Time Series Analysis on Electric Vehicles Adaption in the UK

Thesis submitted in partial fulfilment of the requirements
for the award of the degree of

Master of Science

by

Iftekhirul

(ID No. 2777638)

Under the Guidance of

Dr Waddah Saeed

Data Analtics

School of Computer Science and Informatics

De Montfort University

Leicester, UK

August 30, 2024

Contents

List of Figures

List of Tables

Chapter 1

Background

1.1 Introduction

The transportation sector has a substantial impact on the emission of greenhouse gases, emphasizing the need of electrifying vehicles to achieve net-zero objectives. The UK government's ambitious goal to ban the sale of new petrol and diesel cars by 2035 underscores the urgency for widespread electric vehicle (EV) adoption(1). The growing global concern about climate change, as evidenced by the findings of the Intergovernmental Panel on Climate Change (IPCC), has justified the urgent need to shift towards more ecologically sustainable forms of transportation(2).

Gaining insight into the present level of electric vehicle (EV) adoption in the United Kingdom is crucial. Data from the Society of Motor Manufacturers and Traders (SMMT) provides insights into the market share of EVs compared to traditional vehicles (3). Moreover, the Office for Zero Emission Vehicles (OZEV) website provides comprehensive information on government incentives for electric vehicle (EV) acquisition, including grants and tax breaks, which greatly contribute to the promotion of EV adoption ..

Accurately studying future EV demand is crucial for various stakeholders. Manufacturers can utilize this information for production planning and resource allocation (4). Governments can leverage demand forecasts for infrastructure development and policy design. Similarly, energy providers can use these predictions to plan grid capacity and electricity generation (5).

Identifying patterns and forecasting results are crucial functions of time series analysis. The objective of this thesis is to ascertain the prevailing trend of electric vehicle (EV) adoption in the United Kingdom and examine the patterns observed over several years. One possible data source that this project will investigate is government vehicle registration data obtained from the Department for Transport (6). The synthesis will be described using the SEMMA methodology.

1.2 Problem Statement

The present study investigates the present status of electric vehicle (EV) technology adoption in the United Kingdom. To do this, our objective is to:

1. Examining the temporal patterns in Electric vehicles throughout time using the Vehicle Licencing Statistics: 2023 dataset from the Department for Transport. This analysis will measure the rate of electric vehicle (EV) uptake in recent years.

2. Discovering the insights pertaining to the various categories of electric vehicles.

3. Creating a data-driven model to predict electric vehicle (EV) demand in the United Kingdom by utilising time series methodologies.

This thesis aims to provide relevant insights for policymakers, industry stakeholders, and infrastructure developers by addressing these objectives. This knowledge will be vital for enabling a seamless shift towards a more environmentally friendly transportation system in the UK, where the country's ambitious net-zero targets need extensive electric vehicle (EV) adoption.

1.3 AIM

The aim of this thesis is to extensively analyse the present condition and future prospects of electric vehicle (EV) adoption in the UK using time series analytic methods. This objective will be accomplished by:

1. Measuring the rate of increase in various forms of electric vehicle (EV) adoption in recent years.
2. Utilise time series methodologies such as autocorrelation and partial autocorrelation, as well as seasonal decomposition, to gain a deeper understanding of the trend in electric vehicles (EVs).
3. Constructing a data-centric model to forecast electric vehicle (EV) demand in the United Kingdom.

1.4 Research Question

1. How much has the electric vehicle (EV) market expanded in the UK in recent years?

2. How do time series methods such as autocorrelation, partial autocorrelation, and seasonal decomposition reveal the underlying trends and patterns in the adoption and usage of electric vehicles (EVs) over time?

3. Can we construct a dependable model to forecast electric car demand in the UK by utilising time series methodologies?

1.5 OBJECTIVE

Objective 1: Analyze Trends in EV Adoption

1. Evaluate the rate of electric vehicle (EV) uptake in the UK in relation to conventional fuel vehicles.

Objective 2: Identify underlying trends of EVs

1. Explore the patterns of many categories of electric vehicles using time series methodologies.

Objective 3: Build Forecast Model

1. Develop data-driven models to forecast future demand for EVs.

Chapter 2

Literature Review

2.1 Introduction

We discovered extremely few scholarly articles or research papers pertaining to electric vehicles. The majority of the literature was unrelated to analytical research. I gathered these with the highest percentage of similarities.

2.2 Literature Review

The sporadic and stochastic character of electric vehicle consumption makes energy demand prediction difficult, according to Fatemeh Marzbnai et al. Accurate forecasting is also crucial for grid stability. Linear, nonlinear, and hybrid prediction methods are examined (7). Constance Crozier et al. presented cluster analysis of conventional vehicle travel survey data to construct representative electric vehicle usage profiles in 2018. They proposed a model to map journey and vehicle parameters to energy consumption, applied it to conventional car data, and performed cluster analysis to find typical usage profiles. They compare these features to EV consumption statistics from a smaller dataset (8). Bogdan Ovidiu Varga et al. explored the difficulties of forecasting electric vehicle range (9). Electric car range depends on driver behavior, weather, and traffic. The author also suggested complicated software methods for range prediction in future electric automobiles. To meet the UK's ambitious electric vehicle (EV) targets, Michael Nicholas and Nic Lutsey suggested expanding charging infrastructure (10). Quantifying public, workplace, and quick charging infrastructure was stressed in the report. The analysis also showed that some places require more chargers than others, requiring tailored investment. In 2020, Nele Rietmann et al. utilized the logistic growth model to forecast EV inventory in 26 nations on five continents (11). Real sales data from 2010 to 2018 was used to forecast these nations till 2035. The analysis also showed that EV growth can lower CO2 emissions if governments invest heavily in renewable energy.

2.3 Research Gaps

Although considerable research has been conducted on the adoption patterns, range prediction, and infrastructure required to support the expansion of electric vehicles (EVs), there is still a lack of understanding of the time-dependent patterns and trends in EV adoption utilizing advanced time series approaches. Prior research has predominantly concentrated on linear, nonlinear, and hybrid prediction techniques, cluster analysis for measuring consumption patterns, and logistic growth models for making predictions. Yet, the use of autocorrelation, partial autocorrelation, and seasonal decomposition methods to study the temporal dynamics and predict future demand for electric vehicles (EVs) has been quite restricted. Furthermore, although studying the impact of driver behavior, weather conditions, and traffic on electric vehicle (EV) range, there is a dearth of research that incorporates these elements into comprehensive time series models to enhance the accuracy of predicting EV adoption rates and demand. Mitigating this deficiency would improve the accuracy of demand prediction and help to more efficient design for charging infrastructure and grid stability.

Chapter 3

Study Methodology

The acronym SEMMA represents the stages of Sample, Explore, Modify, Model, and Assess (12). Quantitative data mining is a systematic approach created by the SAS Institute to direct the development of prediction models. SEMMA prioritises the repetitive procedures required to analyse and extract valuable information from data, with a particular emphasis on the quality and improvement of data to construct strong analytical models.

3.1 Sample

The initial stage in applying the SEMMA methodology to the vehicle licensing statistics dataset is to choose a representative subset of data from the original dataset. This guarantees our ability to effectively investigate and simulate the data without incurring excessive computational expenses. In order to capture a wide range of vehicle characteristics (such as various 'BodyType', 'Fuel', and 'LicenceStatus'), we may choose a sample of specific records that are typical of different time periods within the dataset spanning from the 3rd quarter of 2014 to the last quarter of 2023. The sample size should be sufficiently big to ensure statistical power while also being within practical limits for thorough analysis.

In the absence of data for the first two quarters of 2014, it is advisable to exclude 2014 and instead use data from 2015 to 2023. The 'BodyType' column contains several categories of vehicles, including cars, motorcycles, buses, heavy vehicles, and others. We partitioned the data based on cars from BodyType and Licenced from LicenceStatus and grouped the data by fuel.

3.2 Explore

There exist various fuel kinds, including several categories of electric fuel. Given that this project exclusively concentrates on Electric vehicles, any other sort of vehicles is excluded. This filtered dataset consists of seven distinct categories of electric vehicles. The bar chart labeled "Distribution of different types of Electric Vehicles [2015–2023]" depicts the proportion of different electric car fuel types in the market from 2015 to 2023 3.1. With 57.71% (22 million) of the total registrations, hybrid electric (petrol) vehicles lead the market, reflecting a significant preference for hybrids that integrate electric power with petrol adaptability. Fully electric battery electric vehicles, devoid of any internal combustion engine, constitute 20.97% (8 million) and are experiencing substantial acceptance as the market shifts towards zero-emission alternatives. Plug-in hybrid electric (petrol) vehicles also account for a significant portion of 17.08% (6 million), providing car owners the advantages of both electric propulsion and petrol support. Hybrid electric (diesel), range-extended electric, plug-in hybrid electric (diesel), and fuel cell electric vehicles together account for a mere 4.24% (1.6 million) of the market, underscoring their relatively low uptake in comparison to the more widespread hybrid and fully electric vehicles. This pattern indicates a period of change in the market, characterized by an increasing preference for electric and plug-in hybrid alternatives, propelled by technological progress and rising environmental consciousness.

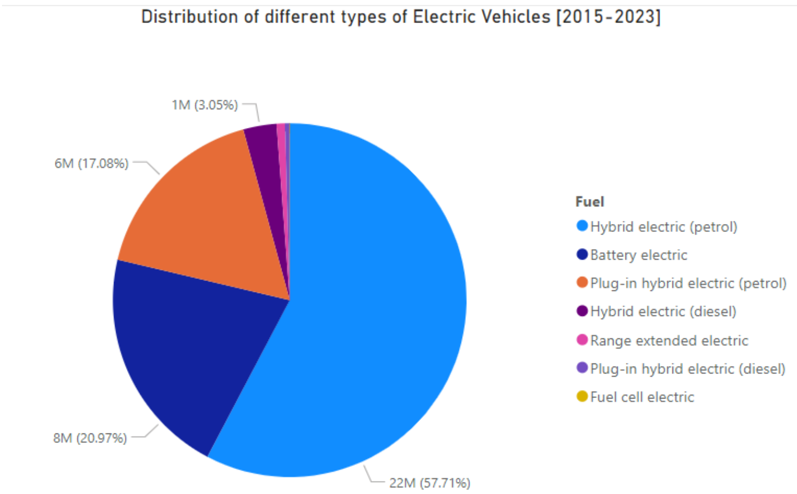

FIGURE 3.1: Licensed Different Electric Vehicles

The bar chart displays data on the production of electric vehicles by the top 10 automobile manufacturers between 2015 and 2023 3.2. Toyota emerges as the unequivocal frontrunner, manufacturing 11.8 million electric vehicles, a figure that surpasses those of its competitors. Following closely are Lexus and BMW, with 3.6 million and 2.8 million vehicles, respectively. Additional manufacturers, including Kia, Hyundai, Tesla, Nissan, Mercedes, Honda, and Mitsubishi, manufactured electric vehicles ranging from 1.3 million to 1.8 million units each. Toyota's graph clearly demonstrates their overwhelming control in the production of electric vehicles, surpassing all other competitors by a significant margin.

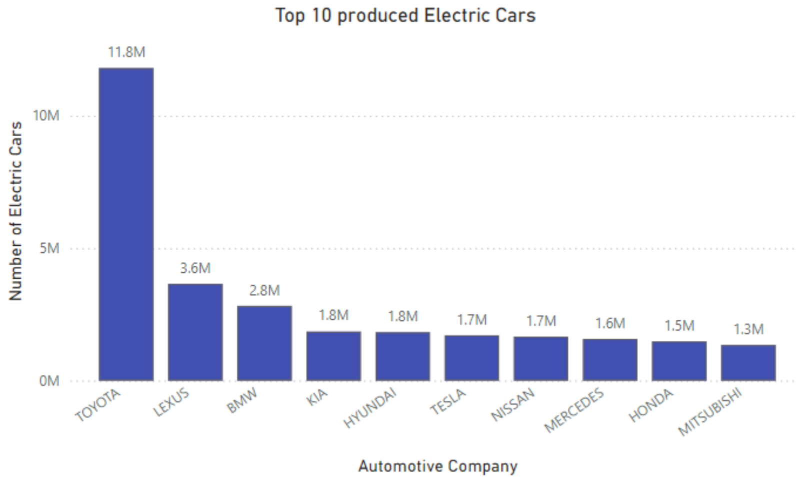

FIGURE 3.2: Top 10 Automative Company

3.2.1 Autocorrelation & Partial Autocorrelation

The presented figures depict the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) (13) for three distinct categories of electric vehicles, namely Battery Electric, Hybrid Electric (Petrol), and Plug-in Hybrid Electric (Petrol versions). The uppermost row comprises the ACF plots, whereas the lowermost row has the matching PACF plots.

3.2.1.1 Battery Electric Vehicle:

- **ACF:** The autocorrelation decays progressively across the lags, suggesting a continuous decrease in the correlation between observations as time progresses. This observation implies that it is typical in non-stationary time series data for past values to exert a persistent influence on future values.
- **PACF:** After the first lag, the partial autocorrelation decreases significantly and continuously falls within the confidence interval for the following delays.

This pattern indicates that the data can be effectively represented by an autoregressive (AR) process, in which only the first stage has a substantial influence on the present value.

3.2.1.2 Hybrid Electric (Petrol):

- **ACF:** The ACF plot shows a progressive reduction akin to that of Battery Electric Vehicles, suggesting a robust autocorrelation at shorter time intervals that progressively diminishes, hence confirming the non-stationarity of the time series.
- **PACF:** The PACF figure displays a notable decrease beginning with the first lag, and the following lags are consistently within the confidence interval. The observed behavior implies that the time series can be effectively represented by an AR(1) process, in which only the first lag has a substantial impact on the current value.

3.2.1.3 Battery Electric Vehicle:

- **ACF:** The ACF exhibits a progressive decrease, akin to the characteristics of the other two vehicle categories, suggesting that previous values have an impact on future values throughout several time periods. Such behavior suggests a gradual decrease in autocorrelation and perhaps non-stationarity.
- **PACF:** The PACF plot presents a statistically significant value at lag 1, and all time lags thereafter lie within the confidence interval. Therefore, similar to other vehicle categories, this implies that an AR(1) model may be suitable.

The ACF plots for all three vehicle types show a steady decrease in autocorrelation, indicating a non-stationary time series. After the first lag, the PACF plots for each vehicle type decline significantly, suggesting an autoregressive model of order 1 (AR(1)) could explain these time series dynamics. This tendency implies that the present value is largely impacted by the previous value and little by following ones.

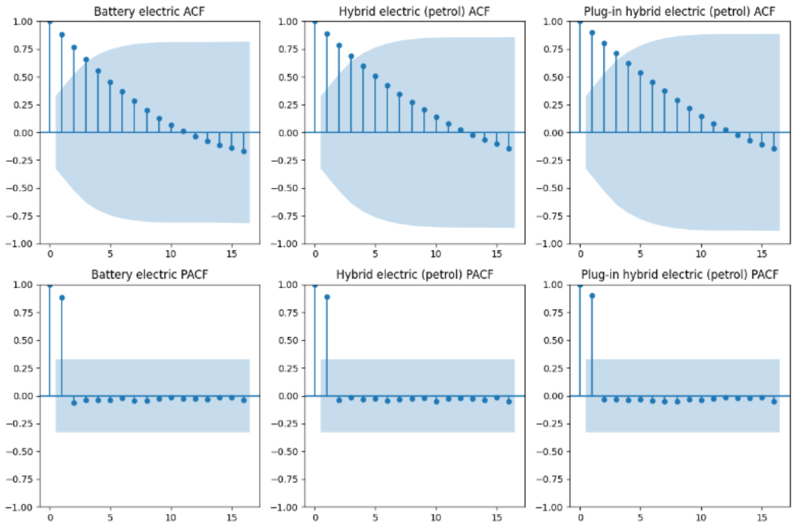

FIGURE 3.3: ACF & PACF of three types of EVs

3.2.1.4 Fuel Cell Electric Vehicle (FCEV):

- **ACF:** Indicates a robust and enduring autocorrelation structure by the steady decline of positive autocorrelations across several delays.
- **PACF:** A substantial surge at lag 1, followed by rapid decline to zero, indicating an autoregressive (AR) process.

3.2.1.5 Hybrid Electric Vehicle (Diesel):

- **ACF:** Demonstrates a progressive decline, albeit at a considerably accelerated rate compared to the FCEV. The ACF undergoes a reverse shift from positive to negative values, suggesting the presence of a cyclical mechanism.

- **PACF:** A strong spike at lag 1, followed by a more rapid decline than FCEV, indicating a weaker autoregressive influence.

3.2.1.6 Plug-in Hybrid Electric Vehicle (Diesel):

- **ACF:** Has a progressive decline in positive autocorrelations and ultimately transitions to negative values, indicating a trend or cyclical pattern with lower persistence compared to FCEV and Hybrid Electric Vehicle (Diesel).
- **PACF:** Similar to the other vehicles, a strong spike at lag 1 with a sharp decline, consistent with an AR(1) model, but the pattern is less pronounced after lag 1 compared to FCEV.

3.2.1.7 Range Extended Electric Vehicle:

- **ACF:** shows a progressive decline, characterized by positive autocorrelations that ultimately approach zero and then transition to negative values, akin to the Plug-in Hybrid Electric Vehicle but with minor fluctuations in the decay rate.
- **PACF:** Characterised by a significant increase at lag 1, followed by a fast decrease and oscillating about zero, indicating an autoregressive (AR)(1) process. However, the oscillations imply possible little effects from higher-order delays.

All four vehicle types' PACF plots have a substantial spike at lag 1 (suggesting an autoregressive model of order 1, AR(1)), but their ACF and PACF charts differ in decay rate and positive-to-negative correlation transition. These discrepancies reflect different time series data persistence, trend, and cyclicality. The general

AR(1) model may apply to everyone, however their ACF and PACF patterns reflect different data behaviors.

3.2.2 Seasonal Decomposition

Seasonal decomposition (14) statistically separates time data into trend, seasonality, and residual. This strategy works well for examining data with recurring patterns throughout time. Seasonal decomposition is additive or multiplicative.

3.2.2.1 Additive Decomposition:

1. **Hybrid Electric Petrol**

   - **Observed (Quantity):** Data demonstrates a clear growing trend, suggesting that hybrid electric petrol has increased dramatically over time. Growth looks to be accelerating late in the decade.
   - **Trend:** Since around 2019, the trend line has shown a persistent upward trajectory. These statistics indicate that hybrid electric petrol consumption has been rising over time. The trend line shows a steady rising growth rate, maybe due to technology advancement, hybrid car use, or consumer behavior changes.
   - **Seasonal:** The seasonal graph has annual peaks and troughs. Seasonal swings in hybrid electric petrol usage or sales may be caused by weather, holidays, or fuel prices. Persistent trend shows seasonal influences affect quantity predictably.
   - **Residual:** The residuals scatter around the zero line without a pattern, suggesting that trend and seasonal components explain much of the data

variability. However, some changes, especially in later years, suggest irregular events or anomalies may be impacting the statistics.

In conclusion, this hybrid electric petrol time series decomposition reveals a strong increase trend and a constant seasonal pattern, indicating rising demand. The residuals show that the model explains most of the variance, but some unpredictability remains, especially toward the conclusion of the term. This decomposition helps predict future trends and make strategic decisions based on observed patterns by explaining the elements that drive quantity.

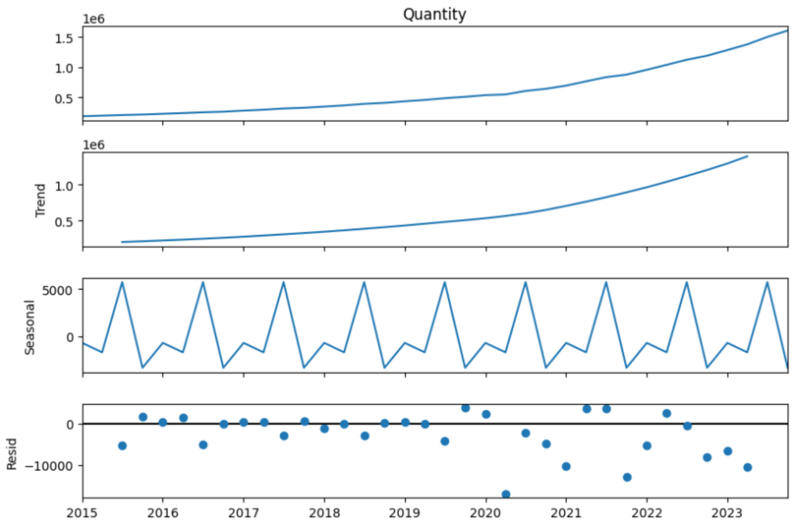

FIGURE 3.4: Additive Decomposition of Hybrid Electric Petrol

3.2.2.2 Multiplicative Decomposition:

1. **Hybrid Electric Petrol**

   - **Observed (Quantity):** The data indicates a strong increasing tendency, like additive decomposition. The increase is consistent and accelerates at the end, indicating a growing amount of hybrid electric petrol.

- **Trend:** Like additive decomposition, this trend line shows a consistent and rising increase in hybrid electric petrol over time. This shows a growing market or usage of hybrid electric petrol over time.
- **Seasonal:** In the multiplicative model, seasonal fluctuations occur around 1. This shows that seasonal effects are proportionate to the trend; as quantity grows, peaks and troughs have a slightly greater impact. Regular seasonal patterns show predictable, cyclic variations in hybrid electric petrol quantity, possibly connected to external factors like seasons or events that affect usage or demand.
- **Residual:** The residuals are densely grouped around 1, indicating that the multiplicative model explains most data variation. With little inexplicable irregularity, the trend and seasonal components appear to have captured much of the time series' structure.

This multiplicative decomposition of hybrid electric petrol shows a strong and fast upward trend with seasonal variations proportionate to the trend. The seasonal component shows regular cyclic behavior, and the residuals suggest that the model reflects the data's dynamics. The multiplicative decomposition allows for the increasing amplitude of seasonal changes along with the expanding trend, offering a complete knowledge of the elements driving hybrid electric petrol demand over time.

3.3 Modify

Individual dataframes are generated for each distinct electric vehicle type from the primary electric cars dataset. Next, the dataframe is transposed using Python's Pandas package [], and the column 'Date' is chosen as the index.

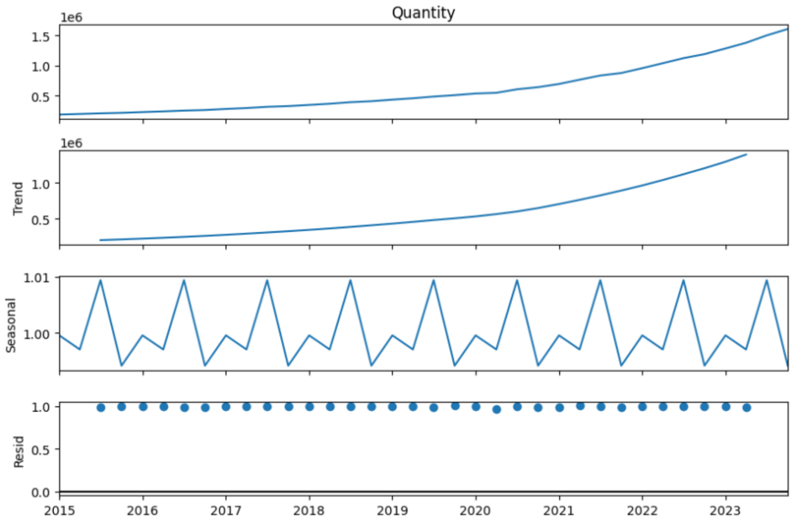

FIGURE 3.5: Multiplicative Decomposition of Hybrid Electric Petrol

ACF of Battery Electric Petrol shows a non-stationary time series. Furthermore, to be accurate, Augmented Dickey-Fuller will be used (15). ADF tests p-values evaluate the evidence against the null hypothesis of non-stationarity. Low p-values strongly suggest stationarity, while high p-values show non-stationarity. We accept p-values below 0.005.

Here is the equation 3.1

$$\Delta Y_t = \alpha + \beta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + ... + \gamma_p \Delta Y_{t-p} + \varepsilon_t \tag{3.1}$$

Applying log transformation can effectively stabilize the variance of a time series, so increasing its stationarity and reducing the skewness of a distribution, thus enhancing its symmetry (16). Log is used using NumPy (17).

$$\log(Y_t) = \log(Y_0) + \alpha t + \varepsilon_t \tag{3.2}$$

Moreover, difference transformation is needed to be used to make a non-stationary time series stationary. This is often necessary because many statistical models, such as ARIMA and SARIMA, assume stationarity.

First Difference transformation equation 3.3

$$\Delta f(x) = f(x+h) - f(x) \tag{3.3}$$

Second Difference transformation equation 3.4

$$\Delta^2 f(x) = \Delta f(x+h) - \Delta f(x) \tag{3.4}$$

After applying the second difference transformation, the p-value decreased to 2.2069e-09, which is below the significance level of 0.005. We partitioned the transformed dataset into a training set and a test set, with sizes of 32 and 4.

3.4 Model

Given the stationarity of the data, we have selected 10 distinct time series models.

3.4.1 AutoARIMA:

Automatically selects the best ARIMA model by optimizing the parameters p, d, and q using information criteria like AIC (18).

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q} + \epsilon_t \tag{3.5}$$

3.4.2 AutoETS:

The optimal selection of the best Exponential Smoothing model among ETS (Error, Trend, Seasonality) models is achieved by an automated process that exploits information criteria (19).

$$\hat{y}_{t+1} = \alpha(y_t - \hat{y}_t) + (1 - \alpha)(L_t + S_t) \tag{3.6}$$

3.4.3 Holt-Winters:

The exponential smoothing method is employed to predict data that exhibits both trend and seasonality, utilizing additive or multiplicative models (20).

$$\hat{y}_{t+h}\& = \hat{l}_t + h\hat{b}_t + \hat{s}_{t+h-m}\hat{l}_t\& = \alpha(y_t - \hat{s}_{t-m}) + (1-\alpha)(\hat{l}_{t-1} + \hat{b}_{t-1})\hat{b}_t\& = \beta(\hat{l}_t - \hat{l}_{t-1}) + (1-\beta)\hat{b}_{t-1}\hat{s}_t\& = \gamma(\tag{3.7}$$

3.4.4 Historic Average (Mean):

Forecasts based on the simple average of all past observations (21).

$$\hat{y}_{t+1} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{3.8}$$

3.4.5 Naive Forecast:

Presumes that the projection for the next period is equivalent to the most recent observed statistic (22).

$$\hat{y}_{t+1} = y_t \tag{3.9}$$

3.4.6 Decision Tree:

A nonlinear model that forecasts results by acquiring decision rules deduced from characteristics of the data (23).

3.4.7 AR (Auto-Regressive):

Forecasts generated by a linear weighted mixture of past values in the time series (24).

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \epsilon_t \tag{3.10}$$

3.4.8 MA (Moving Average):

Forecasts derived from a linear mixture of previous forecast inaccuracies (25).

$$y_t = c + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q} + \epsilon_t \tag{3.11}$$

3.4.9 ARIMA (Auto-Regressive Integrated Moving Average):

Combines AR, I (Integration), and MA to model the time series (26). The selected order =(1, 0, 1) model incorporates a single lagged moment of the time series (AR(1)) and a single lagged error (MA(1)). The inclusion of the value "0" in the middle suggests that no differencing is required to achieve stationarity of the data.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q} + \epsilon_t \tag{3.12}$$

3.4.10 SARIMA (Seasonal ARIMA)):

Expands upon ARIMA by integrating seasonal variations, auto-regression, and moving averages (27). The seasonal ARIMA model SARIMA(1, 0, 1, 12) is characterized by a single lagged seasonal value (SAR(1)), absence of seasonal differencing, a single lagged seasonal error (SMA(1)), and a seasonal period of 12. This is beneficial for data exhibiting recurring seasonal trends that occur every 12 sequential time intervals.

$$y_t = c+\phi_1 y_{t-1}+\ldots+\phi_p y_{t-p}+\theta_1\epsilon_{t-1}+\ldots+\theta_q\epsilon_{t-q}+\phi_1^{(1)} y_{t-s}+\ldots+\phi_p^{(1)} y_{t-ps}+\theta_1^{(1)}\epsilon_{t-s}+\ldots+\theta_q^{(1)}\epsilon_{t-qs}+\epsilon_t \tag{3.13}$$

3.5 Assess

The evaluation of these models is conducted using mean squared error (MSE) and mean absolute error (MAE) functions from sklearn.metrics. The Mean Squared Error (MSE) and Mean Absolute Error (MAE) are widely used measures for assessing the effectiveness of regression models, particularly those applied to time series data (28). They quantify the mean discrepancy between the computed values and the observed values.

Mean Squared Error (MSE) Equation 3.14

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.14}$$

- n: Number of data points
- yi: Actual value for the i-th data point

- yî: Predicted value for the i-th data point

Mean Absolute Error (MAE) Equation 3.15

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{3.15}$$

- n: Number of data points
- yi: The actual value at time i
- yî: The predicted value at time i

Chapter 4

Result

Data spanning from 2015 to 2022 is utilized for model training, while data from 2023 is employed for model evaluation. The figure below shows the temporal time series graph for the comparison of several forecasting models with real data 4.1.

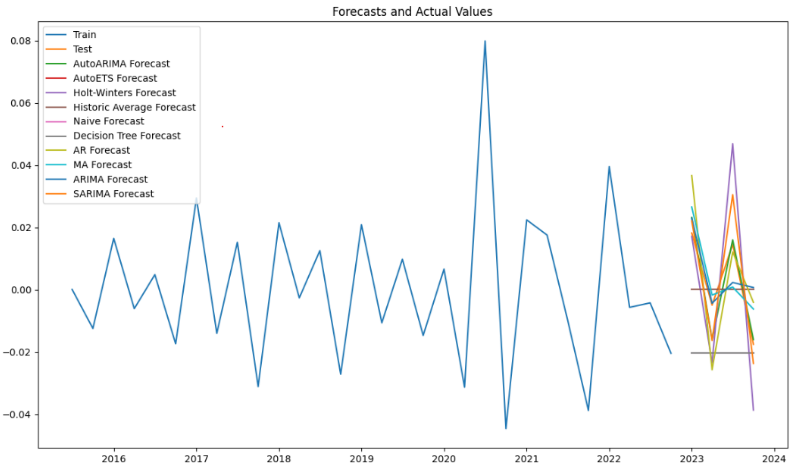

FIGURE 4.1: The comparison of several forecasting models

4.0.1 Best Performing Models

With the lowest RMSE (0.0062) and MAE (0.0048) values among all models, the AutoARIMA model emerges as the top-performing model. These findings indicate that AutoARIMA exhibits the greatest precision in forecasting the future values of the series, successfully encompassing both patterns and seasonal variations. This model is the optimal selection for obtaining the most precise predictions.

In addition, the Moving Average (MA) model, SARIMA, and ARIMA models provide good performance, as evidenced by their low RMSE and MAE values. These models demonstrate superior performance in managing the properties of the data and serve as dependable substitutes for AutoARIMA. The Mean Absolute Error (RMSE) of the MA model is 0.0099, closely trailing AutoARIMA. SARIMA and ARIMA also exhibit strong predictive capability with their commendable low error metrics.

4.0.2 Moderately Performing Models

The Autoregressive (AR), AutoETS, and Historic Average models exhibit intermediate performance, as evidenced by somewhat higher RMSE and MAE values in comparison to the top-performing models. As an example, the AR model has a Root Mean Square Error (RMSE) of 0.0155 and Mean Absolute Error (MAE) of 0.0137. These values are considered acceptable but suggest that these models may not replicate the data as precisely as AutoARIMA or other exceptional performers. These models may nonetheless be valuable in situations when prioritizing model simplicity or interpretability takes precedence over objective accuracy.

4.0.3 Poor Performing Models

The Holt-Winters, Naive, and Decision Tree models have the lowest accuracy, as evidenced by their dramatically larger RMSE and MAE estimates. The Holt-Winters model, with a root mean square error (RMSE) of 0.0214 and mean absolute error (MAE) of 0.0183, has a restricted capacity to precisely forecast future values, partly because of its treatment of seasonal elements. The Naive and Decision Tree models exhibit the lowest performance, as evidenced by their RMSE of 0.0272 and MAE of 0.0230. Consequently, these models fail to adequately consider temporal patterns or variations in the data, resulting in less dependable predictions.

4.0.4 Conclusion

To summarize, the AutoARIMA model is the optimal selection for predicting in this particular situation, since it provides the highest level of accuracy and effectively captures the fundamental patterns present in the data. Furthermore, MA, SARIMA, and ARIMA offer robust alternatives that demonstrate outstanding performance. For less complex or demanding applications, models like as AR, AutoETS, and Historic Average may be worthwhile considerations. Conversely, Holt-Winters, Naive, and Decision Tree models have inadequate predictive efficacy and are less appropriate for accurate forecasting requirements.

Table 4.1: Hybrid Electric Patrol Forecasting Model Performance Metrics

Model	**RMSE**	**MAE**
AutoARIMA	0.0062	0.0048
AutoETS	0.0148	0.0138
Holt-Winters	0.0214	0.0183
Historic Average	0.0147	0.0138
Naive	0.0272	0.0230
Decision Tree	0.0272	0.0230
AR	0.0155	0.0137
MA	0.0099	0.0091
ARIMA	0.0112	0.0089
SARIMA	0.0105	0.0093

Chapter 5

Discussion

The present study primarily concentrated on the time series analysis of electric vehicles in the UK. There were 7 types of categorized electric cars. Among them, Hybrid Electric Petrol was the largest one and was selected as a primary dataset. Preliminary autocorrelation and partial autocorrelation tests indicated that the data was not stationary. To tackle this issue, log and difference transformations were used following the identification of non-stationary and seasonality using ACF and PACF analysis. A total of 10 time series models were then deployed, among which AutoARIMA, MA, SARIMA, and ARIMA exhibited exceptional performance. The analysis of the rest categorized electric types of cars is in the Appendix.

Appendix A

The rest four types of electric vehicles ACF & PACF

In this appendix, we provide the ACF and PACF plots for the remaining four types of electric vehicles.

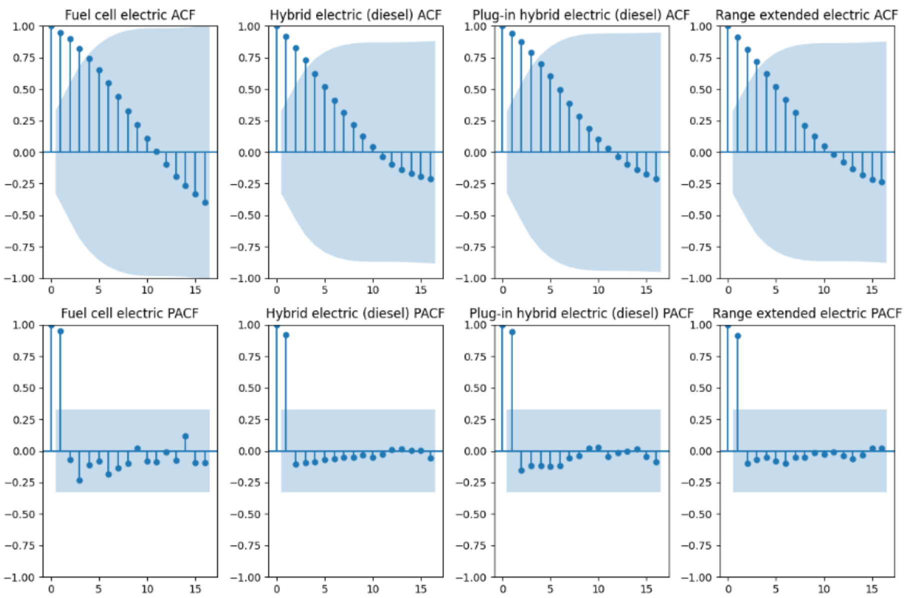

FIGURE A.1: ACF and PACF plots for electric vehicle types.

Appendix B

Seasonal Decomposition

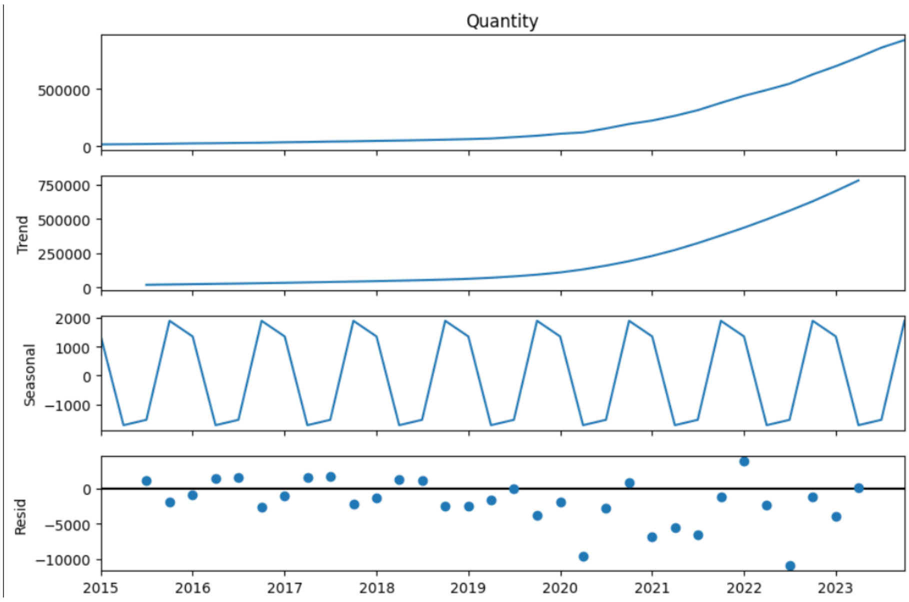

FIGURE B.1: Additive Decomposition of Battery Electric

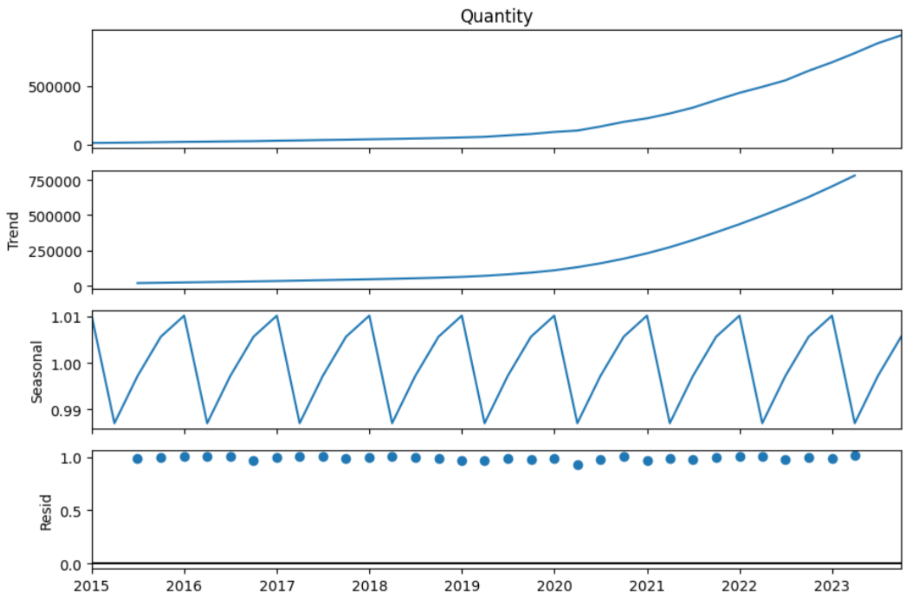

FIGURE B.2: Multiplicative decomposition of Battery Electric

Appendix C

Battery Electric Forecasting Model Performance Metrics

ARIMA has the best overall performance, with RMSE 0.0174 and MAE 0.0140. ARIMA has the lowest prediction errors in this dataset, making it the most accurate and dependable forecasting model.

Following ARIMA, AutoARIMA performs well with RMSE 0.0198 and MAE 0.0163. These indicators are just slightly higher than ARIMA's, demonstrating AutoARIMA is also good for accurate predicting. SARIMA does well with an RMSE of 0.0198 and an MAE of 0.0189, making it a good option to ARIMA although less exact.

AutoETS and Historic Average models perform similarly, with RMSE around 0.0213 and MAE about 0.0168 to 0.0169. Although less accurate than ARIMA-based approaches, these models have modest error rates, making them reasonably dependable for predicting.

The Moving Average (MA) model performs well with an RMSE of 0.0211 and an MAE of 0.0155. Although less exact than ARIMA-based models, these measures show that the MA model is reasonably accurate and might be used for predicting.

Unfortunately, the Holt-Winters model has the greatest RMSE of 0.0661 and MAE of 0.0572. The Holt-Winters model has the highest prediction error values, making it the least dependable. In this case, the Naive and Decision Tree models show significant RMSE (0.0499) and MAE (0.0482), indicating errors and unsuitability for reliable forecasting.

The AutoRegressive (AR) model performs moderately with RMSE 0.0246 and MAE 0.0184. These figures show that AR has higher prediction errors than ARIMA-based models, but it may be suitable for situations when higher error rates are acceptable.

Based on its low RMSE and MAE, the ARIMA model forecasts best, followed by AutoARIMA and SARIMA. Due to their large error rates, the Holt-Winters, Naive, and Decision Tree models are the least dependable for exact forecasting.

Table C.1: Battery Electric Forecasting Model Performance Metrics

Model	**RMSE**	**MAE**
AutoARIMA	0.0198	0.0163
AutoETS	0.0213	0.0168
Holt-Winters	0.0661	0.0572
Historic Average	0.0213	0.0169
Naive	0.0499	0.0482
Decision Tree	0.0499	0.0482
AR	0.0246	0.0184
MA	0.0211	0.0155
ARIMA	0.0174	0.0140
SARIMA	0.0198	0.0189

Bibliography

[1] T. P. Carolin Reiner, George Beard and N. Kinnear, "Driving and accelerating the adoption of electric vehicles in the uk," *GOV.UK*, July 2020.

[2] M. T. E. P. K. M. A. A. M. C. S. L. S. L. V. M. A. O. B. R. e. H.-O. Pörtner, D.C. Roberts, "Climate change 2022: Impacts, adaptation and vulnerability," *IPCC*, 2022.

[3] SMMT, "Automotive sustainability report 2023 data," *SMMT*, 2023.

[4] IEA, "Global ev outlook 2021," 2021.

[5] P. Patil, "An empirical study of the factors influencing the adoption of electric vehicles," *Contemporary Issues in Behavioral and Social Sciences*, vol. 4, p. 1–13, Jan. 2020.

[6] D. for Transport, "Vehicle licensing statistics data tables," 2024.

[7] F. Marzbnai, A. Osman, and M. Hassan, "Electric vehicle energy demand prediction techniques: An in-depth and critical systematic review," *IEEE Access*, vol. PP, pp. 1–1, 01 2023.

[8] C. Crozier, D. Apostolopoulou, and M. Mcculloch, "Clustering of usage profiles for electric vehicle behaviour analysis," pp. 1–6, 10 2018.

[9] B. Varga, A. Sagoian, and F. Mariasiu, "Prediction of electric vehicle range: A comprehensive review of current issues and challenges," *Energies*, vol. 15, p. 946, 03 2019.

[10] M. Nicholas and N. Lutsey, "Quantifying the electric vehicle charging infrastructure gap in the united kingdom," 08 2020.

[11] N. Rietmann, B. Hügler, and T. Lieven, "Forecasting the trajectory of electric vehicle sales and the consequences for worldwide co2 emissions," *Journal of Cleaner Production*, vol. 261, 07 2020.

[12] Tiffani, "Data mining semma," 2018.

[13] Q. Liu, Z. Li, Y. Ji, L. Martinez, Z. Ul-Haq, A. Javaid, W. lu, and J. Wang, "Forecasting the seasonality and trend of pulmonary tuberculosis in jiangsu province of china using advanced statistical time-series analyses," *Infection and Drug Resistance*, vol. 12, pp. 2311–2322, 07 2019.

[14] S. Raghuvanshi, "Assessing the impact of seasonal decomposition on the time series analysis accuracy: A comprehensive study," *International Journal for Research in Applied Science and Engineering Technology*, vol. 12, pp. 2988–2992, 05 2024.

[15] Z. Guo, "Research on the augmented dickey-fuller test for predicting stock prices and returns," *Advances in Economics, Management and Political Sciences*, vol. 44, pp. 101–106, 11 2023.

[16] G. Choi, J. Buckley, J. Kuiper, and A. Keil, "Log-transformation of independent variables: Must we?," *Epidemiology (Cambridge, Mass.)*, vol. 33, pp. 843–853, 10 2022.

[17] P. Gupta and A. Bagchi, *Introduction to NumPy*, pp. 127–159. 01 2024.

[18] pmdarima, “auto$_a rima$,″ 2018.

[19] SKTIME, “autoets,”

[20] OTEXTS, “Holt-winters’ seasonal method,”

[21] A. Bakker, “The early history of average values and implications for education,” *Journal of Statistics Education*, vol. 11, no. 1, 2003.

[22] . K. A. Ahuja, S., “Expectation-based probabilistic naive approach for forecasting involving optimized parameter estimation.,” *Arabian journal for science and engineering*, 2023.

[23] . L. Y. Song, Y. Y., “Decision tree methods: applications for classification and prediction.,” *Shanghai archives of psychiatry*, 2015.

[24] M. C. L. V. L. S. . S. N. W. A. Liu, Y., “Asymmetric autoregressive models: statistical aspects and a financial application under covid-19 pandemic.,” *Journal of applied statistics*, 2015.

[25] . I. S. Lukić, V., “Optimizing moving average control procedures for small-volume laboratories: can it be done?,” *Biochemia medica*, 2019.

[26] C. M. I. N. D. . S. L. Y. Earnest, A., “Using autoregressive integrated moving average (arima) models to predict and monitor the number of beds occupied during a sars outbreak in a tertiary hospital in singapore,” *BMC health services research*, 2005.

[27] T. H. Noor, A. M. Almars, M. Alwateer, M. Almaliki, I. Gad, and E.-S. Atlam, “Sarima: A seasonal autoregressive integrated moving average model for crime analysis in saudi arabia,” *Electronics*, vol. 11, no. 23, 2022.

[28] B. J. B., “Classifiers and their metrics quantified,” *BMC health services research*, 2018.