Technical Report: Forecasting Economic Indicators and FTSE 100 Index

1. Exponential Smoothing

1.1 Data Preparation:

In the initial stage, the datasets were acquired through the provided links to access the ONS information, which encompassed the Average weekly earnings data set, retail sales time series data set, index of production data set, and turnover and orders in the production and services industries. Relevant columns from each dataset were identified, containing essential time series data required for predictive purposes. To streamline the analysis and prediction process, the information was converted into a date format wherein each data point corresponds to a particular time frame (such as a month) and is linked to a numerical value.

1.2 Preliminary Analysis:

The initial phase of the preliminary analysis involved the visual and quantitative examination of the dataset. This encompassed the generation of time series plots, Decomposition, Scatter plot, and ACF to visualize the data over time and detect any trends, seasonality, or anomalies. Time series plots were scrutinized to pinpoint long-term trends and periodic patterns. Trend analysis identified overall upward or downward trends, whereas seasonality analysis uncovered recurring patterns at specific intervals. Seasonal decomposition methods were utilized to segregate the time series into trend, seasonal, and residual elements. Correlation matrices were constructed for each dataset to explore the relationships between variables and unveil potential patterns or interactions within each dataset. Autocorrelation plots were employed to spot patterns between previous and subsequent values in a time series, aiding in the selection of appropriate forecasting models.

1.3 Exponential Smoothing Models:

1.3.1 K54D: Monthly average of private sector weekly pay.

Upon initial inspection, we identified an upward trend and seasonal patterns in the time series graphs, with cycles recurring at regular intervals. The ACF confirmed seasonal patterns every 12 months. Given the presence of trend and seasonality, we opted for the Holt-Winters' model for exponential smoothing. We applied this technique to forecast average weekly earnings data, considering both additive and multiplicative seasonality. Two models were developed: one with additive seasonality (fit1) and the other with multiplicative seasonality (fit2). Forecasts were made for the next 12 periods for each model, and a comparison of fitted values and forecasts was conducted as shown in fig(1.3.1). After calculating RMSE values for both types of forecasting, we determined that using the multiplicative forecasted values for January 2024 to December 2024 was more accurate, as they had a lower RMSE compared to the additive forecasts.

1.3.2 EAFV: Retail sales index, household goods, all businesses:

After analyzing the retail sales index data for household goods, it was observed that there were clear upward trends and seasonal variations in the time series plots. The ACF confirmed the presence of seasonal patterns repeating every 12 months. Therefore, the Holt-Winters' model was chosen as the preferred exponential smoothing technique to account for both trend and seasonality. Two Holt-Winters models were developed: one with additive seasonality (fit1) and the other with multiplicative seasonality (fit2). Forecasts were generated for the next 12 periods using both models as shown in fig(1.3.2). By calculating the RMSE values for the additive and multiplicative forecasts, it was determined that the multiplicative forecasted values for January 2024 to December 2024 were more accurate, as they had a lower RMSE compared to the additive forecasts.

1.3.3 K226: Extraction of crude petroleum and natural gas.

After analyzing the data on crude petroleum and natural gas extraction (K226), we have noticed a consistent downward trend without any noticeable seasonal patterns. Therefore, we have chosen Holt's Linear Exponential Smoothing (LES) Model as our preferred method for exponential smoothing. This model uses an additive trend and applies Holt's linear exponential smoothing (fit_linear) to generate projections for the next 12 periods as shown in fig(1.3.3). To assess the accuracy of the LES forecasts, we have calculated the Root Mean Square Error (RMSE) value. Based on these results, we will be using the Linear Exponential Smoothing forecasted values until December 2024 for our forecasting purposes.

1.3.4 JQ2J: The manufacturing and business sector of Great Britain, total turnover and orders

After analyzing the manufacturing and business sector in Great Britain, we found upward trends and seasonal patterns. The patterns repeat every 6 months, as indicated by the autocorrelation function. To account for both trend and seasonality, we used Holt-Winters' model. We trained two models, one with additive seasonality and one with multiplicative seasonality. We forecasted the next 12 periods using each model as shown in fig(1.3.4). After calculating the RMSE values, we decided to use the additive forecasted values from January 2024 to December 2024 due to their lower RMSE compared to the multiplicative forecasts.

1.4 Graphical Illustrations:

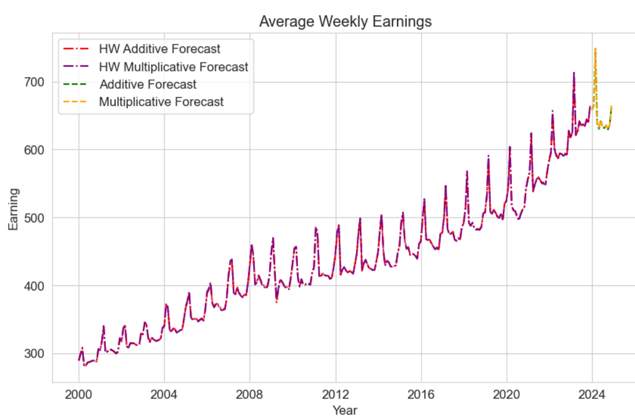

Fig:1.3.1 K54D

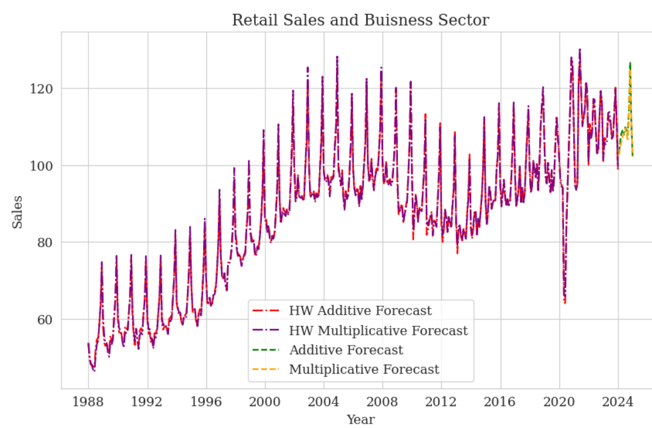

Fig:1.3.2 EAFV

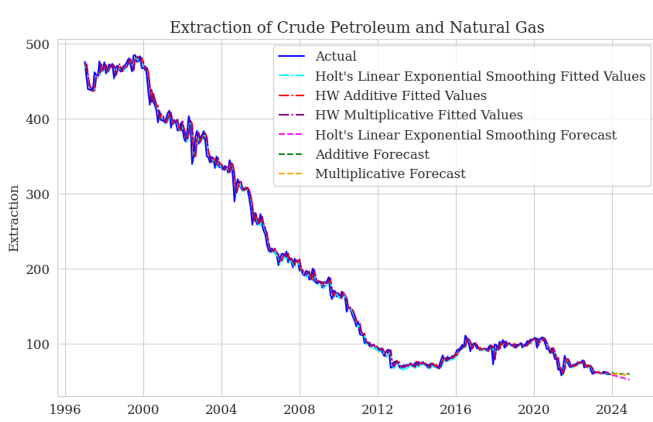

Fig:1.3.3 K226

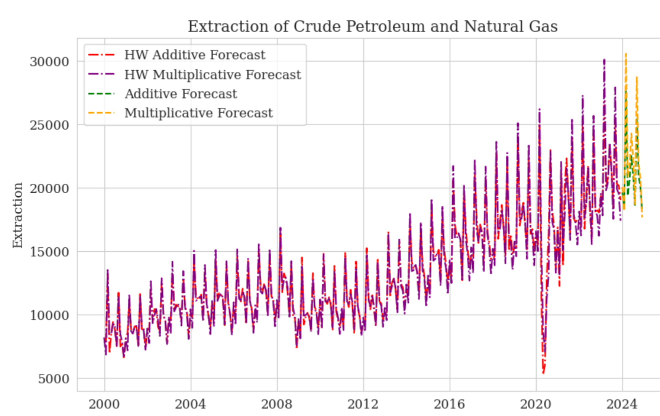

Fig: 1.3.4 JQ2J

2. ARIMA Forecasting

2.1 Preliminary Analysis:

In the initial examination, ACF and PACF time series analysis techniques are used to understand the correlation structure in the dataset, helping determine optimal parameters for models like ARIMA and SARIMA. The ACF Plot shows patterns at seasonal lags, indicating a SARIMA model may be appropriate.

2.2 SARIMA Model:

The SARIMA (Seasonal AutoRegressive Integrated Moving Average) model is an extension of the ARIMA model, Includes seasonality and is optimized using auto_arima from pmdarima. The identified parameters are used to fit a SARIMAX model to the data, capturing both seasonal and non-seasonal components. This model generates forecasts for the next 12 months, with accuracy assessed using RMSE against actual values(fig 2.2).

2.3 Comparison and Graphical Illustrations

The SARIMA forecasts were evaluated against the forecasts generated by exponential smoothing

- SARIMA RMSE: 35.17
- Exponential Smoothing RMSE (Multiplicative Model): 8.77

A lower RMSE means higher accuracy, indicating that the predicted values are closer to the actual values. Therefore, in this scenario, exponential smoothing with multiplicative seasonality has a significantly lower RMSE than SARIMA.

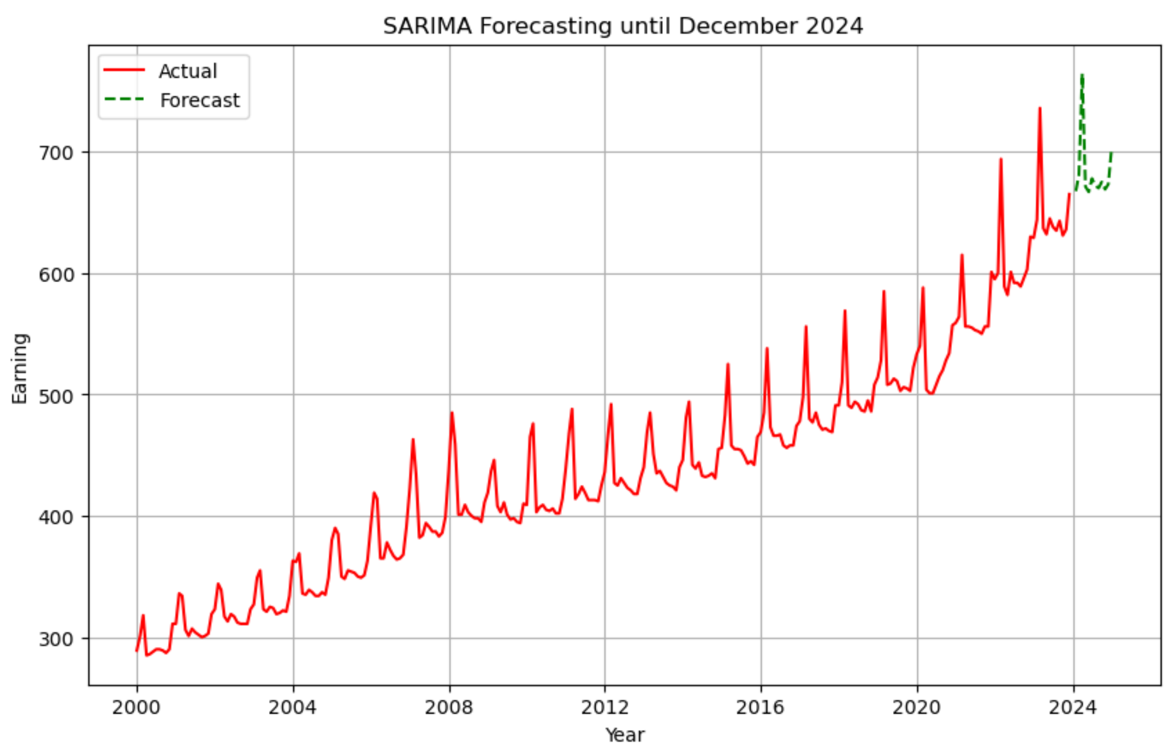

Fig 2.2 SARIMA Model for K54D

3. Regression Prediction

3.1 Data Preparation

The UK Footsie 100 share index dataset, FTSE, was used for regression analysis. A new dataset called "FTSEdata_34812598.xls" was created, including four time series variables (K54D, EAFV, K226, and JQ2J) as predictors in a multivariate regression model to predict the "Open" column in the FTSE dataset. Outliers were removed from the EAFV and K226 datasets, and the analysis focused on data from January 2000 onwards to align with the other datasets. This approach aimed to maintain consistency in the forecasting process while preserving the model's trend and seasonality. Attempts to replace outliers with mean or median values were unsuccessful as they disrupted the overall trend. Therefore, outliers were completely excluded from the analysis to ensure the integrity of the forecasting process.

3.2 Regression Model

The FTSE dataset was used for regression analysis. A new dataset called "FTSEdata_34812598.xls" was created, including four variables as predictors in a multivariate regression model to predict the "Open" column in the FTSE dataset. Outliers were removed from two of the variables, and the analysis focused on data from January 2000 onwards. This approach aimed to maintain consistency in the forecasting process while preserving the model's trend and seasonality. Attempts to replace outliers with mean or median values were unsuccessful, so outliers were completely excluded from the analysis to ensure the integrity of the forecasting process.

3.3 Graphical Illustrations and Results

The model generates graphs showing predictions for multiple variables using various techniques(fig 3.3). It also includes a merged 'Open' regression forecast using OLS regression coefficients. The projected values for the Open Column are displayed until December 2024, with an RMSE of 258.85.

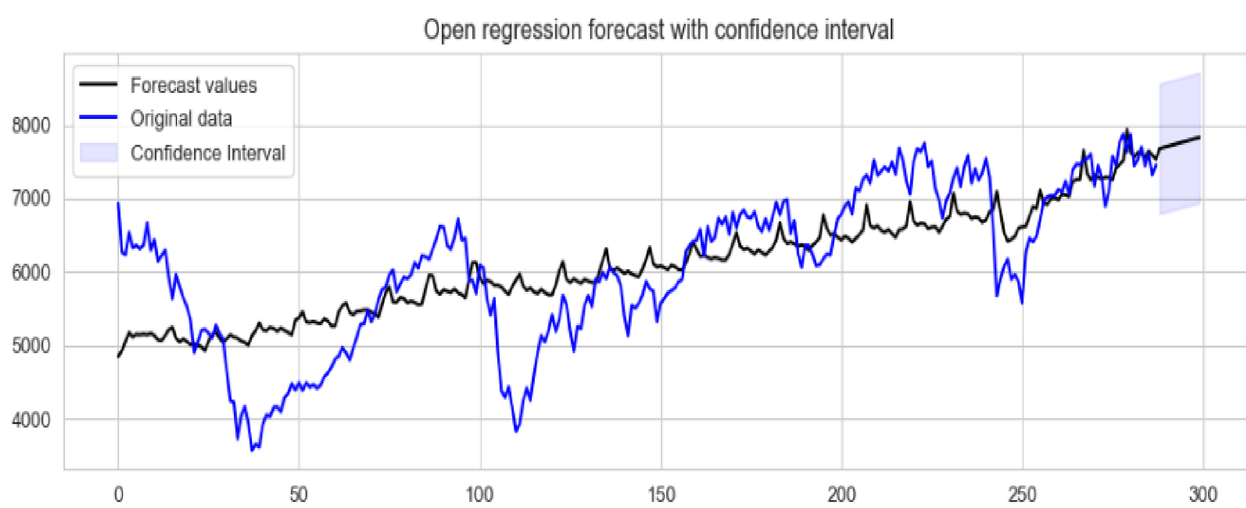

Fig 3.3 Regression Model

4. Appendix

4.1 Exponential Smoothing: The common description of the provided Python files, namely ExponentialSmoothing_K54D_34812598.py,ExponentialSmoothing_EAFV_34812598.py, ExponentialSmoothing_K226_34812598.py, and ExponentialSmoothing_JQ2J_34812598.py, is as follows:

- The code imports required libraries such as pandas, matplotlib.pyplot, numpy, seaborn, and functions from statsmodels for time series analysis.
- Time series data from an Excel file is loaded into a pandas DataFrame, with the 'Date' column transformed to datetime format and used as the index.
- Trend Estimator code creates a plot that displays the original time series data, as well as seasonal and trend components.
- The time series data is divided into trend, seasonal, and residual components using the statsmodels library's seasonal decomposition, and each component is shown individually.
- Calculates the correlation matrix for the dataset and displays it as a heatmap with seaborn. The scatter plot and autocorrelation graph of time series data help us find noteworthy trends and seasonality.
- For forecasting, the code applies the Holt-Winters method to the time series data from K54D.py, EAFV.py, and JQ2J.py, considering both additive and multiplicative seasonality. However, for one dataset, it uses Holt's Linear Exponential Smoothing approach. The code generates forecasts for the next 12 periods and plots the fitted values and forecasts for both models, allowing for visual comparison.
- The code computes the Root Mean Squared Error (RMSE) for both additive and multiplicative Holt-Winters models, as well as Holt's Linear Exponential Smoothing model.
- Lastly, the model prints the predicted values for each month and creates forecasts for the period of January 2024 to December 2024.

4.2 ARIMA: The Python file ARIMA_SARIMA_34812598.py contains the code description provided below.

- The Python code imports a time series dataset from an Excel file and creates visual representations of the ACF and PACF to identify parameters for a SARIMA model.
- The Python code utilizes the auto_arima function from pmdarima to determine optimal parameters for a SARIMA model. These parameters are applied to fit a SARIMAX model and produce forecasts for the upcoming 12 months. The original data is displayed alongside the forecasted values up to December 2024.
- The code snippet retrieves predicted values for January to December 2024 and calculates the RMSE for these forecasts.

4.3 Regression: The Python file REGRESSION_FTSE_34812598.py contains the code description provided below.

- The code imports data from an Excel file with two sheets ('Sheet1' and 'Sheet2'). It uses Holt's Linear Model Forecasting and Linear Exponential Smoothing (LES) to calculate fitted and forecast values. These values are then used for further analysis.
- The code snippet performs OLS regression analysis, extracts coefficients, combines predicted and forecasted values, calculates forecast error, and produces subplots showing forecasts for individual variables with confidence intervals.
- The code snippet shows the predicted values for the 'Open' variable and calculates the Root Mean Squared Error (RMSE) for these predictions.