Proceedings of the 8th International Conference on Economic Management and Green Development
DOI: 10.54254/2754-1169/126/2024.18340

UK Battery Electric Vehicle Sales Prediction Using ARIMA and SARIMA Models

Yourong Lu[1,a,*]

[1]*Department of Mathematics, University College London, Gower Street, London, United Kingdom*
a. zcahuhx@ucl.ac.uk
**corresponding author*

Abstract: As a crucial part of fulfilling its commitment to the Paris Agreement, the adoption of electric vehicles (EVs) is a milestone of the UK's strategy to achieve environmental sustainability. The UK government has been employing tactics like providing subsidies to EV buyers and exempting EVs from congestion charge and road tax to promote the sales of clean fuel vehicles. Among these varieties of EVs, only battery electric vehicles (BEVs) are running completely on electric power with zero emission, which makes it more worthy to study the sales pattern and predict the future sales figures of BEVs. This passage employs Auto Regressive Integrated Moving Average (ARIMA) and Seasonal Auto Regressive Integrated Moving Average (SARIMA) models to forecast the future sales figures based on the past sales of UK's BEVs from 2019 to 2024. It can be concluded from the prediction that sales figures are about to keep the trend of raising for the following year meanwhile strong fluctuations could be witnessed between March and September. This conclusion should be to provide a foresight of UK's BEV market for investors and BEV manufacturers, as well as a better understanding of UK's transition to sustainable transportation system for policy makers.

Keywords: Sales prediction, ARIMA, SARIMA, Time series.

1. Introduction

As the world keeps engaging in the crucial issue of climate change, the transition from traditional fossil fuels to sustainable clean energy sources has become a critical priority. In the UK, road transport is a significant contributor to carbon dioxide (CO_2) emissions, approximately 27% of the UK's total greenhouse gas emissions. Specifically, passenger cars are the largest source, contributing between 55% and 60% of road transport CO emissions [1]. Electrification is recognized as the most effective way to decarbonize the transportation sector globally [2]. The European Green Deal outlines a set of policy initiatives with the goal of making Europe climate neutral by 2050, including zero greenhouse gas emissions from transport, which requires a shift away from petroleum-based vehicles [3]. China has been actively formulating subsidies policies, vigorously supporting the deployment of charging infrastructure, and encouraging car manufacturers to research and produce electric vehicles [4]. The United Kingdom, however, as a significant player in the international auto market and a pioneer of environmentalism, has also devoted itself to an ecological transformation in which the promotion of electric vehicles is an essential part.

In the UK, the electric vehicle (EV) market includes a variety of vehicle types, including Battery Electric Vehicles (BEVs), Plug-in Hybrid Electric Vehicles (PHEVs), Hybrid Electric Vehicles

DOI: 10.54254/2754-1169/126/2024.18340

(HEVs), and Fuel Cell Electric Vehicles (FCEVs) [5]. Since PHEVs and HEVs combine internal combustion engines which still operate with fuels, and FCEVs occupy an exiguous part of the market due to their immaturity, BEVs are the ones that run solely on renewable energy and produce zero pipeline emissions. Therefore, BEVs are more likely to experience sustained development and play a more significant role in the future.

In the past decade, the rapid adoption of BEVs in the United Kingdom accelerated due to breakthroughs in battery technology and a rise in the number of charging stations [6]. The total number of BEV registrations in the UK increased from less than ten thousand in 2018 to more than 300 thousand in 2023 [7]. Thus, establishing a forecast model based on sales figures can provide a more profound and numeric understanding of the demand for the UK's BEV market, as well as provide a valuable reference for automotive enterprises' strategies.

This paper utilizes the monthly registration figures of BEVs in the past five years and fits the Autoregressive Integrated Moving Average (ARIMA) model and Seasonal Autoregressive Integrated Moving Average (SARIMA) model to forecast due to their ability to featly handle autocorrelation, non-stationarity and seasonality time series.

There are many studies on the sales forecasting using time series models, Guo and Gong used SARIMA and Holt-winters methods to forecast pulmonary tuberculosis cases in China [8]. Zhang used hybrid ARIMA combined with neural network to predict [9]. Contreras et.al applied ARIMA models to predict next-day electricity price [10]. However most didn't make a comparison between ARIMA and SARIMA model. In this study besides making a prediction, we also compare the estimated performance of these two models.

2. Methodology

2.1. ARIMA

ARIMA model is a popular statistical method for analyzing and forecasting time series data, which is a collection of numbers taken at exactly spaced intervals. ARIMA model is determined by three key parameters of the form ARIMA(p,d,q) , where:

AR (autoregressive) part (p): indicates the number of time lags included in the observation.

I (integrated) part (d): refers to the number of differencing steps required to eliminate trends and stabilize the variance of time series.

MA (moving average) part (q): captures the order of the lagged forecast errors included in the model.

Arima model is implemented by first differencing the time series by d times to obtain stability. Next, an AR(p) model is fitted to the differenced series. Finally, an MA(q) model is applied to the residuals of the AR model.

ARIMA uses the following Equations 1-2 and describes the autocorrelation in the data.

$$Y_t = (1 - L)^d X_t \tag{1}$$

$$Y_t\left(1 - \Sigma_{i=1}^{p} \Phi_i L^i\right) = \left(1 + \Sigma_{i=1}^{q} \theta_i L^i\right)\varepsilon_t \tag{2}$$

The parameter p is the number of time lags while q is the order of the MA model. L represents the lag operator and θ is the parameter of the MA part, Φ is the parameter of the AR part of the model.

2.2. SARIMA

SARIMA model works better with seasonal data by taking consideration of additional four seasonal factors. SARIMA model is represented in the form of SARIMA(p,d,q)(P,D,Q)[s], where

SAR (seasonal autoregressive) part (P): indicates how many seasonal lag observations are there in the observation.

SI (seasonal integrated) part (D): the number of seasonal differencing steps required to attain variance stability.

SMA (seasonal moving average) part (Q): denotes the seasonal error terms as a linear combination of error terms from the preceding season.

s (seasonal period) part (s): the number of periods per season. For example, for monthly data with a season of 12 months, s = 12.

SARIMA uses the following Equations 3-4-5.

$$\Phi(B)\ \Delta^D X\_t\ =\ \Theta(B)\alpha\_t \quad (3)$$

$$\Phi_s(B^s)\Delta_s^D \alpha_t = s\Theta(B^s)\alpha_t \quad (4)$$

$$\Phi(B)_s \Phi(B^s)\Delta_s^D \Delta^d X_t = \Theta(B_s)\Theta(B^s)\alpha_t \quad (5)$$

The SARIMA model is an extension of the ARIMA model, adding seasonal factors into the ARIMA framework. For instance, $\Theta(B_s)$ represents non-seasonal AR polynomial while $\Theta(B^s)$ represents seasonal AR polynomial. Δ^D is the non-seasonal differencing operator while Δ_s^D is the seasonal differencing operator.

3. Experimental results

3.1. Source of Data

The registration figures of BEVs in the UK are sourced from Society of Motor Manufacturers and Traders (SMMT: https://www.smmt.co.uk/vehicle-data/evs-and-afvs-registrations/). By picking the monthly data from Jan 2019 to Apr 2024, a time series consisting of 64 figures is generated. The modelling process is performed using R.

3.2. Preparing and Processing Data

3.2.1. Plotting the Data as a Time Series

After plotting the time series, as displayed in Figure 1, a clear upward trend with seasonality could be seen. The sales figures tend to peak in March and September, however, remains low in the rest of the year. This is due to UK's unique number plate registration system, which issues new number plates in March and September, and this leads to more customers purchasing cars as they can get an up-to-date number plate. As a result of this obvious seasonality of lag 12 and upward trend, it's necessary to perform differencing to the time series to achieve stationarity.

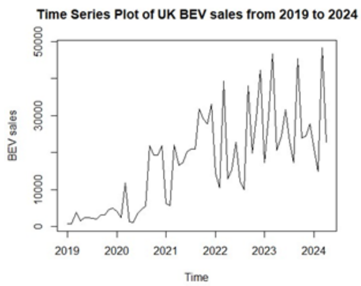

Figure 1: Time series plot of UK's BEV sales from 2019 to 2024

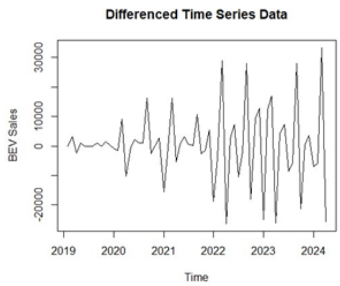

Figure 2: Differenced Time Series Data of UK's BEV sales from 2019 to 2024

DOI: 10.54254/2754-1169/126/2024.18340

3.2.2. Stationarity Test and Differencing Time Series

Conducting the ADF root test to original time series yields a p-value of 0.11 (see Table 1), larger than 0.05. This numerically indicates that differencing is indeed required. The varying trend in sales can be removed with first-order differencing with lag = 12 using formula $Y_t' = Y_t - Y_{t-12}$. After performing differencing, the curve is now fluctuating around the horizontal line of coordinate y=0, just as displayed in Figure 2. Another ADF test for differenced time series yields a p-value of 0.01, which is within the range of 0.05 (see Table 2).

Table 1: ADF test before differencing

Statistic	Dickey-Fuller	Lag order	p-value
Value	-3.1246	3	0.1193

Table 2: ADF test after differencing

Statistic	Dickey-Fuller	Lag order	p-value
Value	-4.3398	3	0.01

Table 3: Ljung Box white noise check

Data	Q*	df	p-value
Residuals from ARIMA (2,1,0)	40.784	11	2.625e-05

3.2.3. White Noise Check

After conducting the stationarity test, the Ljung-Box test is performed to ensure that the differenced time series is not white noise. The p-value is almost 0, less than 0.05, indicates that it is not white noise as can be seen from Table 3.

3.3. Find the Best Fit Model and Forecast

The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots can give a rough estimate for the parameters of the ARIMA and SARIMA models. From the PACF plot (Figure 3), all lags are insignificant, suggesting p=0. The ACF plot shows a spike at lag 1, with subsequent lags being insignificant, indicating q=1. Meanwhile, the ACF plot (Figure 4) does not show significant spikes in seasonal lag multiples, suggesting Q=0. It is certain that d = 1, D = 1, and s = 12.

With the help of auto. Arima function in R we can finally determine the choice of SARIMA (0,1,1)(1,1,0)(12). Similarly, the choice of ARIMA (2,1,0) is obtained.

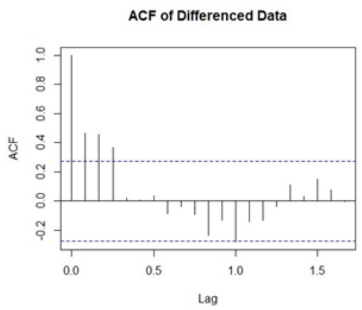

Figure 3: ACF of differenced time series

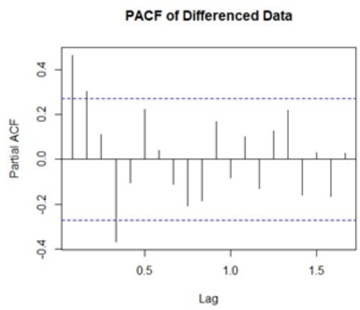

Figure 4: PACF of differenced time series

Use the plot feature of R to draw the forecast of UK's BEV sales figures in ARIMA and SARIMA model (Figure 5 and 6). The black line represents the historical data of UK BEV sales in the past while the blue line represents the forecasted value. The grey shaded area represents the confidence interval indicating the range of uncertainty in the prediction.

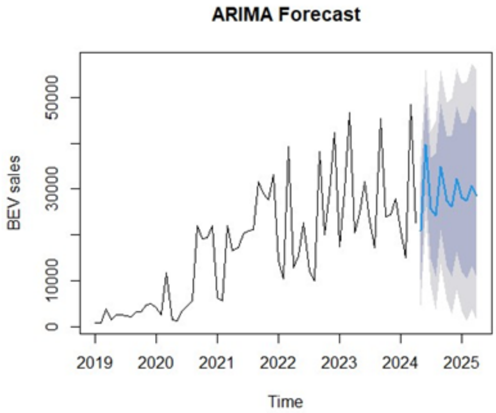

Figure 5: ARIMA Forecast of UK BEV market

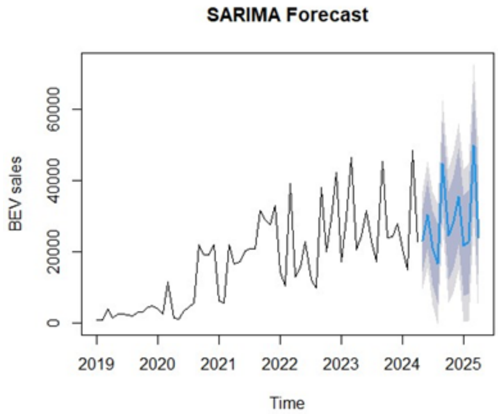

Figure 6: SARIMA Forecast of UK BEV market

3.4. Comparison between ARIMA and SARIMA models

From the diagnoses of the ARIMA and SARIMA models (Table 4 and Table 5), it can be seen that MAE, RMSE and MAPE of the SARIMA model are all smaller than those of ARIMA. In this case, this indicates a superior fitting and forecasting performance of the SARIMA model, being consistent with the strong seasonality of UK's BEV sales figures.

Table 4: Diagnosis of ARIMA model

ME	RMSE	MAE	MPE	MAPE	MASE
1078.608	8084.906	5935.63	-11.631	47.80383	0.7135

Table 5: Diagnosis of SARIMA model

ME	RMSE	MAE	MPE	MAPE	MASE
-72.43848	6102.027	4431.521	-17.74743	36.37715	0.5327

4. Conclusion

The successful forecast of UK's BEV sales has shown time series methods to be a handy tool with trended figures with seasonality. In this case, SARIMA has shown better forecast performance due to its consideration of seasonality features. The forecast output indicates that the sales figures of UK's BEVs will still show an upward trend in the coming few months, going along with significant fluctuations between March and September. This forecast aligns with automotive market future trend as well as UK's environmental policies which promotes zero emission technologies. As indicated from the forecast, the car manufacturers could devote more to the research of BEVs and accelerate technology upgrades, such as designing more durable batteries and faster charging methods, so as to attract more customers to choose BEVs. The UK government should also keep on making more incentive policies to encourage people choosing BEVs, meanwhile keeping pace on building charging infrastructures to meet the daily growing needs.

References

[1] *Department for Transport (2023). Transport statistics great Britain. Retrieved from: https://www.gov.uk/government/statistics/transport-statistics-great-britain-2023. Accessed June 26, 2024.*
[2] *IEA (2019). Tracking transport. Retrieved from: https://origin.iea.org/reports/renewables-2019/transport. Accessed June 26, 2024.*
[3] *Kovacs, Z. (2019). The end of the fossil fuel car is on the eu agenda. Retrieved from https://www.transportenvironment.org/articles/end-fossil-fuel-car-eu-agenda. Accessed June 26, 2014.*
[4] *Miao, H., Tang, C., and Luo, L. (2020). Forecasting new energy vehicle sales based on the ARIMA model. Enterprise Science and Technology & Development, 10, 97–98.*
[5] *Küfeoğlu, S., & Hong, D. K. K. (2020). Emissions performance of electric vehicles: A case study from the United Kingdom. Applied Energy, 260, 114241.*
[6] *Hidrue, M.K., Parsons, G.R., Kempton, W., and Gardner, M.P. (2011). Willingness to pay for electric vehicles and their attributes. Resource & Energy Economics, 33(3), 686–705.*
[7] *Society of Motor Manufacturers and Traders December 2023 new car registrations. Retrieved from https://media.smmt.co.uk/december-2023-new-car-registrations/. Accessed July 17, 2014.*
[8] *Guo, Z.J., Gong, H., and Zhou, L.J. (2022). Application of SARIMA Model and Holt-Winters Exponential Smoothing Method in Forecasting Tuberculosis Incidence in Jiangsu Province. Disease Surveillance, 37(08), 1042–1047.*
[9] *Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.*
[10] *Contreras, J., Espinola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA models to predict next-day electricity prices. IEEE Transactions on Power Systems, 18(3), 1014-1020.*