A Comprehensive Technical Report on Time Series Forecasting and Regression Analysis

1. Exponential smoothing

1.1 Data Preparation. Identification of Key Variables: The necessary data was extracted from each dataset by identifying key variables such as Average weekly earnings (K54D column from the average weekly earnings dataset), Retail sales index for household goods (EAFV column from the retail sales time series dataset), Extraction of crude petroleum and natural gas (K226 column from the index of production dataset), and Total turnover and orders for the manufacturing and business sector (JQ2J column from the turnover and orders dataset).

1.2 Preliminary Analysis

1.2.1 Trend Estimator:

Upon graphing the initial time series data for 'Earning', 'Production', 'Retail Sales', and 'Turnover' alongside the calculated moving averages (MA7 for seasonality and MA2x12 for trend), various observations were made.

'Earning' data exhibited seasonal variations and a rising trend.
'Production' did not show distinct seasonality but displayed a declining trend.
'Retail Sales' closely resembled 'Earning' in terms of seasonality and an upward trend. 'Turnover' demonstrated seasonal fluctuations and an upward trend.

1.2.2 Autocorrelation Function (ACF)

An Autocorrelation Function (ACF) plot was generated for each dataset to examine the autocorrelation patterns within the time series data: -
'Earning': A spike was observed after every 12 lags, indicating a seasonal pattern in earnings.
'Production': No significant spikes were observed, suggesting a lack of autocorrelation and no clear seasonal pattern.
'Retail Sales': A spike was evident after 12 lags, indicating a seasonal pattern in retail sales.
'Turnover': A spike appeared after 6 lags, indicating a seasonal pattern in turnover activities.

1.2.3 Decomposition:

Through seasonal decomposition with the 'additive' model, the time series data is decomposed into its trend, seasonal, and residual components. This analytical approach facilitates the identification of underlying patterns and components present in the time series data, including long-term trends, seasonal fluctuations, and irregular variations or residuals.

1.2.4 Correlation

The correlation matrix is a visual representation of correlation coefficients among variables, with each cell indicating the correlation coefficient between two specific variables. This coefficient quantifies the magnitude and direction of a linear association between the two variables.

1.3 Exponential smoothing.

1.3.1 Monthly average of private sector weekly pay. (Earning)

Upon completion of the initial analysis, it was noted that seasonality was present, leading to the application of Holt-Winter exponential smoothing. The additive method is recommended for series with relatively stable seasonal patterns, while the multiplicative method is more appropriate for series where seasonal variations change in relation to the series' level.

The RMSE for the additive model is 9.027, while for the multiplicative model it is 8.771. The slightly lower RMSE of the multiplicative model indicates that it fits the data better than the additive model. Therefore, in this instance, the multiplicative model with seasonality is the preferred option.

1.3.2 Retail sales index, household goods, all businesses.(Retail sales index)

Following an initial analysis, I identified seasonality and decided to implement Holt-Winter exponential smoothing.

The RMSE for the additive model was 4.75, whereas the multiplicative model yielded an RMSE of 4.74. The marginally lower RMSE of the multiplicative model suggests that it offers a more accurate representation of the data than the additive model. Hence, the multiplicative model with seasonality is the recommended approach in this instance.

1.3.3 Extraction of crude petroleum and natural gas.(Production)

Upon examining the graphs, it is clear that there is an absence of seasonality. Therefore, I decided to apply Holt-Winter exponential smoothing.

Both the Holt-Winters and Holt Linear exponential smoothing techniques were applied to the data. Upon evaluating their respective RMSE values, it is evident that the Holt Linear method exhibited a slightly lower RMSE of 8.78, while the Holt-Winters method yielded an RMSE of 8.85. This suggests that, in this specific instance, the Holt Linear method showcased a slightly superior forecasting accuracy

1.3.4 The manufacturing and business sector of Great Britain, total turnover and orders.(Turnover)

Upon conducting an initial analysis, I identified seasonality in the data, prompting me to employ Holt-Winter exponential smoothing.

The RMSE for the additive model is 1334.74, whereas for the multiplicative model it is 1424.81. With a slightly lower RMSE, the additive model demonstrates a better fit to the data compared to the multiplicative model. Therefore, in this instance, the additive model with seasonality is the recommended approach.

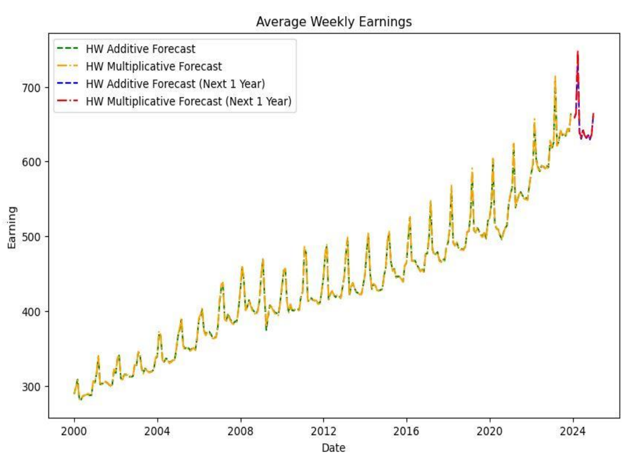

Fig 1.1 Average Weekly Earnings

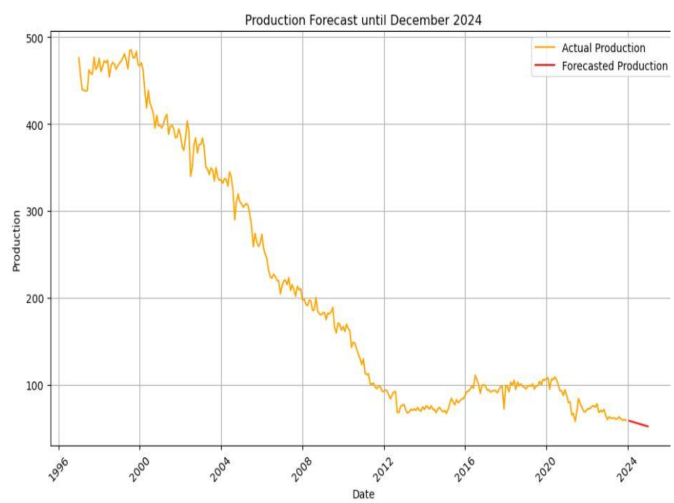

Fig 1.2 Production Forecast until Dec 2024

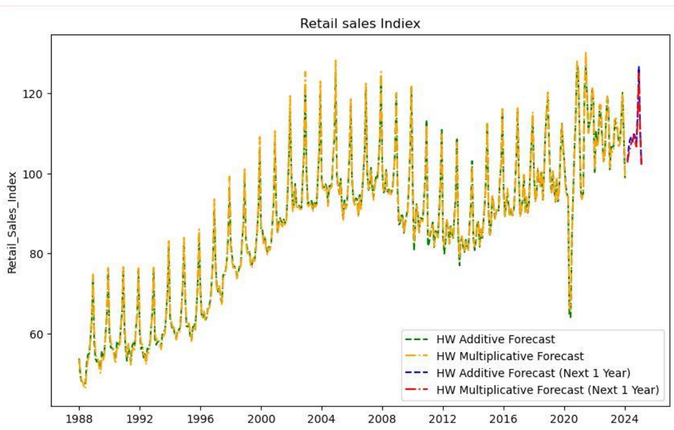

Fig 1.3 Retail sales Index

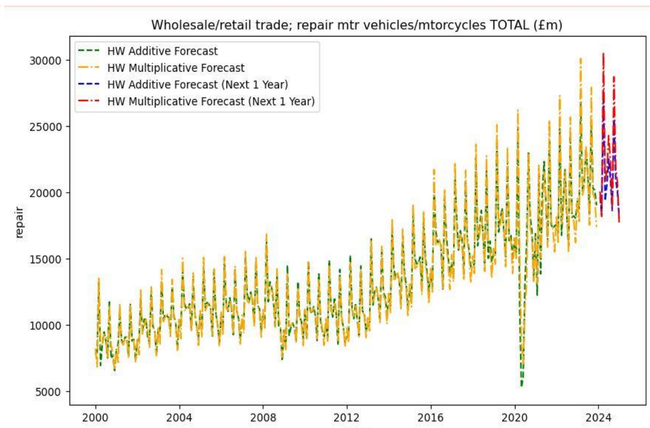

Fig 1.4 Wholesale/Retail trade

2. ARIMA

2.1PreliminaryAnalysis:

ACF and PACF plots are commonly used to analyze the correlation structure of time series data. These plots play a crucial role in determining the appropriate parameters for the SARIMA model, including the order of differencing (d), autoregressive (p), and moving average (q) parameters. By examining the ACF and PACF plots, significant spikes at specific lags can provide valuable insights into potential values for the p and q parameters, respectively. These plots serve as valuable tools in the process of parameter selection for SARIMA modeling.

2.2Exponential Smoothing Model Selection:

Section 1.2.1 provides evidence of seasonality within the data. To accommodate this recurring pattern, the SARIMA (Seasonal Autoregressive Integrated Moving Average) model is utilized.

2.3Model Identification:

The identification of the order of differencing (d), autoregressive (p), and moving average (q) parameters was conducted by analyzing the ACF and PACF plots. Additionally, seasonal patterns were detected to determine the seasonal order (P, D, Q, S).

To automatically select the optimal parameters for the ARIMA model, the 'auto_arima' function from the 'pmdarima' library was utilized. This selection was based on the Akaike Information Criterion (AIC), which is a widely used criterion for model selection.
Best model parameters (p, d, q): (0, 1, 2)
Best seasonal parameters (P, D, Q, S): (0, 1, 2, 12)

2.4 Compare ARIMA and Exponential Smoothing

2.4.1 ARIMA models are utilized to identify and capture linear relationships present in time series data. Conversely, exponential smoothing models are specifically designed to capture exponential trends and seasonality, making them more suitable for different types of data patterns. Furthermore, ARIMA models require more extensive parameter tuning, whereas exponential smoothing models offer a simpler implementation process and often produce accurate forecasts with fewer parameters.

2.4.2 In terms of forecasting accuracy for the given dataset, the SARIMA and Holt-Winters exponential smoothing models produce RMSE values of 35.16 and 8.77, respectively. The substantial difference in these RMSE values clearly indicates that the Holt-Winters exponential smoothing model exhibits superior forecasting accuracy compared to the SARIMA model.

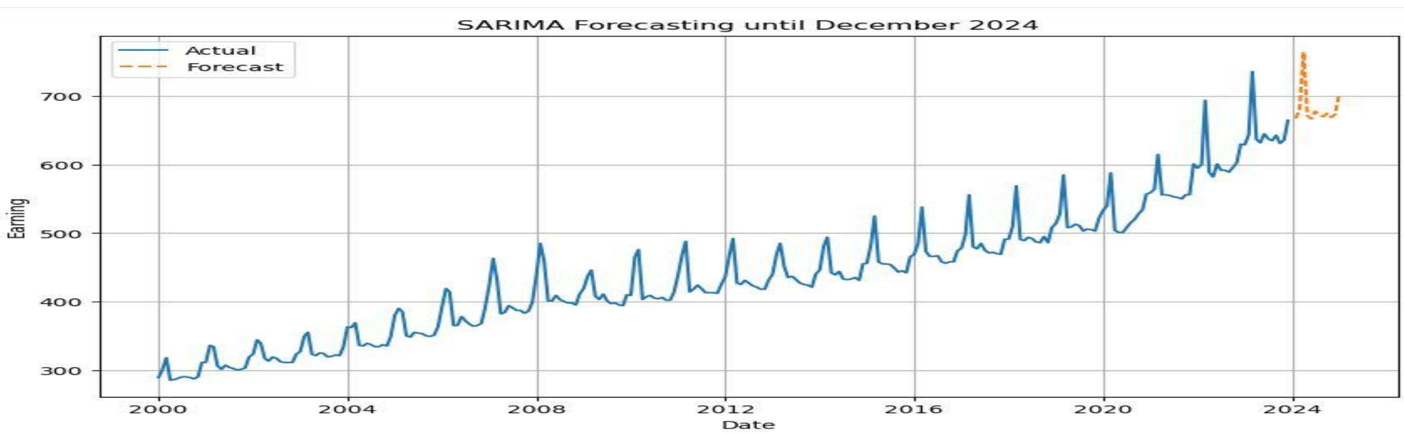

Fig 2.1 The forecasted values until December 2024 are illustrated in the figure through the utilization of the ARIMA model.

3.Regression Forecasting Analysis

3.1 Aim:

To forecast the Open stock prices, a regression model is utilized, wherein earnings (Earning), retail sales (Retail), production (Production), and turnover (Turnover) are considered as the independent variables.

3.2 Data Preparation:

An Excel spreadsheet has been formulated, comprising of two distinct sheets. The initial sheet, labeled as "Sheet1", encompasses a broad array of data including Open stock prices, earnings, retail sales, production, and turnover. In contrast, the second sheet, titled "Sheet2", exclusively holds details concerning Open stock prices. The date range has been standardized from January 2000 to December 2023 to guarantee uniformity and aid in accurate forecasting. This standardized timeframe promotes consistency and aids in precise forecasting. The refined and standardized dataset is subsequently employed to forecast the variable of interest for the upcoming 12 months.

3.3Holt's Linear Model Forecasting:

Separate Holt's Linear Models are initialized and fitted for every variable, facilitating individual predictions.

3.4Post-processing Tasks:

3.4.1 Deriving regression coefficients from an Ordinary Least Squares (OLS) regression model. 3.4.2 Determining fitted values for Open by utilizing regression coefficients and Holt's linear model forecasts.
3.4.3 Estimating forecast values for Open using regression coefficients and Holt's linear model.
3.4.3 Integrating fitted and forecasted values, assessing forecasting errors, and establishing confidence intervals.

3.5 Regression Analysis:

Regression coefficients are derived from an ordinary least squares (OLS) regression model in order to depict the associations between variables. By utilizing these coefficients and Holt's linear model forecasts, the fitted and projected values can be calculated. The accuracy of the forecasts is assessed by evaluating the forecasting errors, and confidence intervals are established to provide a measure of uncertainty.

3.6 Accuracy

The Root Mean Squared Error (RMSE) calculated for the estimated values is 262.09

3.7Plotting:

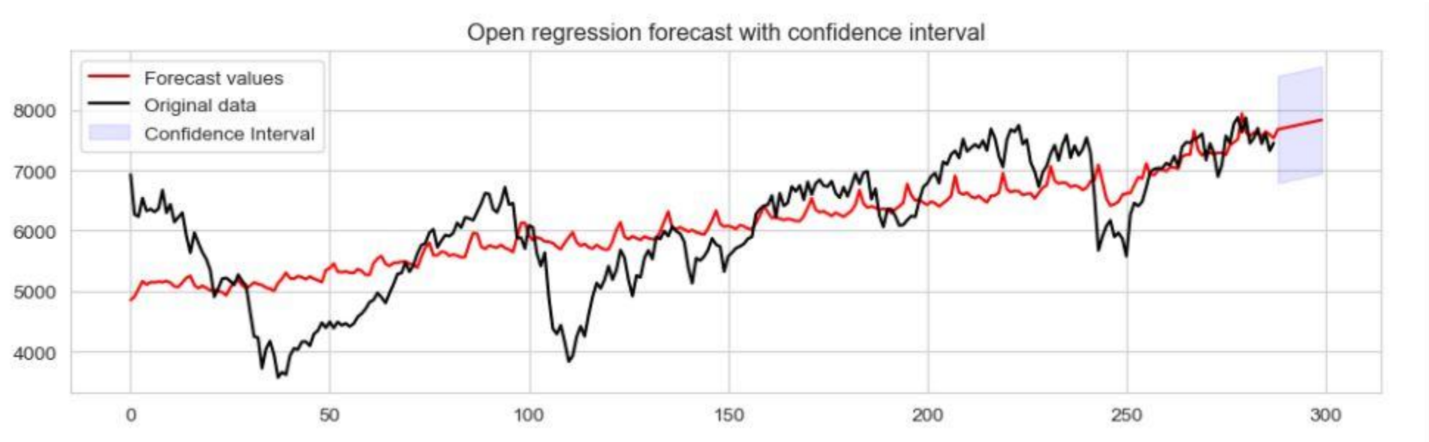

Fig 3.1 Regression Forecast and Confidence Interval for Open Variable

4. Appendix

4.1 For Exponential forecasting

1. The process of exponential forecasting begins with an initial examination of the time series data, which involves analyzing trends, calculating monthly and yearly averages, and decomposing the time series into trend, seasonality, and residual components. Additionally, correlation analysis is performed to understand the relationships between variables, and autocorrelation patterns are identified through the autocorrelation function (ACF).

Subsequent to the initial analysis, the Holt-Winters exponential smoothing models are applied to the data, incorporating both additive and multiplicative smoothing techniques.

Following the fitting of the Holt-Winters models to the data, their performance is evaluated using the root mean squared error (RMSE) metric. This metric helps assess how well the models can capture data variation and generate accurate forecasts. Ultimately, forecasts are generated using either the multiplicative or additive Holt-Winters model, based on the model that produces a lower RMSE for future time points, typically up to December 2024. These forecasts provide valuable insights into future data trends.

4.2 For ARIMA

In this section, the application of the ARIMA model is demonstrated for time series forecasting. The analysis begins by visually examining and summarizing the original data to understand its patterns and variability. To select appropriate parameters for the SARIMA model, ACF and PACF plots are then utilized.

To simplify the process of selecting optimal SARIMA model parameters, the script utilizes the auto_arima function. This function helps in determining the most suitable values for the model's orders and seasonal orders. Once the SARIMAX model is fitted to the data, forecasts are generated for future time points, projecting until December 2024.

The effectiveness of the SARIMAX model is evaluated using the root mean squared error (RMSE) metric. This metric provides insights into how accurately the model can forecast future values based on historical data.

4.3 For regression analysis

In this section, we apply regression analysis to predict financial metrics by utilizing Holt's Linear Model and Ordinary Least Squares (OLS) regression. The process begins with data preparation, involving the extraction of financial indicators from an Excel file and verifying their suitability for analysis.

Subsequently, Holt's Linear Models are used to forecast each specific metric of interest. These models help in identifying trends and seasonality in the data, offering valuable insights into future patterns. Additionally, OLS regression is employed for further analysis to extract coefficients and compute fitted and forecast values.

For better understanding of the predictions, visualization techniques are utilized with matplotlib, enabling a clear interpretation of the outcomes. The section ends by displaying the anticipated values for a particular variable, facilitating easy analysis and decision-making.