Truc Nguyen
Econ 138 - Spring 2021
Professor Sanchita Mukherjee
Note: Thank my teammates, Mai Vu and Chih-yu Huang(Tony) for their contributions to my term paper.

Sales of Electric Vehicles in the US Forecast to 2030

Abstract: We use ARIMA model to forecast the sale of electric vehicles (EVs) in the US in the next 10 years based on the previous sales from January 2011 to December 2020. Our main findings are that monthly EVs' sales will not exceed 40,000 units sold in 2030 and that the EV sales in 2030 will increase by 34.5% when compared to the sales in 2020.

Introduction:

We are interested in forecasting sales of EVs because of their great potential in reducing global warming emissions and embracing environmental sustainability. According to "Fact Sheet: The American Jobs Plan" (2021) published by the White House, the government plans to build a network of 500,000 EV chargers by 2030, achieve 100% carbon-free electricity by 2035, and net-zero emission by 2050, and make electric vehicles a major focus. Thus, in the next years, we expect to see an increase in sales of EVs due to President Biden's plans to fight climate change. The purpose of this project is to forecast the sale of EVs in the US in the next 10 years based on previous monthly sales during January 2011 - December 2020. Our approach is comparing between an ARIMA and ETS model to check which one is the best model to forecast and proceed with that model.

Several existing studies have addressed the methods to forecast EV sales. An Iowa University's study uses a multiple linear regression model based on four corresponding explanatory factors: average fuel cost savings (difference in average fuel cost between EVs and conventional vehicles (CVs)) in months, average tax credits for EVs in months, the average price difference between EVs and CVs in months, and numbers of EVs models in months (Duan, Gutierrez, and Wang, 2014). Another study published by Simon Frazer University (Wolinetz and Axsen, 2017) examines a latent discrete model to provide a prediction for three situations: no-policy, supply-focused policy, and demand-focused policy. The model predicts that in 2030, the sales share will be higher when the government applies supply-focused policy than demand-focused policy by around 2-17%, and both of those situations achieve a much higher result in annual sales than the "no-policy" situations by 4 times. Both studies emphasize the importance of policy from the government in the sale of EVs. Our study's limitation is not taking new policies into the forecast because they are an unknown variable.

EV Sales Forecast Model:

Our data is compiled by the Alliance for Automotive Innovation provided by IHS Markit and Hedges & Co. Data is collected by recording the number of registered electric vehicles in the US each month from January 2011 to December 2020 (Table 1). Electric vehicles include vehicles running on electricity from hydrogen fuel, and running exclusively on electricity. Table 2 provides a codebook of data. Table 3 shows descriptive statistics for the sales. The time plot of the data from Figure 1 shows that there is an increasing trend for sales from 2011 to 2020. We do not observe the seasonal pattern in this time plot.

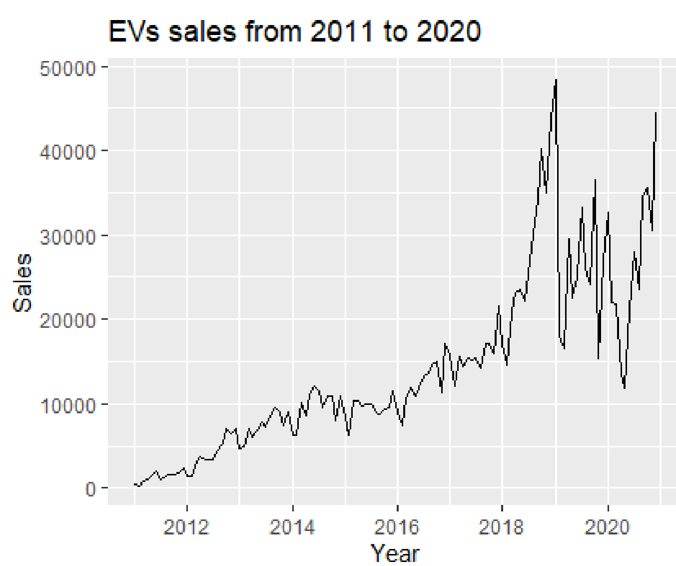


Figure 1. Time plot of EV sales

First, we compare between ARIMA model and ETS model to choose the best fit one. To do this, we split data into training and testing sets, two partitions are January 2011 - December 2019 and January 2020 - December 2020. Because the EV sales experience a huge increase at the end of 2018, if we use data in 2018 and 2019 in the testing set, the forecast will be overestimated, we choose to include 2018 and 2019 in the training set. We use ARIMA() function to fit the data and get ARIMA$((0,1,5)(0,0,1))_{12}$ as the best estimated ARIMA model. The model includes a non-seasonal MA(5) term, a seasonal MA(1) term, one first difference, no AR terms, and the seasonal period is S = 12. The residuals as shown in [Figure 2](#) appear white noise and are distributed as a normal distribution. The model also passes the Ljung-box test with p-value of 0.146. We fit the ARIMA model to the data to check training against the testing set.

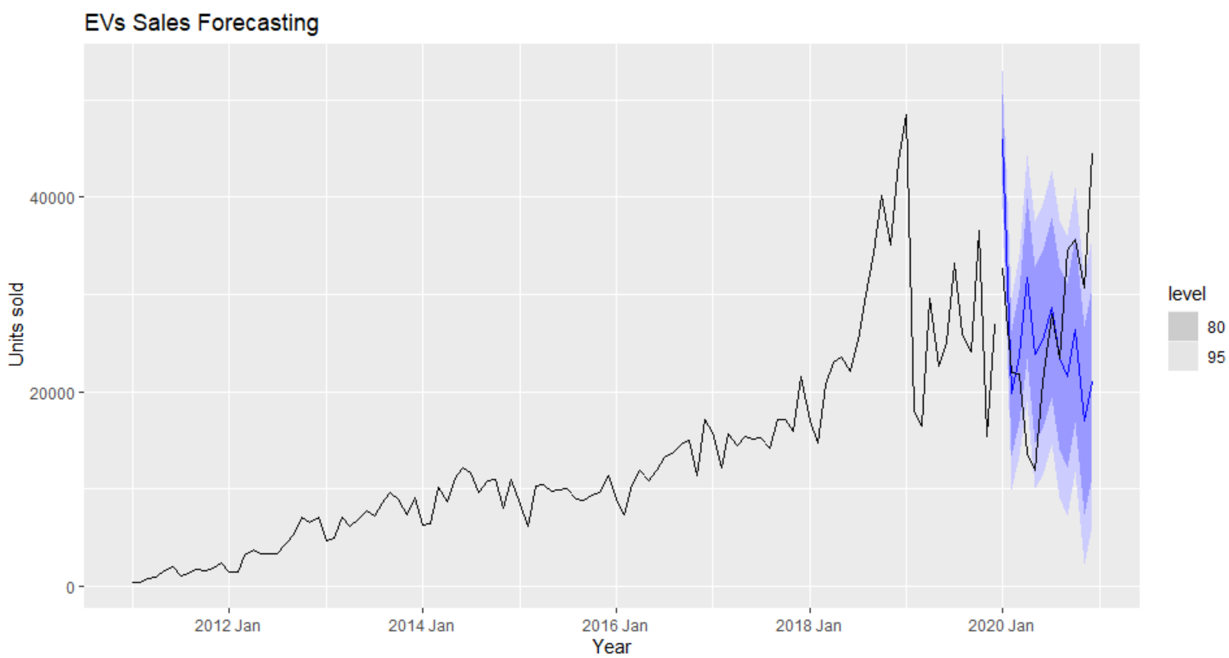


Figure 3. EV Sales Forecasting training against testing set

The ARIMA model does a good job in forecasting the trend and seasonal pattern of the EVs sale. Because of the decrease in sales in 2019, the model predicts a slight decrease in sales in 2020. The model predicts an increase following a decrease in sales for Spring 2020 because of the pattern of the previous data. However, because of Covid 19, the sales went downhill in Spring 2020, which the ARIMA is unable to capture. The ARIMA model forecast correctly the sales of EVs for the middle of the year. One of the reasons for the huge increase in EV sales at the end of the year is Governor of California, one of the place having the largest EV consumption in the US, Galvin Newsom set “a goal to end sales of pollution-spewing light-duty vehicles by 2035" (Weisbrod, December 2020). The ARIMA model fails to cover new policies in its forecast.

When we fit an ETS model, the residuals however are not white noise by the Ljung-box test (p-value of 0.0173). There are more spikes in the ACF plots from ETS(M, A, N) model than in the ARIMA model, as shown in Figure 4. The residuals do not appear white noise in this case. So, we choose the ARIMA model over the ETS model.

To fit in the ARIMA model, we will do one first difference as the `ndriffs()` function from R suggests, to obtain stationary data. Figure 5 shows the time plot before and after differencing the data.

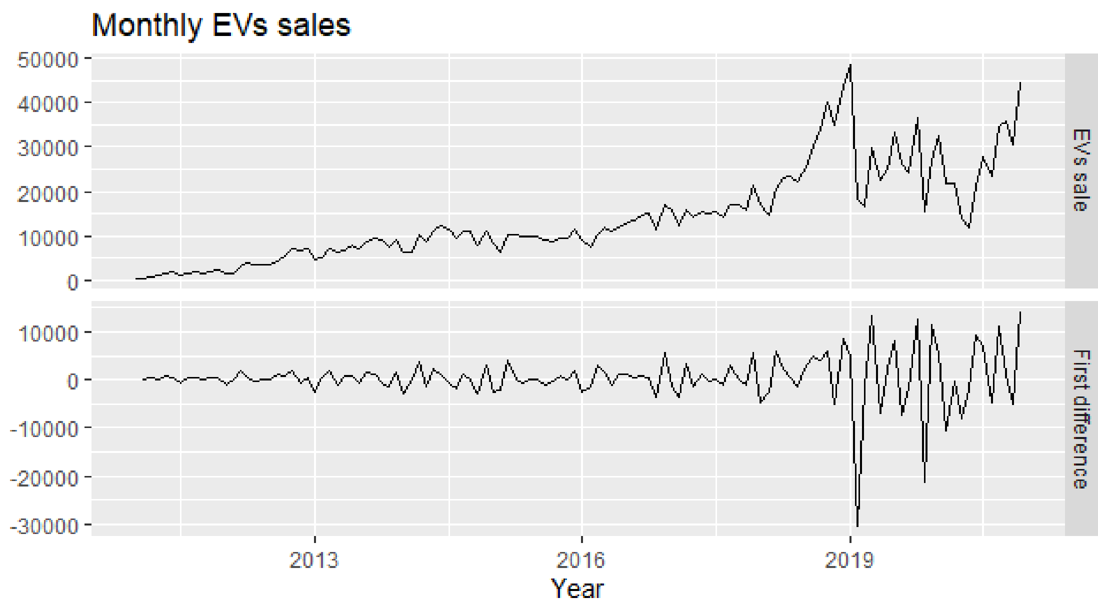


Figure 5. Top panel: Time plot of EV Sales from 2011 to 2020; bottom pane: Time plot after one first difference

The ACF plot shows the autocorrelations which measure the relationship between y_t and y_{t-k} for different k-values. The PACF plot shows the relationship between y_t and y_{t-k} after removing the effect of lags 1, 2, 3, …, k-1 (Hyndman & Athanasopoulos, 2018). Figure 6 provides ACF and PACF plots of differenced data. We use `ARIMA()` function to find the best fit ARIMA models: ARIMA(0,1,4)(0,0,2)$_{12}$ and ARIMA((0,1,4)(2,0,0))$_{12}$.

For ARIMA(0,1,4)(0,0,2)$_{12}$, the model includes a non-seasonal MA(4) term, a seasonal MA(2) term, one first difference, no AR terms, and the seasonal period is S = 12. The model could be written as :

$$(1 - B)Y_t = (1 - \Theta_1 B^{12} - \Theta_2 B^{24})(1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \theta_4 B^4)\varepsilon_t$$

For ARIMA((0,1,4)(2,0,0))$_{12}$, the model includes a non-seasonal MA(4) term, a seasonal AR(2) term, one first difference, and the seasonal period is S = 12.

Similarly, the model could be written as:

$$(1 - B)(1 - \Phi_1 B^{12} - \Phi_2 B^{24})Y_t = (1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \theta_4 B^4)\varepsilon_t$$

The AICc for ARIMA((0,1,4)(2,0,0))$_{12}$ is 2328, which is smaller than the AICc for ARIMA(0,1,4)(0,0,2)$_{12}$ (2332). Figure 7 shows the residuals from ARIMA(0,1,4)(2,0,0)$_{12}$ Model. The ACF plot looks good, and the residuals appear white noise and distribute as a normal distribution. We check with the Ljung-box test to ensure the residuals are white noise. The p-value for the Ljung-box test is 0.118, thus we fail to reject the null hypothesis, and there is enough evidence to conclude the residuals are white noise.

Now, we have a seasonal ARIMA model that is ready for forecasting. We forecast for the next 10 years, and the forecast is shown in Figure 8.

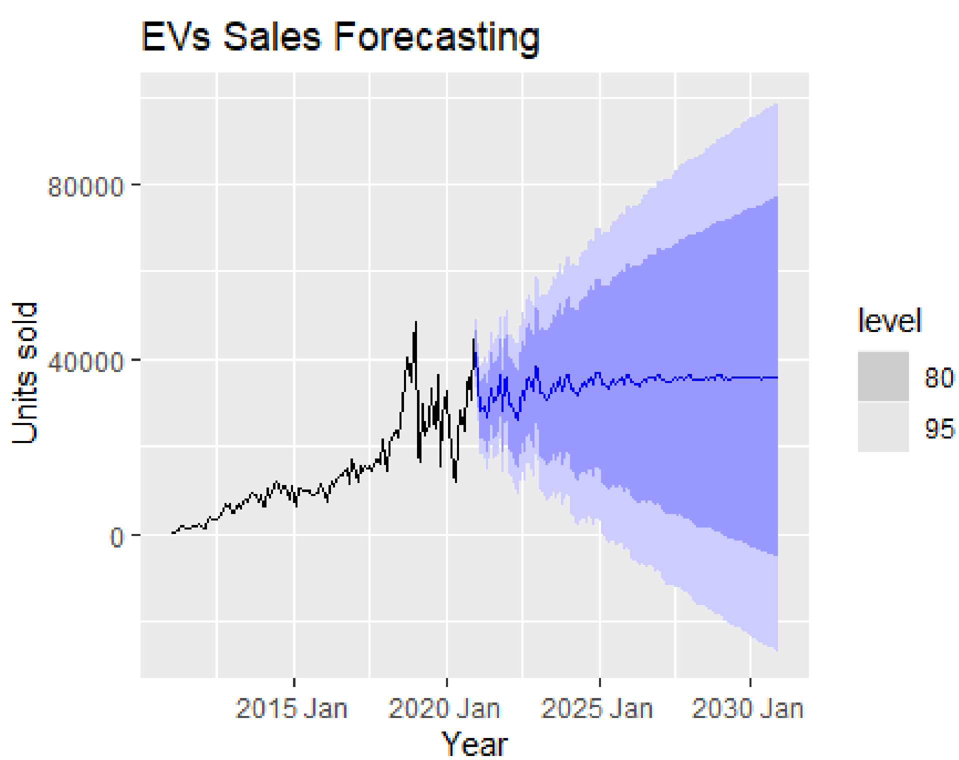


Figure 8. EV Sales Forecasting using ARIMA((0,1,4)(0,0,2))$_{12}$ model

The model does a good job to catch on to the trend and seasonal pattern of the previous data. Because there will be new policies to encourage EVs consumption, so the ARIMA model may not do well to cover those factors.

Results and Discussion:

Based on the ARIMA model, the total EV sales for the year 2030 is 429761, increasing by 34.5% compared with the sale for the year 2020 (319611 units sold). Our forecast's result is less optimistic than other forecasts. For example, the forecast from Fraunhofer Institute for Systems and Innovation Research (Dallinger, et. al., 2013) predicts that the EV sales for only California in 2030 will reach over 1 million units sold. The reason for our conservative result is that we predict solely on the previous data and do not take into other scenarios, such as new supporting policies, new tax incentives, or an increase in supply, into account.

Another main finding is the monthly sales of EVs will not exceeding 40,000 units. However, with the upcoming plans from the recent Government, we believe the monthly sales will reach above 40,000 units.

Many unknown variables may positively increase the sales of EVs. However, in general, all EV sales forecasts agree that EVs will experience a huge increase in the number of units sold in the next future.

References:

Dallinger D., Gerda S., Wietschel M., "Integration of intermittent renewable power supply using grid-connected vehicles – A 2030 case study for California and Germany", Applied Energy, Volume 104, 2013, Pages 666-682, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2012.10.065.

Duan Z., Gutierrez B., and Wang L., "Forecasting Plug-In Electric Vehicle Sales and the Diurnal Recharging Load Curve," in IEEE Transactions on Smart Grid, vol. 5, no. 1, pp. 527-535, Jan. 2014, DOI: 10.1109/TSG.2013.2294436.

Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2

The United States Government. (2021, May 4). *FACT SHEET: The American Jobs Plan*. The White House. https://www.whitehouse.gov/briefing-room/statements-releases/2021/03/31/fact-sheet-the-american-jobs-plan/.

Weisbrod, K., "Was 2020 The Year That EVs Hit It Big? Almost, But Not Quite." Inside Climate News, 23 Dec. 2020, insideclimatenews.org/news/21122020/electric-vehicles-in-2020.

Wolinetz M., and Axsen J., "How Policy Can Build the Plug-in Electric Vehicle Market: Insights from the REspondent-Based Preference And Constraints (REPAC) Model." Technological Forecasting & Social Change, vol. 117, 2017, pp. 238–250.

Tables and Figures:

Date	Sales (units sold)
Jan 2011	428
Feb 2011	367
.....	......
Nov 2020	30,606
Dec 2020	44,522

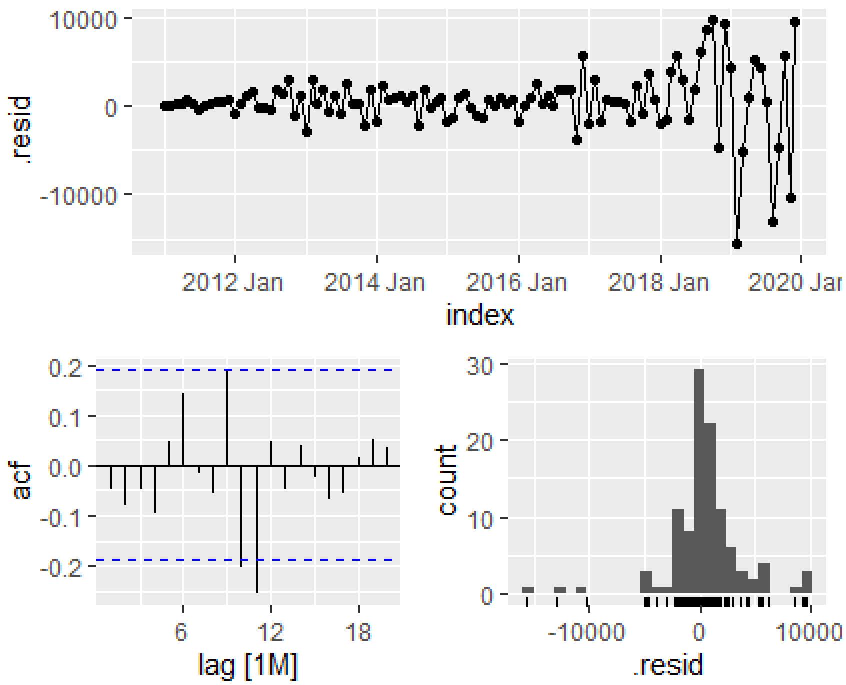
Table 1. Sales data by AAI

Variable name	Description
Date	The first date of every month to represent each month of the year
Sales	Number of electric vehicles (units) sold in each month

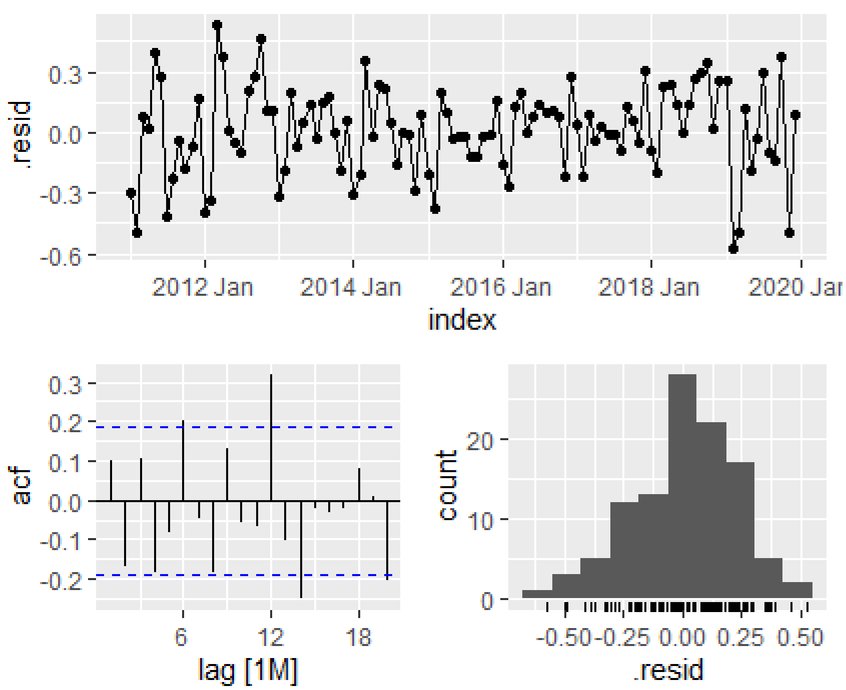
Table 2. Codebook for the Data

Total units sold	Mean	Standard Deviation	Median	Minimum	Maximum
1,693,906	14,116	10,595.49	11,060	367	48,454

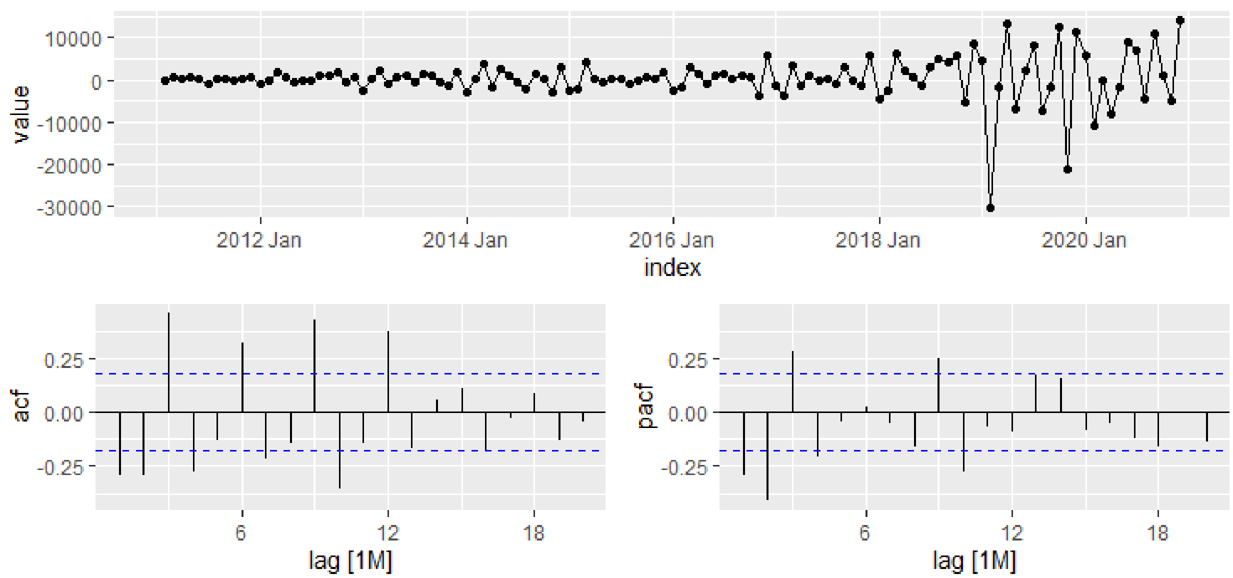
Table 3. Descriptive Statistics for Sales Data



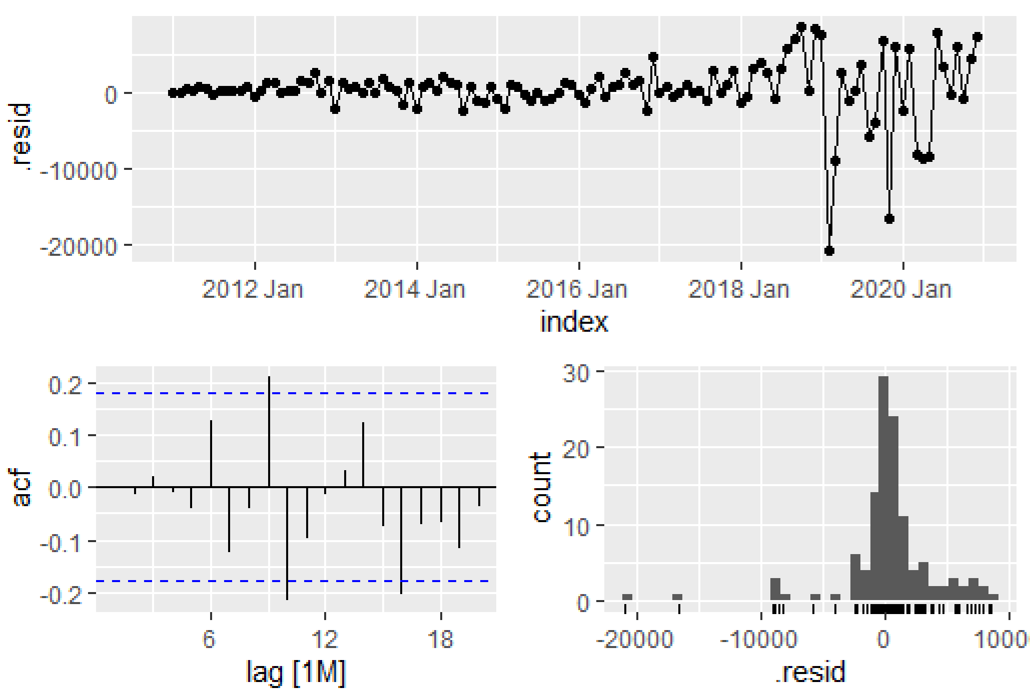
[Figure 2](#). Residuals from ARIMA((0,1,5)(1,0,0))$_{12}$ applied to the training set



[Figure 4](#). Residuals from ETS(M, A, N)



[Figure 6.](#) Top panel: Time plot after one first difference; left panel: ACF plot, right panel: PACF plot



[Figure 7](#). Residuals from ARIMA((0,0,4)(2,0,0)$_{12}$ model applied to EVs sale