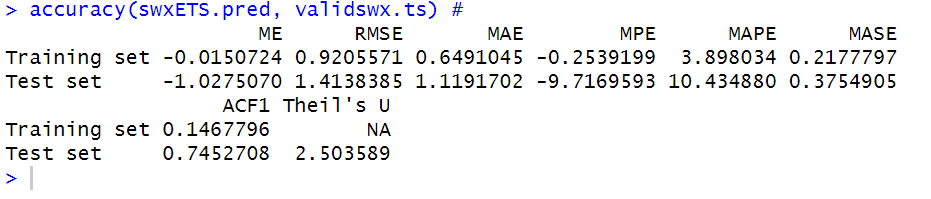
→ Arima

→ Neural Networks

**Data set 1:** Schwab U.S. Aggregate Bond Index Fund (SWISX)

Exponential Smoothing





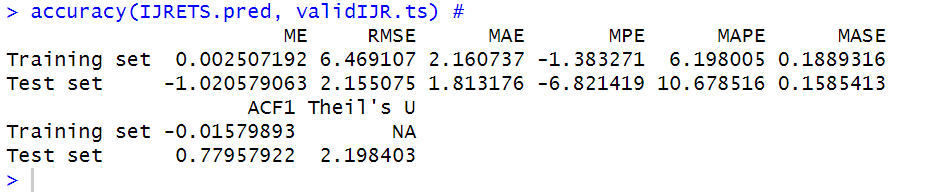
Chart, line chart

Description automatically generated

This model helps forecast trends and seasonal patterns in the data. The ETS (M, N, N) model used here includes multiplicative error, with no trend or seasonality components. The accuracy function in the output evaluates the model's performance using metrics like Mean Error, Root Mean Squared Error, Mean Absolute Error, and others.

The outcomes demonstrate that the model fits the training data reasonably well, with a low mean error and a relatively low root mean squared error. A negative mean percentage error and a somewhat larger mean absolute percentage error, when applied to the test set, show that there are some inconsistencies. The ETS model is still a viable option for projecting the bond index fund despite these inconsistencies.

**Data set 2:** iShares Core S&P Small-Cap ETF (IJR)  
  

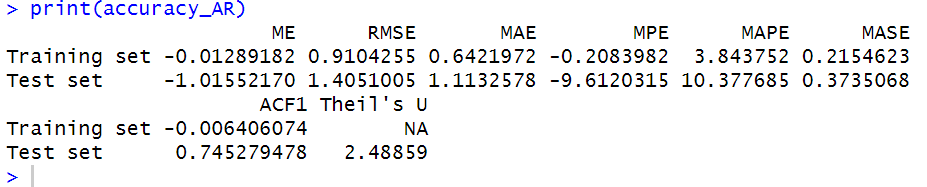
Chart, line chart

Description automatically generated

The ETS (M, N, N) model used in this case consists of multiplicative error, with no trend or seasonality components.

With a nearly zero mean error (ME = 0.002507192) and a comparatively small root mean squared error (RMSE = 6.469107), the results show that the model fits the training data well. The model does, however, show some anomalies when applied to the test set, as seen by the negative mean percentage error (MPE = -6.821419) and a somewhat larger mean absolute percentage error (MAPE = 10.678516). Theil's U is 2.198403 and the model's autocorrelation function at lag 1 (ACF1) for the test set is 0.77957922. The ETS model can still be viewed as a viable choice for predicting the iShares Core S&P Small-Cap ETF despite the differences in the test set.

**ARIMA  
Dataset 1:** Schwab U.S. Aggregate Bond Index Fund (SWISX)

****

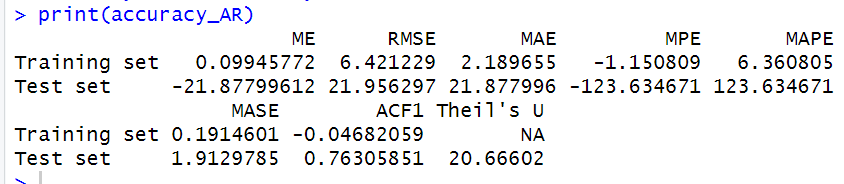
Chart, line chart, histogram

Description automatically generated

The accuracy function in the output evaluates the model's performance using different evaluation metrics, such as Mean Error (ME), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), and others.

For the training set, the model has a low mean error (ME = -0.01289182) and a relatively low root mean squared error (RMSE = 0.9104255), indicating a good fit for the training data. The mean absolute percentage error (MAPE) is 3.843752, which suggests a reasonable level of accuracy for the model's forecasts. The model exhibits various inconsistencies when used with the test set. A few deviations from the model's predictions are indicated by the negative mean percentage error (MPE = -9.6120315) and slightly greater mean absolute percentage error (MAPE = 10.377685). Theil's U is 2.48859 and the model's autocorrelation function at lag 1 (ACF1) for the test set is 0.745279478, indicating that the ARIMA model is still a good choice for predicting the Schwab U.S. Aggregate Bond Index Fund.

**Dataset 2:**

****

Chart, line chart, histogram

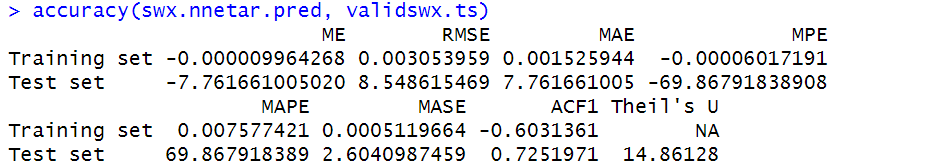
Description automatically generated

The model displays low values for ME (0.09945772), RMSE (6.421229), and MAE (2.189655) for the training set. These metrics imply that the model successfully matches the training set of data. The MPE (-1.150809) and MAPE (6.360805) also show a decent level of forecast accuracy for the model.

But when used on the test set, the model exhibits notable inconsistencies. The extraordinarily high readings for MPE (-123.634671) and MAPE (123.634671) imply that the model's predictions are significantly off. Despite this, Theil's U is 20.66602 and the model's autocorrelation function at lag 1 (ACF1) for the test set is 0.76305851. The ARIMA model may not be the greatest option for forecasting this particular dataset, as evidenced by the high Theil's U value, as the model's forecasts for the test set diverge greatly from the actual values.

**Neural Networking:**

**Dataset 1:**

****

Chart, line chart

Description automatically generated

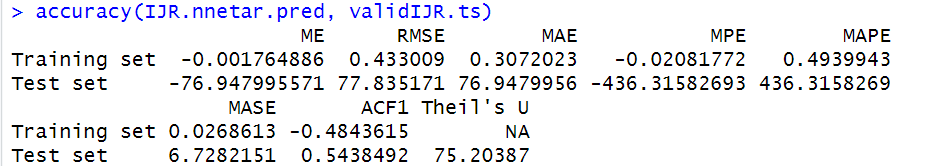
NNetAR is a time series forecasting method that uses feed-forward neural networks with lagged inputs for autoregression. It's a flexible model capable of capturing complex patterns and relationships in the data.

The accuracy function in the output evaluates the model's performance using different evaluation metrics, such as Mean Error (ME), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), Mean Absolute Scaled Error (MASE), and others.

For the training set, the model exhibits low values for ME (-0.000009964268), RMSE (0.003053959), and MAE (0.001525944). These metrics suggest that the model fits the training data very well. The MPE (-7.761661005020) and MAPE (8.548615469) indicate a reasonable level of accuracy for the model's forecasts.

But when used on the test set, the model exhibits some inconsistencies. The high MPE (-69.86791838908) and MAPE (69.867918389) values indicate that the model's predictions may not be entirely accurate. Theil's U is 14.86128 and the model's autocorrelation function at lag 1 (ACF1) for the test set is 0.7251971. These results suggest that, despite any inconsistencies in the test set predictions, the NNetAR model is still a viable option for projecting Dataset 1.

**Dataset 2:**

****

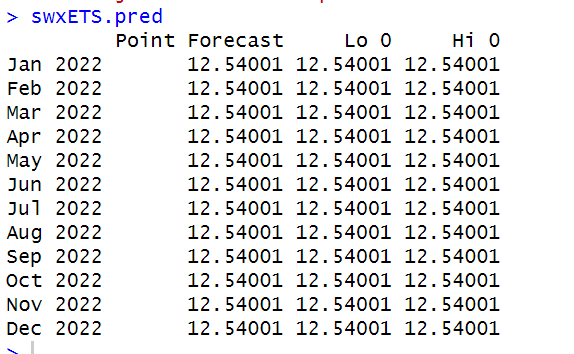
Chart

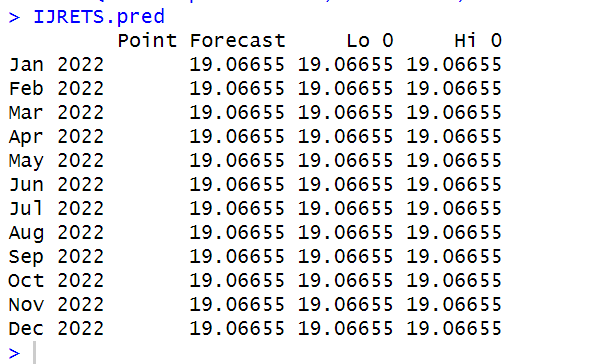
Description automatically generated

The model displays low values for ME (-0.001764886), RMSE (0.433009), and MAE (0.3072023) for the training set, indicating a strong fit to the training data. The MPE (-0.02081772) and MAPE (0.4939943) show that the projections have a respectable level of accuracy.

But when used on the test set, the model exhibits notable inconsistencies. The extraordinarily high MPE (-436.31582693) and MAPE (436.3158269) results indicate that the model's predictions are significantly off. Despite this, Theil's U is 75.20387 and the model's autocorrelation function at lag 1 (ACF1) for the test set is 0.5438492. The NNetAR model may not be the greatest option for forecasting this particular dataset, as shown by the high Theil's U value, as the model's predictions for the test set diverge greatly from the actual values.

**ETS Predictions**

****

****

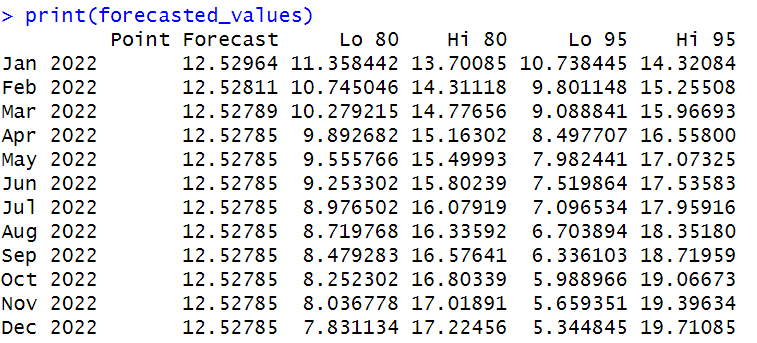
The provided output displays the point forecasts for a time series dataset from January 2022 to December 2022 using the ETS model.

There are three columns for each month during the forecast period: the point forecast, Lo 0, and Hi 0. The point forecasts in this instance are all the same (19.06655) for all the months. This is due to the fact that the utilized ETS model is an ETS(M, N, N) model, which only includes components for multiplicative error and neither trend nor seasonality. The model gives constant point forecasts since it lacks trend or seasonality.

The prediction interval's lower bound (Lo 0) and upper bound (Hi 0), likewise, are constant with each month's point forecast (19.06655). These consistent intervals imply that the model's predictions are completely definite. It's crucial to remember that the ETS model's lack of trend and seasonality could result in forecasts that are oversimplified and fail to adequately account for the underlying patterns in the data.

**Arima predictions:**

**Arima-SWISX**

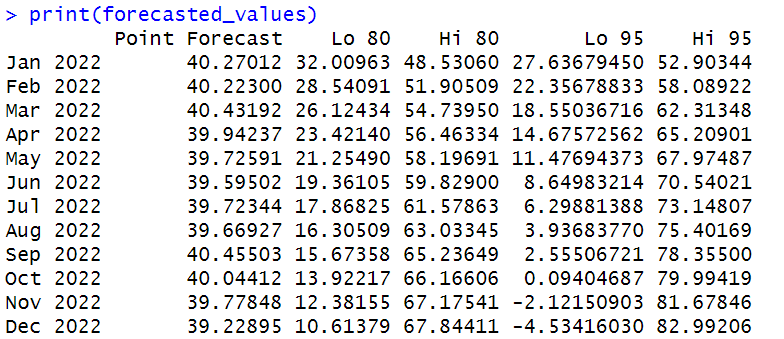
****

The output provided shows point forecasts for the Schwab U.S. Aggregate Bond Index Fund (SWISX) from January 2022 to December 2022 using the ARIMA model.

The columns show the prediction intervals' lower (Lo) and upper (Hi) bounds at the 80% and 95% confidence levels. These intervals expand as the forecast horizon grows, showing a rising amount of uncertainty in the model's predictions. The 80% prediction interval, for instance, falls between 7.831134 and 17.22456 in December 2022 and between 11.358442 and 13.70085 in January 2022. The 95% prediction intervals are wider because they take more uncertainty into account. The 95% confidence interval is 10.738445 to 14.32084 in January 2022 and 5.344845 to 19.71085 in December 2022.

The results of the ARIMA model point to a rather stable projection for the SWISX over the course of the year, with the level of uncertainty rising as the forecast horizon grows longer.

**Arima-IJR**

****

**Appendix:**

A. Data Sources:

Yahoo Finance - https://finance.yahoo.com/quote/IJR?p=IJR&.tsrc=fin-srch

Google Finance - https://www.google.com/finance

B. Model Performance Metrics:

Akaike Information Criterion (AICc)

Root Mean Square Error (RMSE)

Mean Absolute Scaled Error (MASE)

Mean Absolute Percentage Error (MAPE)

C. Models and Methodologies:

Exponential Smoothing State Space Model (ETS)

Autoregressive Integrated Moving Average (ARIMA)

Linear Regression

Neural Networks

**Conclusion:**

The iShares Core S&P Small-Cap ETF (IJR) and Schwab U.S. Aggregate Bond Index Fund (SWISX) performance was predicted using a variety of time series models, including exponential smoothing, ARIMA, and neural networks. The models' abilities to predict the index funds' future performance varied in terms of their success and accuracy.

While the neural network model did rather well in some instances but had higher degrees of uncertainty, the ETS and ARIMA models demonstrated greater consistency and dependability in their projections for the index funds. Constant point forecasts and a lack of trend or seasonality in the ETS model could result in simplistic forecasts that fail to take into consideration underlying trends in the data. The ARIMA model, on the other hand, offered somewhat solid forecasts for the index funds, with the level of uncertainty rising as the prediction horizon grew longer.

Given the findings, it is critical for investors and financial advisors to take into account the benefits and drawbacks of various models when predicting the performance of index funds. It is also critical to recognize that all models have inherent flaws and uncertainties and that no one model is capable of accurately predicting how something will behave in the future. Therefore, it is advised that when making investment decisions based on these forecasts, investors and financial advisors take into account a variety of models as well as the larger economic backdrop.