Performances of the Moving Average Model and the Artificial Neural Network on the Forecast of Stock Market Indices

Peilin Jing, Jiachen Liu, Yaoyuan Zhang, Rashi Lodhi

Abstract

This project delves into the theoretical framework of two types of models. It aims to compare and contrast the performances of the moving average model and the artificial neural network on the prediction of stock market indices. We used the data collected from Yahoo Finance with daily frequency for the period from 1 January 2000 to 31 December 2021. By using a rolling window approach, we evaluated ARIMA-SGARCH to reflect the specific time-series characteristics and have better predictive power than the simple ARIMA model and Recurrent Neural Network models. In order to assess the precision and quality of these models in forecasting, we compared their equity lines, their forecasting error metrics and their performance metrics. The main contribution of this research is to show that the hybrid ARIMA SGARCH model outperforms the other models over the long term.

1. Introduction

Importance of this topic

The importance of this topic can be condensed to 4 points:

- **Interest in the area**: attracted attention of researchers, investors, speculators, and governments
- **ARIMA hybrid over ARIMA**: financial time series often do not follow ARIMA assumptions
- **newest ML techniques to improve models**: We use Recurrent Neural Network(RNN) model to determine if it can reflect the specific time series characteristics and predict better.

2. Literature Review

3. Methodology and Data

3.1 Data Analysis

The first step in the process was cleaning the data. Then, we transformed the adjusted price into a daily logarithmic return, which was calculated according to the following formula:

$$r_t = ln\frac{P_t}{P_{t-1}}$$

Reasons to choose log returns:

- can be added across time periods to create cumulative returns
- easy to convert between log return and simple return
- log return follows normal distribution

Advantages to log return having normal distribution: - Distribution only dependent on mean and sd of sample - forecast with higher accuracy (log return) - Stock prices cannot be normal distribution

3.2.1 Descriptive Satistics - Stock prices

Figure 1 below presents the descriptive statistics of the adjusted closing prices

```
knitr::include_graphics("/Users/rashilodhi/desktop/Descriptive Statistics.png")
```

Statistics <chr>	Value <dbl>
Min Value	676.530029
1st Quantile	1173.805054
Median	1409.119995
3rd Quantile	2125.030029
Max Value	4793.060059
Mean	1772.378729
Skewness	1.400488
Kurtosis	1.457625

Figure 1: Descriptive Statistics for Prices

As seen in Figure 2 there are a few periods, such as 2008, 2011, 2015, and 2018, that show high volatility of returns. We can expect to build more accurate forecasting models if we are able to mitigate and "smooth" such periods.

```r
knitr::include_graphics("/Users/rashilodhi/desktop/PricesPlot.png")
```

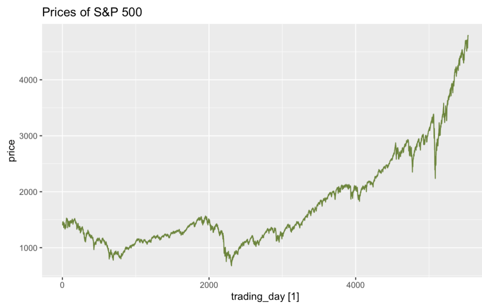

Figure 2: Prices Plot

Stock prices of SP 500 is that it is not normally distributed

3.2.2 Descriptive Satistics - Log Returns

Figure 3 presents the descriptive statistics of the adjusted closing prices

```r
knitr::include_graphics("/Users/rashilodhi/desktop/LogReturnDescriptive.png")
```

Statistics <chr>	Value <dbl>
Min Value	-0.1276521976
1st Quantile	-0.0047060675
Median	0.0006391235
3rd Quantile	0.0058129464
Max Value	0.1095719677
Mean	0.0002148568
Skewness	-0.4004359781
Kurtosis	11.0560760822

Figure 3: Descriptive Statistics for Log Returns

We use log returns to build models

3.3 Methodology

ARIMA (p,d,q)

The ARMA process is the combination of the autoregressive model and moving average [2] designed for a stationary time series. Autoregression (AR) describes a stochastic process, and AR(p) can be denoted as shown below:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t,$$

The moving average process of order q is denoted as MA(q) and the created time series contains a mean of q lagged white noise variables shifting along the series.

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

d is the number of differencing done to the series to achieve stationarity with I (d) so the ARIMA model can be expressed as

$$\text{Expand:} \qquad y_t = c + y_{t-1} + \phi_1 y_{t-1} - \phi_1 y_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

p is the number of autoregressive terms (AR) q is the number of moving average terms (MA)

ARCH(p), GARCH(r,s) and Hybrid ARIMA-SGARCH

The `ARCH(p)` model is given:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i u_{t-i}^2$$

- Most volatility models derive from this - Returns have a conditional distribution (here assumed to be normal) ARCH is not a very good model and almost nobody uses it. - The reason is that it needs to use information from many days before t to calculate volatility on day t. That is, it needs a lot of lags. - The solution is to write it as an ARMA model. - That is, add one component to the equation, $\beta\sigma_{t-1}$.

The `GARCH(p,q)` model is

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

Where: α is news. β is memory. The size of $(\alpha + \beta)$ determines how quickly the predictability of the process dies out.

This leads us to lastly, ARIMA-SGarch

ARIMA SGARCH - Overview

Stock prices can be tremendously volatile during economic growth as well as recessions. When homoskedasticity presumption is violated, it affects the validity or power of statistical tests when using ARIMA models. We consider the SGARCH effect. The error term of the ARIMA model in this process follows SGARCH(1,1) instead of being assumed constant like the ARIMA model.

ARIMA SCGARCH - Steps

1) We conduct a rolling forecast based on an ARIMA-SGARCH model with window size(s) equal to 1000.
2) The optimized combination of p and q which has the lowest AIC is used to predict return for the next point. At the end, the vector of forecasted values has the length of 3530 elements
3) We describe and review our implementation of dynamic ARIMA(p,1,q)-SGARCH(1,1) models with GED distribution and window size(s) equal to 1000.
4) we evaluate the results based on error metrics, performance metrics, and equity curves.

Iteration of the forecasting model ARIMA(p,1,q)-SGARCH(1,1)

```r
knitr::include_graphics("/Users/rashilodhi/desktop/ARIMA_iteration.png")
```

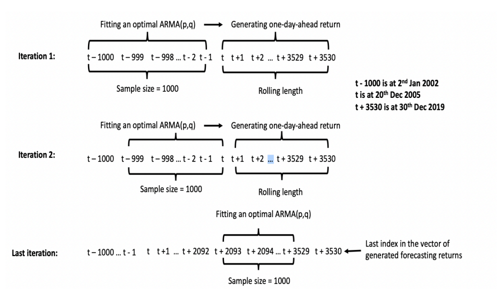

Figure 4: ARIMA iteration

Flowchart of the forecasting model ARIMA(p,1,q)-SGARCH(1,1).

```r
knitr::include_graphics("/Users/rashilodhi/desktop/ARIMA_SGARCH_methodology.png")
```

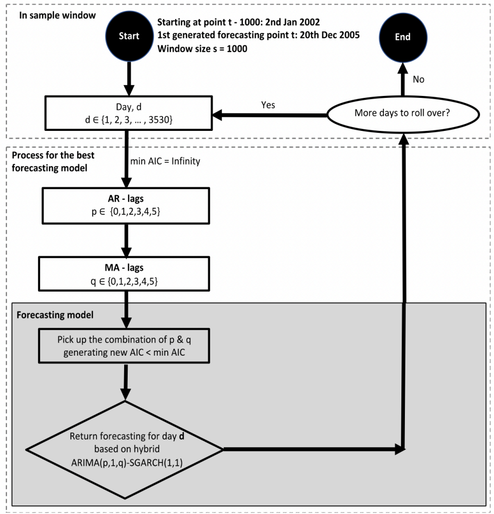

Figure 5: ARIMA_SGARCH methodology

Flowchart of the forecasting model ARIMA(p,1,q)-SGARCH(1,1). This flowchart is for models with window size s = 1000.