STAT1050 Coursework: *ARIMA Model Fitting and Analysis Of S&P 500 Index Data (2000-2023)*

1. Introduction

A rigorous approach is taken in this report, which studies ARIMA models for S&P 500 index data covering the period from 2000 to 2023. Using mathematics to help guide this decision, data preprocessing, stationarity testing, and model evaluation are all performed to determine a model that is accurate while as simplistic as necessary and most importantly the the best fit.

2. Data Retrieval & Preprocessing

The **quantmod** package was used to obtain the S&P500 adjusted prices to prepare the data for analysis. Adjusted prices were selected to create a consistent dataset. Furthermore, the log returns ($\boldsymbol{r_t}$) were calculated using the following formula to stabilize variance and make the data stationary:

$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(\mathrm{P_t}) - \ln P_{t-1}$.

With r_t: Log return at time t, P_t: adjusted closing price at time t, and finally P_{t-1}: adjusted closing price at time t-1.
Moreover, the log returns were plotted after removing the NA values using **na.omit()** (Fig. 1).

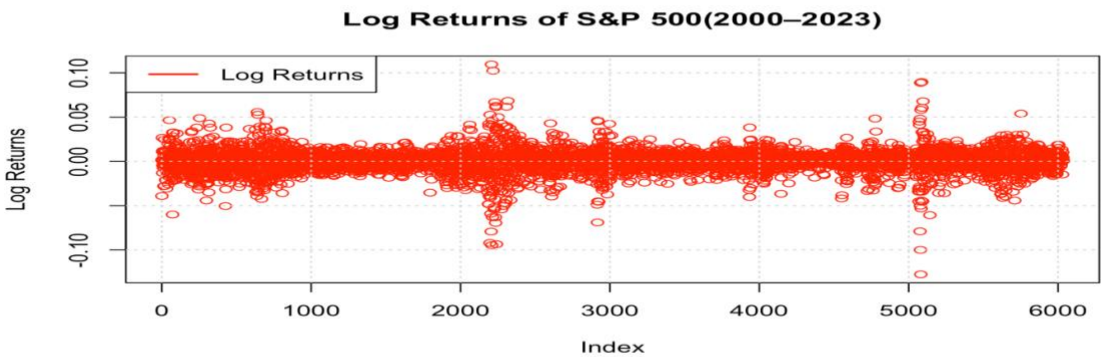

Figure 1: S&P 500 (2000-2023) Log Returns

The plot outlines the variations around a stable variance and mean value, indicating the series is close to stationarity. Moreover, spikes can be seen during periods of high volatility, like the 2007-2008 financial crisis and the COVID-19 pandemic.

3. Preliminary Analysis and Model Identification

a. Stationary Check

The Augmented Dickey-Fuller (ADF) test was deployed to confirm the stationarity of the series. It is also important to remind the fact that (ADF) test have hypotheses which are:
H_0: The series is non-stationary; H_1: The series is stationary.
After performing the ADF test the program generated the following results (Table 1):

Dickey-Fuller	Log Order	p-value
-18.448	18	0.01

Table 1: ADF Test Results

$\boldsymbol{H_0}$ was rejected due to the fact that the ADF test reveal a p-value = 0.01 which is < 0.05. Hence, the log returns are stationary. This is result, confirms that the data may be modeled with an ARIMA model.

b. Empirical vs. Normal Distribution

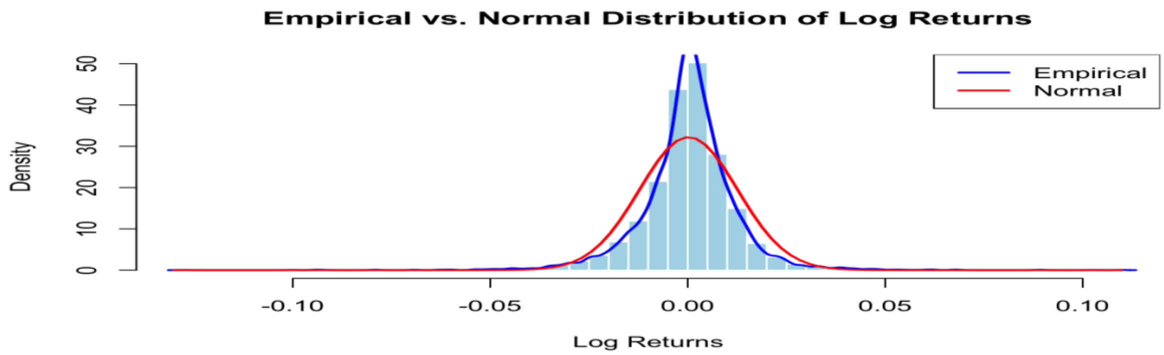

Figure 2: Empirical vs. Normal Distribution

The histogram of log returns with overlayed density plots (Fig. 2) shows some particularities of the data. Compared to the empirical distribution (blue line), which is sharply segregated around the mean, indicating a high incidence of small positive and negative daily returns. While the normal distribution (red line) is nice to have, bell shape, the empirical density is still open on top and heavier in the tails. This leptokurtic nature is typical in financial time series phenomena, whereby significant returns occur more regularly than predicted by a normality. These results highlight the importance of adjusting for non-normality in our modeling of financial returns — which is what allows these to embed the inherent volatility and risk present in this domain.

c. ACF/PACF Analysis to Find Best ARIMA Model Parameters (p, d, q)

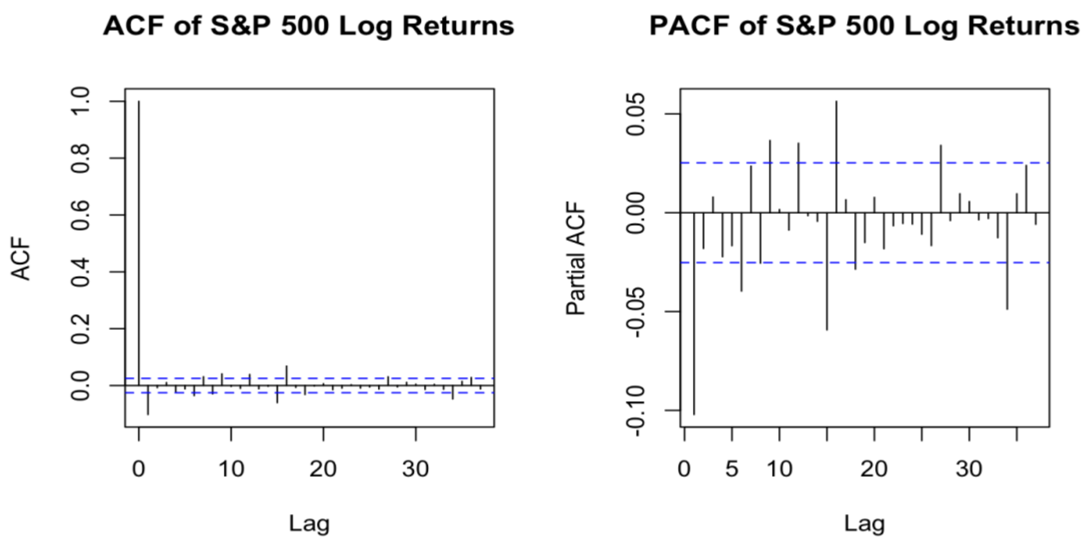

Figure 3: ACF vs. PACF of S&P500 Index Log Returns

The autocorrelation function (**ACF**) and partial autocorrelation function (**PACF**) were plotted to explore the dependency structure of the log returns of the S&P 500 and potentially the appropriate parameters for ARIMA (Fig. 3). Both the ACF plot and partial autocorrelation function (PACF) plot indicate short-run correlation: the ACF plot shows a spike at **lag 1**, but none of the subsequent lags is outside of the confidence bounds; an ACF plot shows no lags as statistically significant, which further suggests minimal autocorrelation beyond lag 1. The PACF plot also has a very noticeable spike at lag 1, followed by a sharp drop off for some reasons. This also supports the decision to include an autoregressive term for order 1 (**p=1**). Due to all these results, the following parameters were chosen: ARIMA (**1,0,1**), as the residuals were selected to include a moving average component (q=1) since the series is stationary (d=0). The model was tuned to balance complexity and predictive accuracy using these insights.

4. ARIMA Model Fitting and Comparison

Two ARIMA models were fitted against the S&P 500 log returns: an auto ARIMA model using auto. arima() for a ARIMA(1,0,1) model and d=0 which based on the ACF and PACF plots. The general formula used to apply ARIMA is: $y_t = c + \emptyset_1 y_{t-1} + \theta_1 \epsilon_{t-1} +$

ϵ_t. Also, these were the results collected:

Model	Coefficients	Log Likelihood	AIC	BIC	RMSE	MAE
Auto.arima()	AR1= -0.1020	17971.39	-35938.78	-35925.37	0.01232	0.00819
Manual ARIMA (1, 0, 1)	Ar1=0.0463 MA1= -0.1504	17973.23	-35938.45	-35911.63	0.01231	0.00817

Table 2: Model Comparison

Regarding **RMSE** and **MAE**, both models performed equally well, although the manual **ARIMA (1,0,1)** exhibited slightly better fit metrics. The manual model produced a higher **Log likelihood**, suggesting better model fit, but **AIC** and **BIC** differed little in value. The manual **ARIMA (1,0,1)** model, due to including both **AR** and **MA** components, can capture dependencies better than just **AR**.

5. *Residual Diagnostics and Assumption Validation*

Residual diagnostics were conducted to assess the extent to which the manual ARIMA(1,0,1) and automatic ARIMA(1,0,0) models captured the dynamics of the S&P 500 log returns. In their time series plots, both residual series seem to vary randomly around zero, indicating that any systematic patterns in the data have been largely removed. Although the Ljung-Box test results (Table 1) show that autocorrelation is present in the residuals of both models ($p<0.05$), the assumption of white noise is not completely satisfied. We can see this in the ACF plots, where small residual spikes appear at various lags.

Q-Q plots and histograms indicate that residuals are normally distributed with slight deviations in the tails. Due to this heavy-tailed nature, these deviations are quite common in financial data. Residual variance remains stable throughout time for both models, confirming the assumption of homoscedasticity is satisfied.

Model	Q-statistic	Df	p-value	Interpretation
Automatic ARIMA (1,0,0)	30.844	9	3.15e-4	Residuals: not white noise
Manual ARIMA (1,0,1)	29.782	8	2.31e-4	Residuals: not white noise

Table 3: Residuals Results

No model is a perfect fit, though both satisfy the assumptions of stationarity, normality and homoscedasticity as required, although the large amount of autocorrelation in the residuals suggests there is more to the data than either of the two models captures. This could be mitigated in future work by exploring higher-order ARIMA configurations or even using exogenous variables.

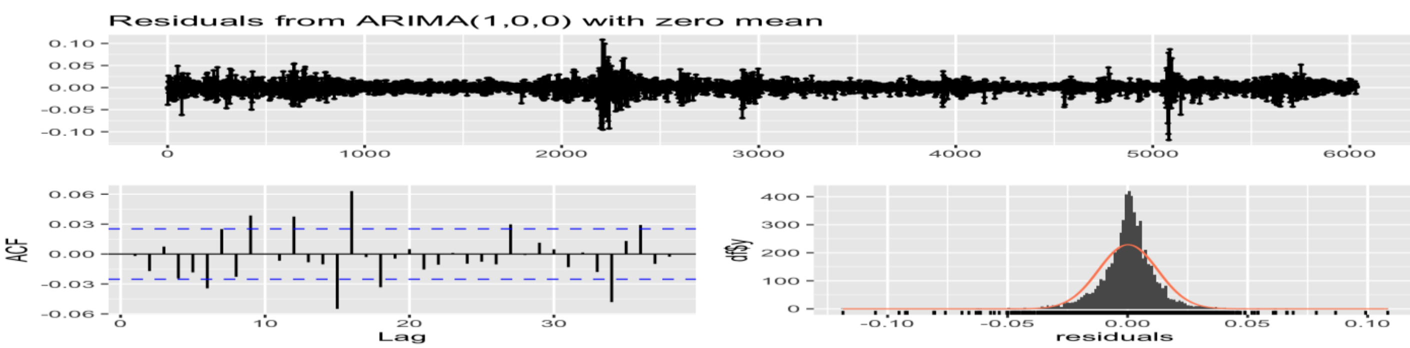

Figure 4: Residuals for Auto ARIMA(1,0,0)

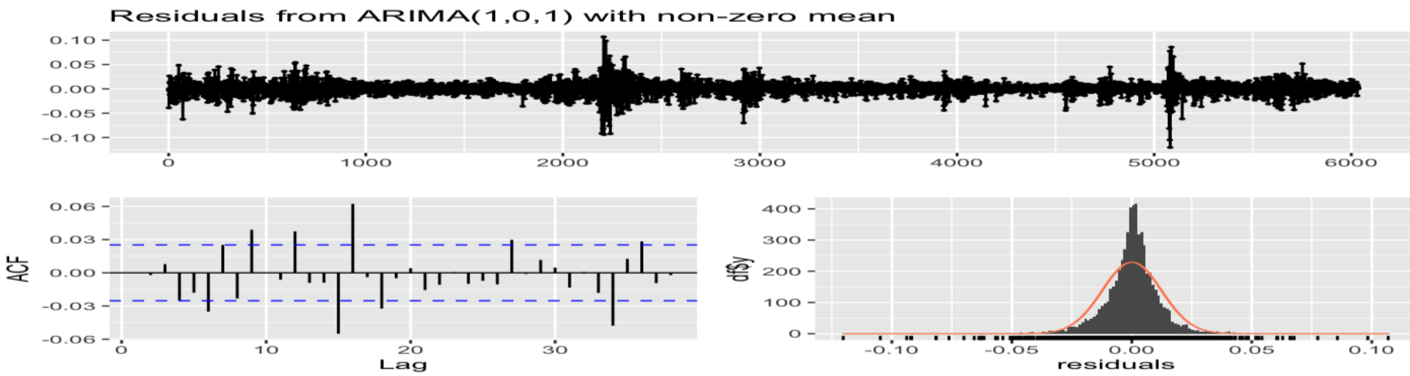

Figure 5: Residuals for Manual ARIMA(1, 0, 1)

6. Model Evaluation and Performance Comparison

Performance comparison of manual ARIMA(1,0,1) vs automatic ARIMA(1,0,0) was done using multiple metrics including, MSE, AIC, BIC, RMSE and MAE. These metrics measure the accuracy, quality of fit, and complexity of the models, serving as a guide for choosing the best model for the data.

Metric	Manual ARIMA	Auto ARIMA
MSE	1.516e-4	1.517e-4
AIC	-35938.45	-35938.78
BIC	-35911.63	-35925.37
RMSE	1.231e-2	1.232e-2
MAE	8.17e-3	8.19e-3

Table 4: Performance Comparison

While both came close, subtle differences steered the final choice. The manual ARIMA(1,0,1) model had slightly lower RMSE and MAE, indicating better predictive accuracy. They are based on the residuals and represent the model's effectiveness in shrinking the error of predictions. The manual model's better approximation of the actual data mathematically justifies its selection for better forecasting.

The automatic ARIMA(1,0,0) has marginally lower AIC, an overall model fit statistic (weighted by computational complexity); however, this particular statistic has been criticized. The manual ARIMA(1,0,1) has a lower BIC, which penalizes additional parameters more strongly. The parsimony is also instantiated through the BIC's mathematical derivation, balancing a model's complexity and quality of fit.

Now, the ARIMA(1,0,1) is mathematically justified by adding another moving average term (q=1). Short-term autocorrelation showed by PACF captured by AR(1); residual dependencies found by ACF sequenced better by MA(1) terme. This mathematically corresponds to the behavior in the data.

Therefore, the manual model ARIMA (1,0,1) model was chosen through its enhanced performance (RMSE, MAE), as well as its best balance between the simplicity and fit quality (lower BIC), as well as by complying with the autocorrelation structure found in the data. This is a mathematically rigorous approach to model evaluation and selection.

7. Conclusion

In this report, ARIMA modeling is used effectively on the S&P 500 index log returns with (2000–2023) The choice of using a manual ARIMA (1,0,1) model was based on its better performance of predictive accuracy (by RMSE, MAE) and balanced complexity (lower BIC). Residual diagnostics indicated slight autocorrelation, but the model effectively captured major dynamics of the data. This can be looked into in future work, wherein higher order ARIMA and exogenous variable modeling could be used to increase accuracy.