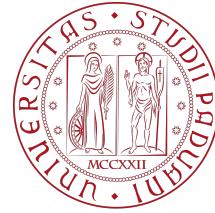




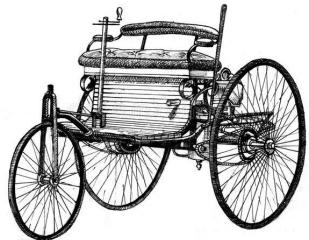
BUSINESS, ECONOMIC AND FINANCIAL DATA



Market of electric and non-electric cars in USA: analysis and forecasting

ALBERTO SINIGAGLIA, BEATRICE SOFIA BERTIPAGLIA, CHIARA COLATO

USA car market

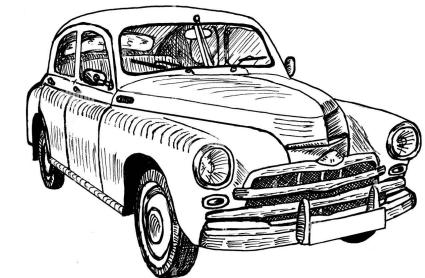


The first **electric car** was created in 1884, the first **gasoline-powered car** in 1885. In the first years of the **1900** in USA 40% of cars were powered by steam, **38%** by electricity, and only **22%** by gasoline.

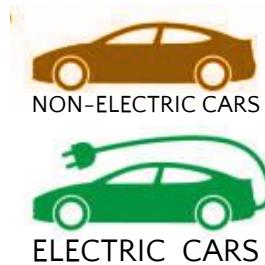
The **automotive industry** in the United States of America began in the **1890s**. USA was the first country in the world to have a mass market for car production and sales.

In **1935 electric cars disappeared** because gasoline-powered cars became easy to drive, cheaper to buy, and the discovery of Texas crude oil made gas very cheap.

Currently USA is second in the world for number of cars produced by year, more than 9 millions in 2021 (total sales were more than 15 millions), following China.

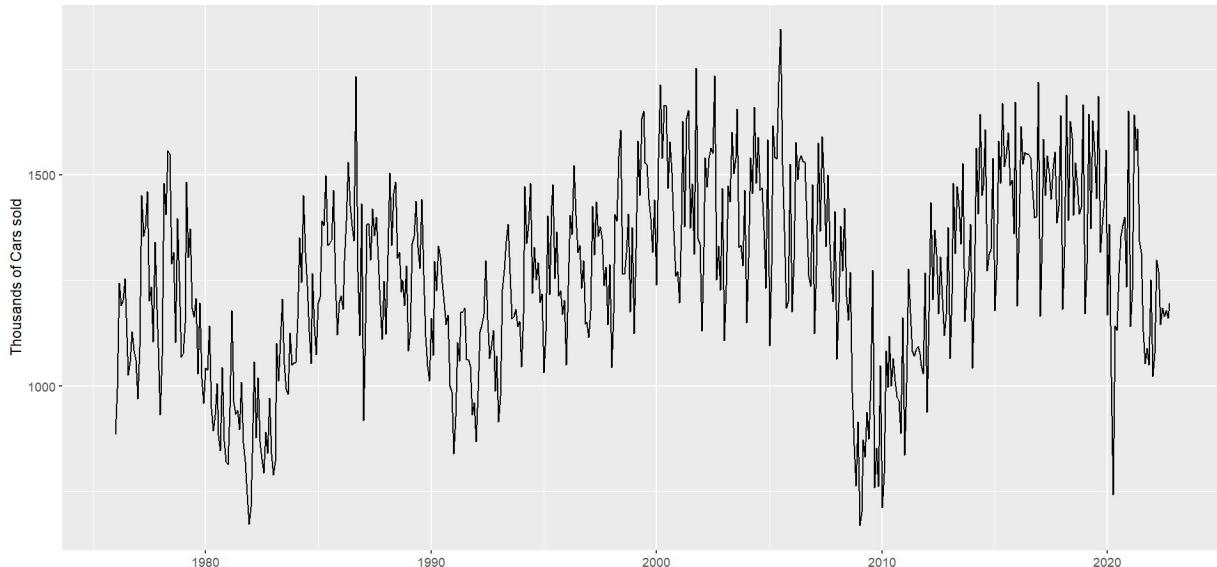


Recently the **interest in electric cars** is coming back, caused by higher oil prices and carbon pollution, but the percentage of electric cars is only **9%** of total sales.



Aim: **analysis and forecasting** of the time series of car sales in USA

Time series of
electric and
non-electric car
sales in USA
(1976-2022)



1980: USA economy crisis with high inflation, high unemployment, and high interest rates. The automakers suffered large losses.
1985: oil prices had fallen sharply, helping the revitalization of the USA auto industry.

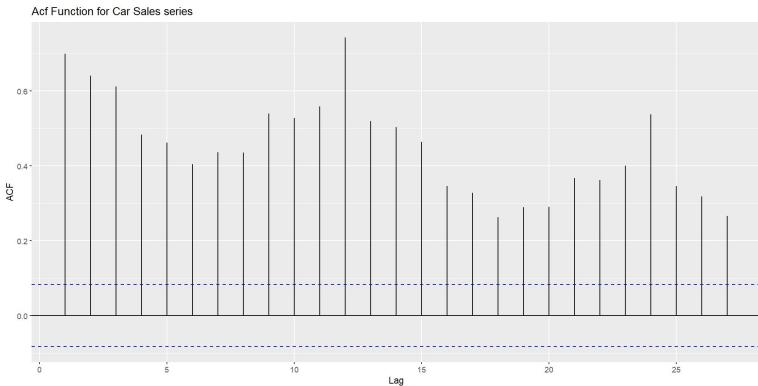
2008: automotive industry crisis caused by the 2007-2008 financial crisis and the resulting Great Recession.
2009: government assistance helped the recovery.

2020: covid.

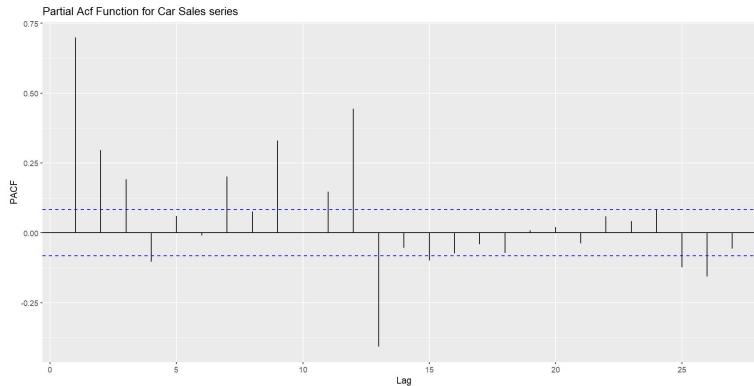


Exploratory analysis

Autocorrelation function



Partial autocorrelation function



There is a strong seasonality.

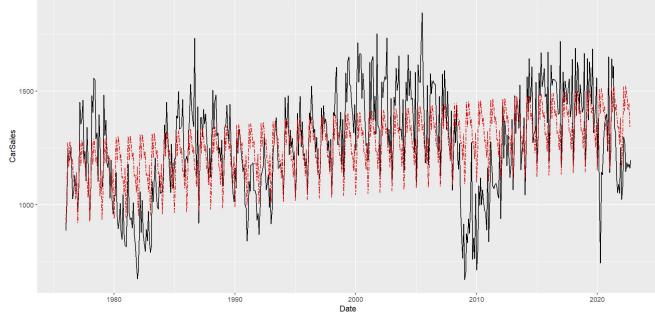
Car sales for the next month are highly correlated with the current month car sales.

Car sales of the same month but one year lagged are highly correlated.

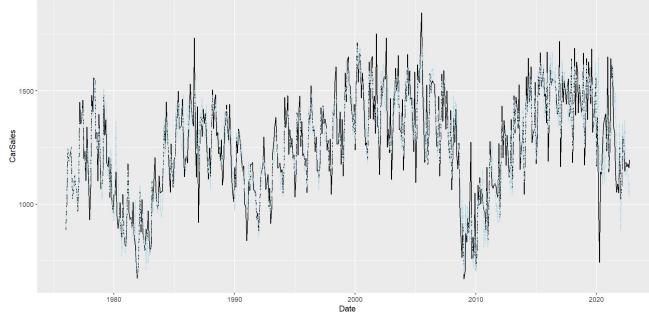


Modelling the time serie

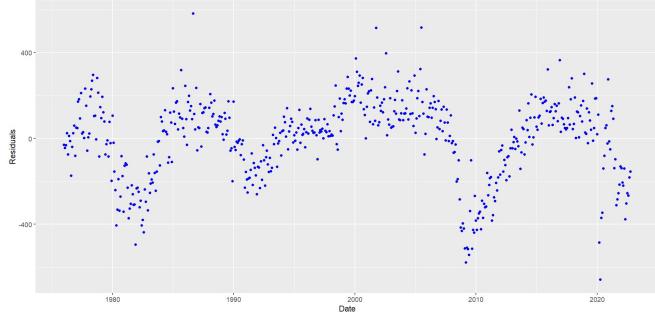
Real Car Sales values VS Fitted Values with TSLM



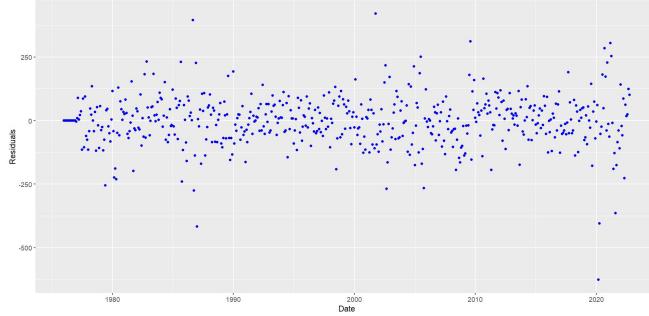
Real Car Sales values VS Fitted Values with Arima



Residuals of TSLM with Trend and Seasonality

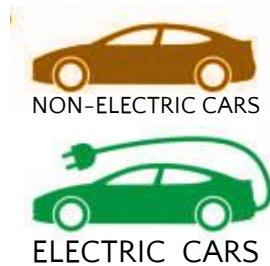


Residuals of Arima



First analysis of car total sales time series with **TSLM** and **ARIMA** models, without other explanatory variables.

A Bass model is not fitted because the assumption that there is a product growth is not shown by the plot.



Modelling car sales

An explanatory variable is added: electric car sales time series.

Car total sales time series is divided in **electric car sales** and **non-electric car sales** time series.

Focus on the time period **2012-2022**, both for compatibility with e-car sales time series and to avoid the presence of economic cycles.

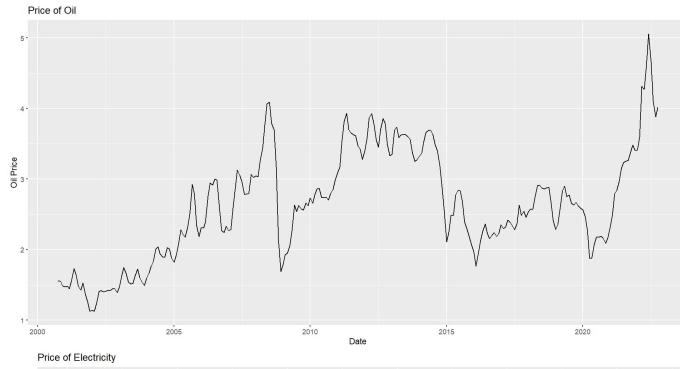




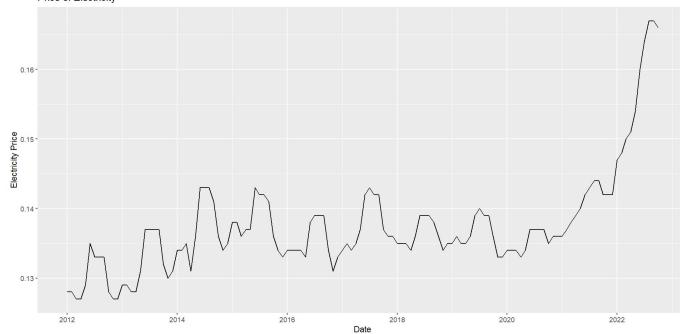
Modelling car sales

Other explanatory variables are added:

Oil
prices



Electricity
prices





Modelling car sales



TSLM models

(to get an idea of the relations between variables)

```
(Intercept) **  
trend ***  
season2 **  
season3 ***  
season4 ***  
season5 ***  
season6 ***  
season7 ***  
season8 ***  
season9 ***  
season10 ***  
season11 ***  
season12 ***  
oil_prices_12$OilPrice ***  
ecar_sales_12$ECarSales  
electricity_prices$ElectricityPrice  
---  
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 136.1 on 114 degrees of freedom  
Multiple R-squared: 0.5802, Adjusted R-squared: 0.525  
F-statistic: 10.5 on 15 and 114 DF, p-value: 2.149e-15
```

```
(Intercept) **  
trend ***  
season2 **  
season3 ***  
season4 ***  
season5 ***  
season6 ***  
season7 ***  
season8 ***  
season9 ***  
season10 ***  
season11 ***  
season12 ***  
oil_prices_12$OilPrice ***  
electricity_prices$ElectricityPrice .  
---  
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 136.4 on 115 degrees of freedom  
Multiple R-squared: 0.5749, Adjusted R-squared: 0.5232  
F-statistic: 11.11 on 14 and 115 DF, p-value: 1.207e-15
```

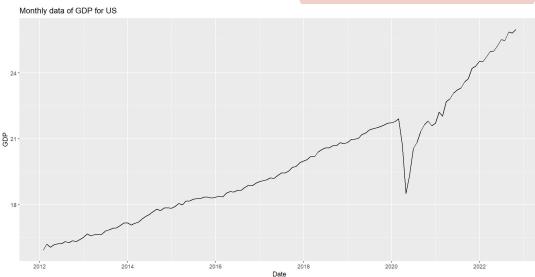
Those are good models because almost all the coefficients are significant, but they are not great models because the R-squared value is low.



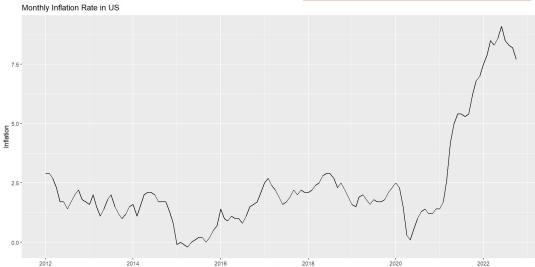
Other explanatory variables

added in order to have more informations:

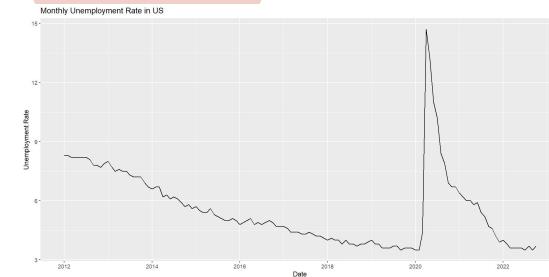
GDP



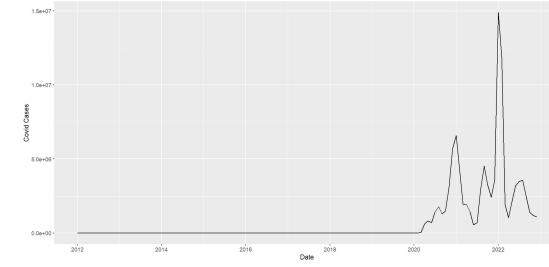
Inflation rate



Unemployment rate



Covid cases



Correlations between explanatory variables

	OilPrice	ElectricityPrice	GDP	Inflation	UnRate	CovidCases	ECarSales	NECarSales
OilPrice	1.00	0.36	0.09	0.62	0.04	0.18	0.45	-0.35
ElectricityPrice	0.36	1.00	0.72	0.72	-0.42	0.41	0.72	-0.23
GDP	0.09	0.72	1.00	0.75	-0.52	0.56	0.85	-0.28
Inflation	0.62	0.72	0.75	1.00	-0.36	0.55	0.85	-0.46
UnRate	0.04	-0.42	-0.52	-0.36	1.00	-0.09	-0.29	-0.27
CovidCases	0.18	0.41	0.56	0.55	-0.09	1.00	0.58	-0.42
ECarSales	0.45	0.72	0.85	0.85	-0.29	0.58	1.00	-0.29
NECarSales	-0.35	-0.23	-0.28	-0.46	-0.27	-0.42	-0.29	1.00



Best TSLM model

```

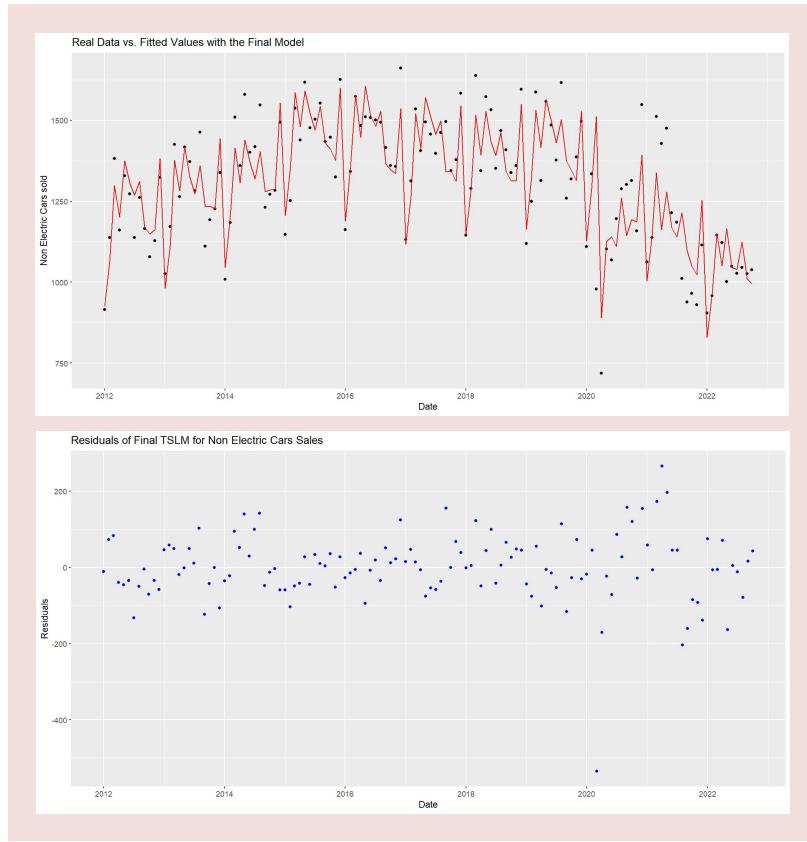
Call:
tslm(formula = necar_sales_12_ts ~ trend + season + OilPrice +
    Inflation + UnRate, data = variables)

Residuals:
    Min      1Q  Median      3Q     Max 
-534.46 -44.51  -2.72  47.06 266.33 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1671.4275   85.0637 19.649 < 2e-16 ***
trend        -1.2869    0.4721 -2.726 0.007420 **  
season2       149.8589   41.1806  3.639 0.000413 *** 
season3       384.2330   41.7785  9.197 2.06e-15 *** 
season4       275.3584   42.0428  6.549 1.73e-09 *** 
season5       420.0725   42.5574  9.871 < 2e-16 *** 
season6       342.3448   42.6997  8.018 1.04e-12 *** 
season7       289.1704   42.4637  6.810 4.83e-10 *** 
season8       351.5727   42.1994  8.331 2.02e-13 *** 
season9       212.9398   42.1996  5.046 1.72e-06 *** 
season10      197.8756   41.9942  4.712 6.98e-06 *** 
season11      176.8089   42.4452  4.166 6.07e-05 *** 
season12      397.2953   42.2330  9.487 6.69e-16 *** 
OilPrice      -48.0136   28.8037 -1.667 0.098274 .  
Inflation     -39.6900   11.1297 -3.566 0.000531 *** 
UnRate        -56.2292   5.0963 -11.033 < 2e-16 *** 
...
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 96.43 on 114 degrees of freedom
Multiple R-squared:  0.7894,  Adjusted R-squared:  0.7616 
F-statistic: 28.48 on 15 and 114 DF,  p-value: < 2.2e-16

```



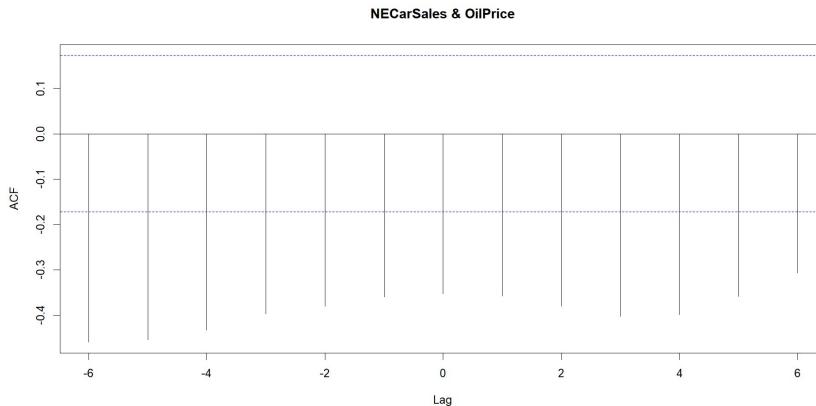
Variables removed one by one in order to have the best combination of significant coefficients.

Like before almost all the coefficients are significant, but this time the R-squared value is greater.



Best model for non-electric car sales: MODEL WITH LAGGED REGRESSORS

For each explanatory variable the positive lag that gives the greatest correlation with non-electric cars is found.



The “lagged” dataset is created:

- **non-electric car sales**
- electric car sales (4 months lagged)
- oil prices (3 months lagged)
- electricity prices (1 month lagged)
- GDP (4 months lagged)
- inflation rate (2 months lagged)
- unemployment rate (not lagged)
- covid cases (1 month lagged)



Best model for non-electric car sales: MODEL WITH LAGGED REGRESSORS

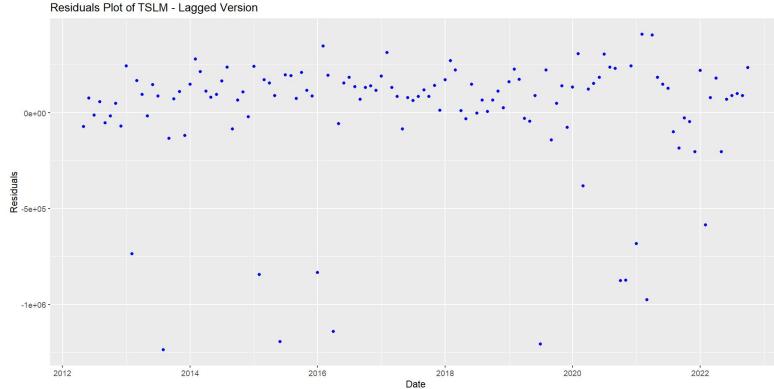
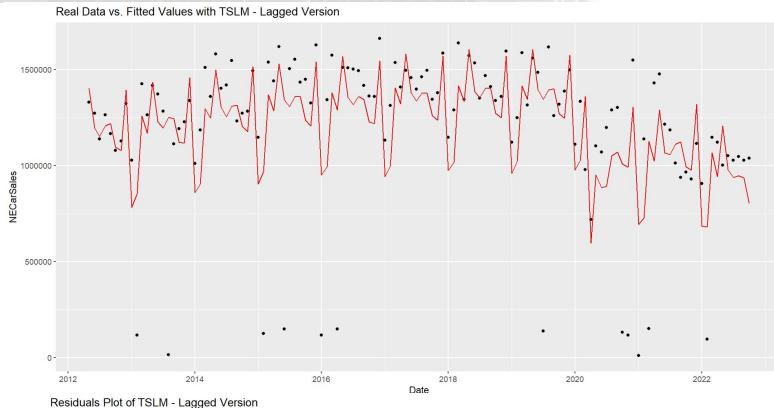
TSLM model

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.946e+06  1.785e+05 10.902 < 2e-16 ***
trend       -1.231e+03  1.166e+03 -1.056 0.293306  
season2     -2.054e+05  1.512e+05 -1.359 0.176851  
season3     -2.486e+05  1.512e+05 -1.644 0.103034  
season4     -1.989e+05  1.514e+05 -1.314 0.191648  
season5     -2.045e+05  1.516e+05 -1.348 0.180283  
season6     -3.266e+05  1.518e+05 -2.152 0.033573 *  
season7     -3.485e+05  1.558e+05 -2.237 0.027286 *  
season8     -2.049e+04  1.559e+05 -0.131 0.896110  
season9     -6.229e+05  1.558e+05 -3.997 0.000116 *** 
season10    -5.715e+05  1.559e+05 -3.666 0.000379 *** 
season11    -1.782e+05  1.558e+05 -1.143 0.255370  
season12    -2.592e+05  1.549e+05 -1.673 0.097104 .  
UnRate      -6.628e+04  1.818e+04 -3.642 0.000413 *** 
CovidCases_1 -1.456e-04  5.833e-05 -2.497 0.013991 *  
...
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 354400 on 111 degrees of freedom
 Multiple R-squared: 0.311, Adjusted R-squared: 0.2241
 F-statistic: 3.578 on 14 and 111 DF, p-value: 7.169e-05

Explanatory variables removed one by one in order to have the best combination of significant coefficients.

This model is way worse than the others with not lagged dataset.





Best model for non-electric car sales: MODEL WITH LAGGED REGRESSORS

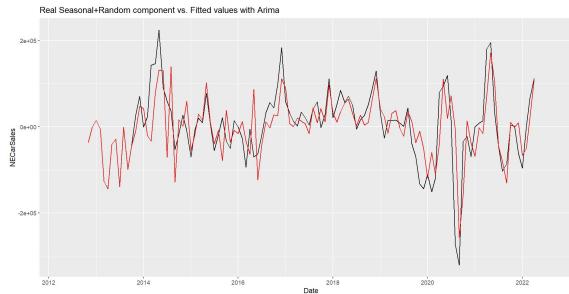
Composed model :

TREND

(TSLM model
with explanatory
variables)

SEASONALITY
+ RANDOM

(arima model)



OilPrice_3	-6.305e+04	2.097e+04	-3.007	0.003293	**
ElectricityPrice_1	8.312e+06	2.295e+06	3.621	0.000452	***
GDP_4	-1.894e+04	7.351e+03	-2.576	0.011358	*
Inflation_2	8.709e+04	1.180e+04	7.381	3.70e-11	***
UnRate	-3.419e+04	4.690e+03	-7.290	5.82e-11	***
CovidCases_1	-3.050e-04	4.361e-05	-6.994	2.50e-10	***
ECarSales_4	-2.367e+00	6.035e-01	-3.922	0.000156	***

Residual standard error: 79300 on 106 degrees of freedom
(12 osservazioni eliminate a causa di valori mancanti)
Multiple R-squared: 0.8025, Adjusted R-squared: 0.7895
F-statistic: 61.53 on 7 and 106 DF, p-value: < 2.2e-16

Series: res_season_random
ARIMA(1,0,2)(1,1,1)[12]

Coefficients:

	ar1	ma1	ma2	sar1	sma1
0.6080	0.2416	-0.2263	0.2102	-0.9961	
s.e.	0.2482	0.2751	0.2199	0.1183	1.0038

sigma^2 = 3.855e+09: log likelihood = -1279.34
AIC=2570.68 AICc=2571.56 BIC=2586.43

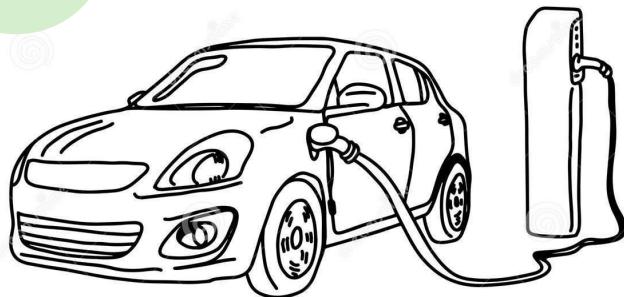


Focus on ELECTRIC CARS : Why?

More interesting
to analyze future
developments

Current topic,
it is a growing
market

Emissions
are reduced



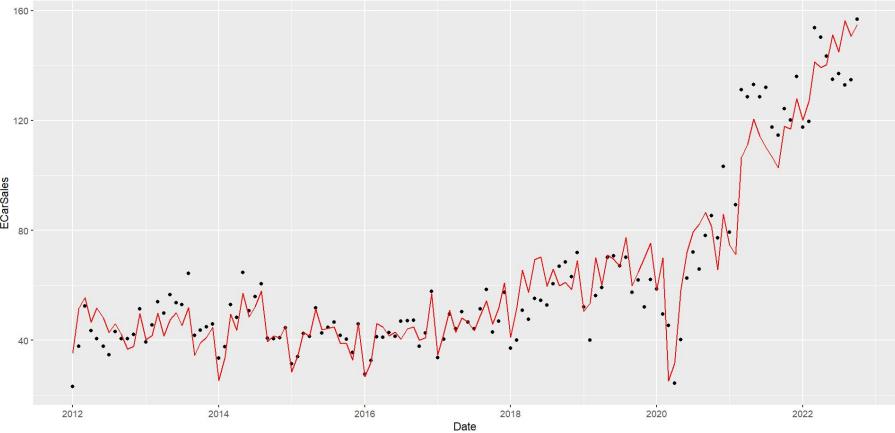
Lower costs
for a mile

Quiet
motors



ELECTRIC CARS

Real ECarSales vs. Fitted values with TSLM



TSLM model

trend	-1.100e+00	1.421e-01	-7.737	3.05e-12	***
GDP	2.752e+01	2.614e+00	10.525	< 2e-16	***
Inflation	3.625e+00	1.003e+00	3.615	0.000434	***
UnRate	8.244e+00	6.525e-01	12.634	< 2e-16	***
NECarSales	4.454e-02	5.308e-03	8.391	9.16e-14	***

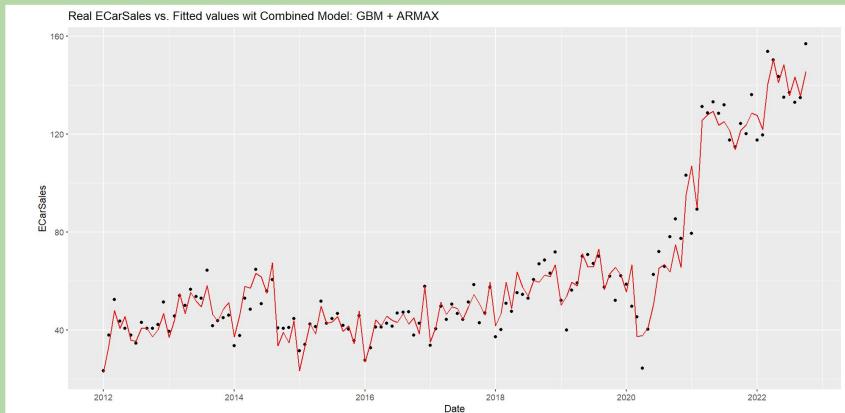
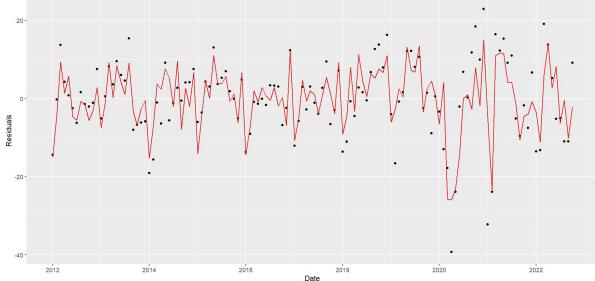
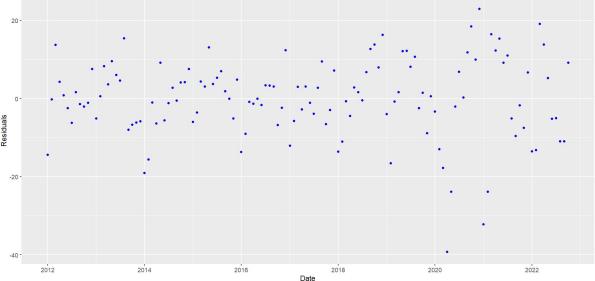
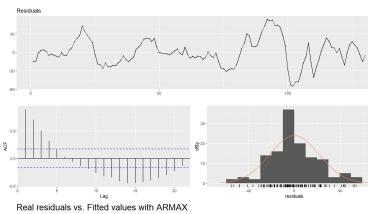
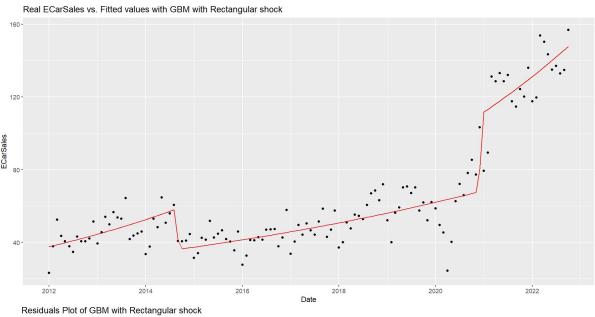
Residual standard error: 8.874 on 124 degrees of freedom
Multiple R-squared: 0.9285, Adjusted R-squared: 0.9256
F-statistic: 322.1 on 5 and 124 DF, p-value: < 2.2e-16

Only significant explanatory variables are used to fit this model that describe electric car sales.



ELECTRIC CARS

GBM with rectangular shock + ARMAX model



Thus, residuals are modelled with ARMAX process



ELECTRIC CARS

GBM with exponential shock + ARMAX model

(same as before)



Comparison
between GBM
with rectangular
and exponential
shock + ARMAX
models

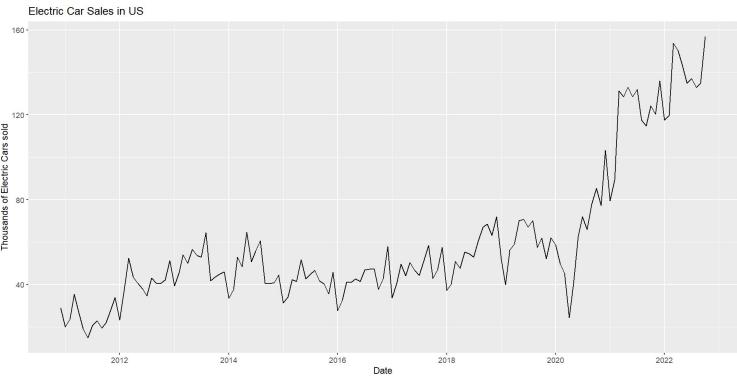
CBM rectangular
shock + ARMAX

CBM exponential
shock + ARMAX



Final model

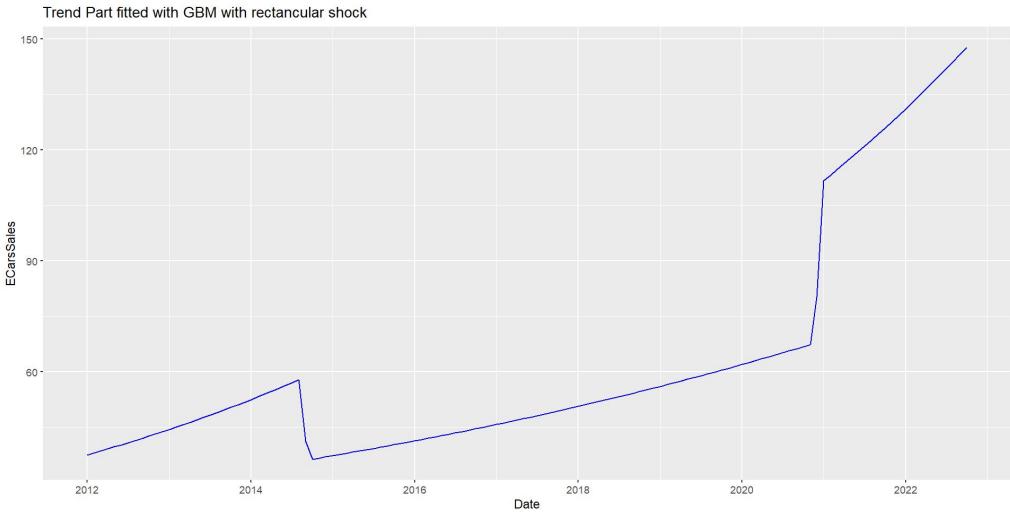
Electric car sales time
serie is decomposed in:





Final model

Electric car sales time serie is decomposed in:



Trend

+ Seasonality

+ External factors influence

Modelled with GBM with rectangular shook

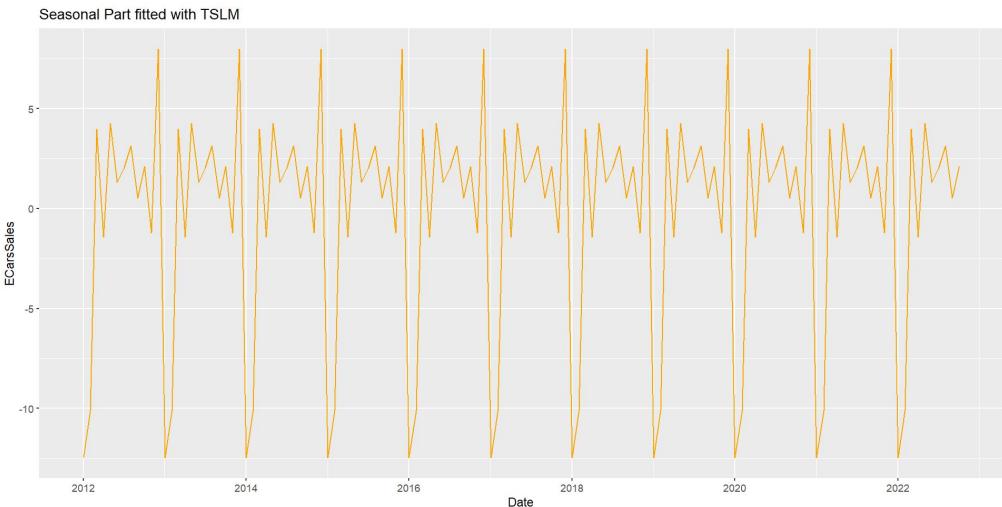
	Estimate	Std.Error	Lower	Upper	p-value
m	3.045165e+05	1.045342e+06	-1.744316e+06	2.353349e+06	7.71e-01
p	1.226309e-04	4.172196e-04	-6.951044e-04	9.493662e-04	7.69e-01
q	1.409579e-02	1.607812e-03	1.094453e-02	1.724704e-02	1.18e-14 ***
a1	3.223029e+01	4.598527e-01	3.132899e+01	3.313158e+01	1.32e-101 ***
b1	1.077116e+02	2.518951e-01	1.072179e+02	1.082053e+02	2.39e-198 ***
c1	-3.836285e-01	1.148881e-02	-4.061461e-01	-3.611108e-01	7.92e-64 ***

Residual standard error 21.42967 on 124 degrees of freedom
 Multiple R-squared: 0.9998999 Residual squared sum: 59241.08



Final model

Electric car sales time serie is decomposed in:



Modelled with **TSLM**

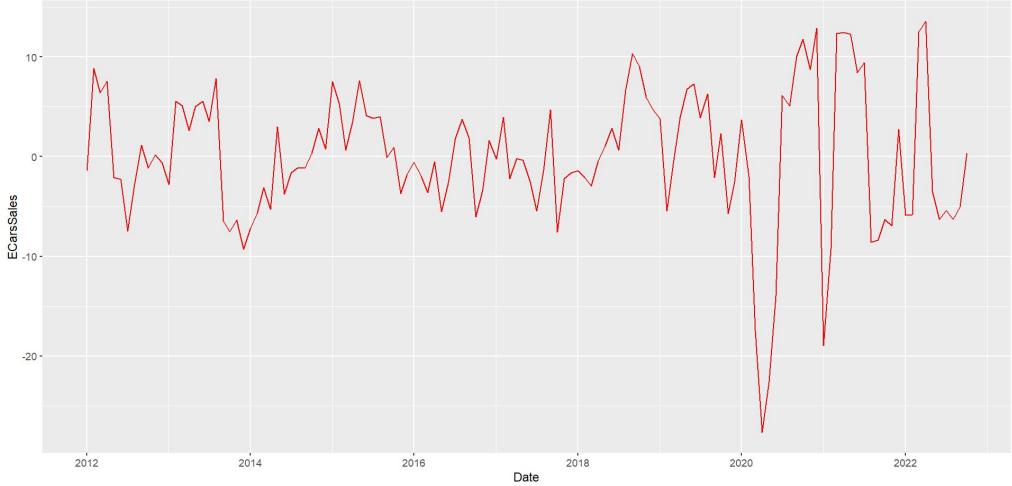
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-12.464	2.628	-4.743	5.95e-06 ***
season2	2.331	3.717	0.627	0.531666
season3	16.431	3.717	4.421	2.20e-05 ***
season4	11.042	3.717	2.971	0.003596 **
season5	16.742	3.717	4.505	1.57e-05 ***
season6	13.787	3.717	3.710	0.000318 ***
season7	14.499	3.717	3.901	0.000160 ***
season8	15.586	3.717	4.194	5.33e-05 ***
season9	12.988	3.717	3.495	0.000670 ***
season10	14.558	3.717	3.917	0.000151 ***
season11	11.256	3.808	2.956	0.003768 **
season12	20.459	3.808	5.372	3.96e-07 ***

Residual standard error: 8.716 on 118 degrees of freedom
 Multiple R-squared: 0.3145, Adjusted R-squared: 0.2505
 F-statistic: 4.92 on 11 and 118 DF, p-value: 3.005e-06



Final model

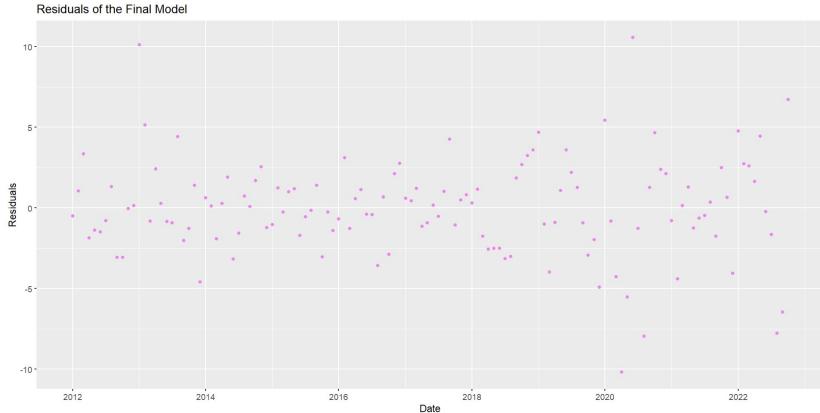
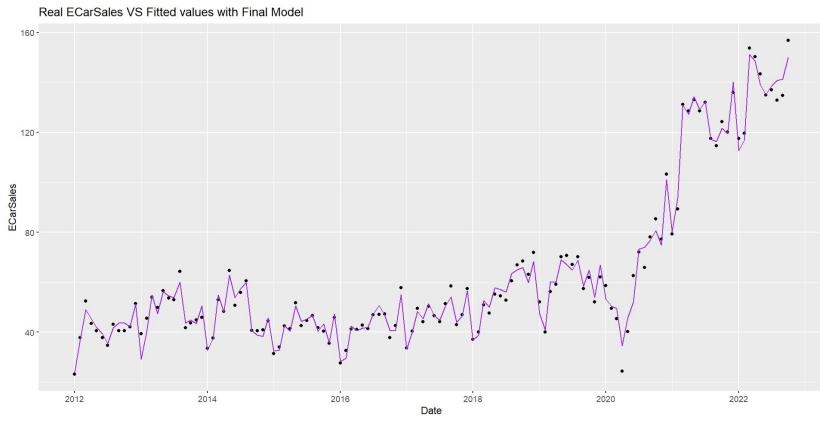
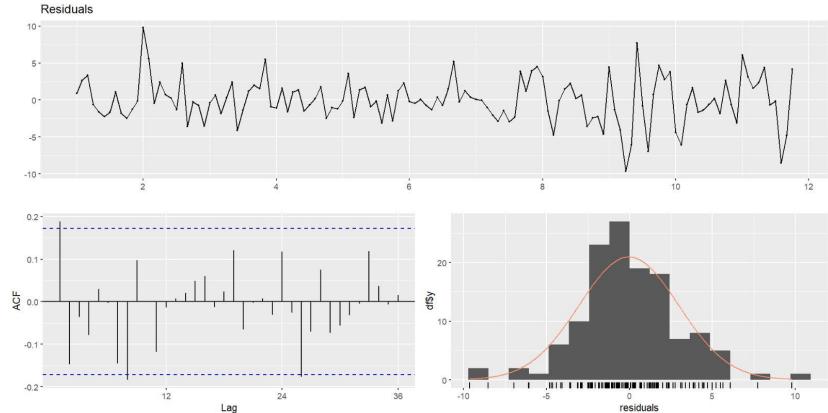
Electric car sales time serie is decomposed in:



External factors influence (TSLM residuals) is time independent, modelled with cross validated **Gradient Boosting** to avoid overfitting

Final model

The final composition of the three models seems to capture all the possible informations of the time series because **residuals are noise**.

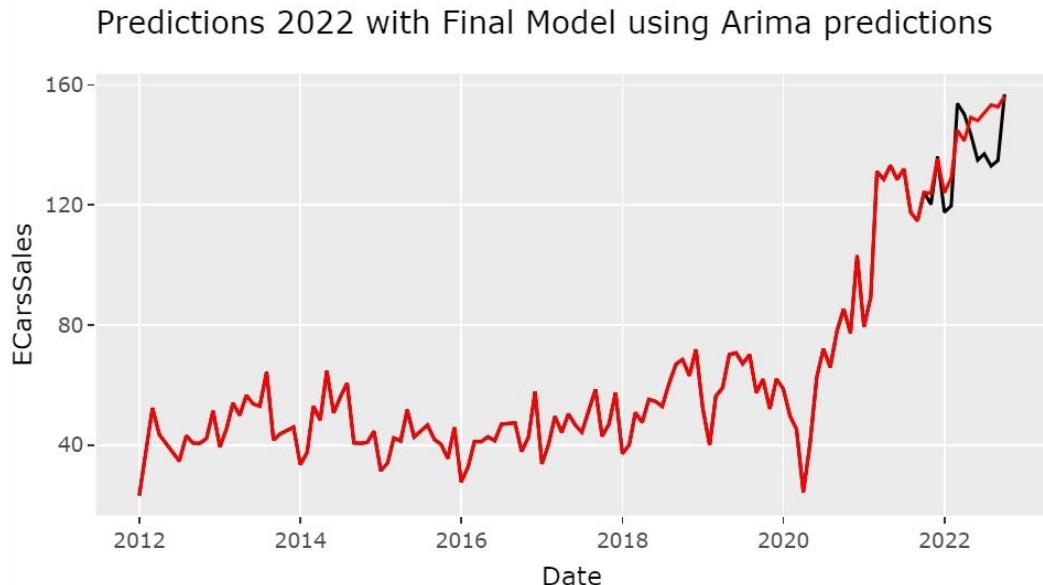




Electric car sales FORECASTING

Predict 2022 e-car sales by training models on data up to and including 2021.

ARIMA models are used to predict data of 2022.

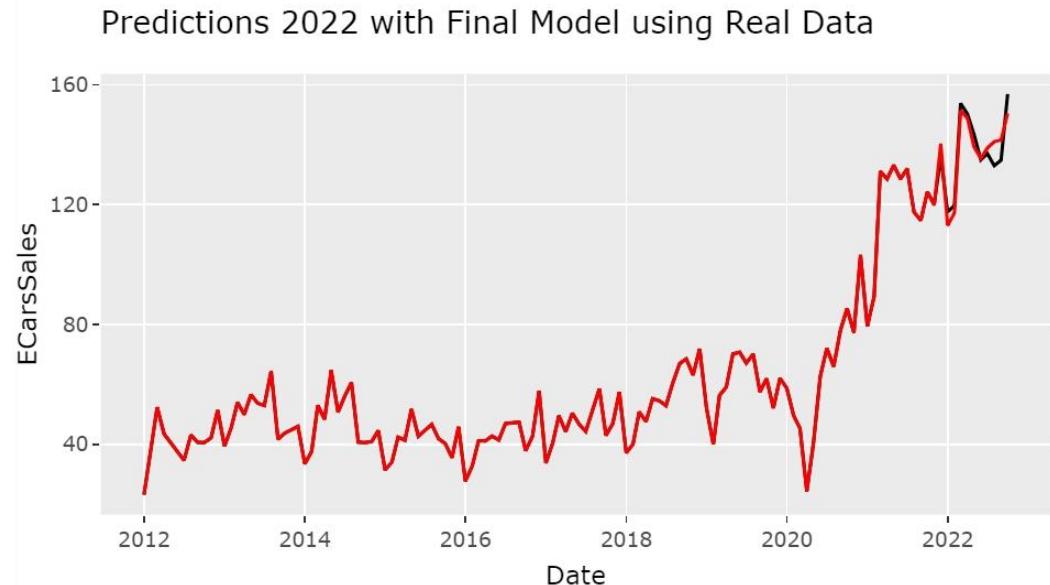




Electric car sales FORECASTING

Predict 2022 e-car sales by training models on data up to and including 2021.

Data of 2022 is used as explanatory variables.

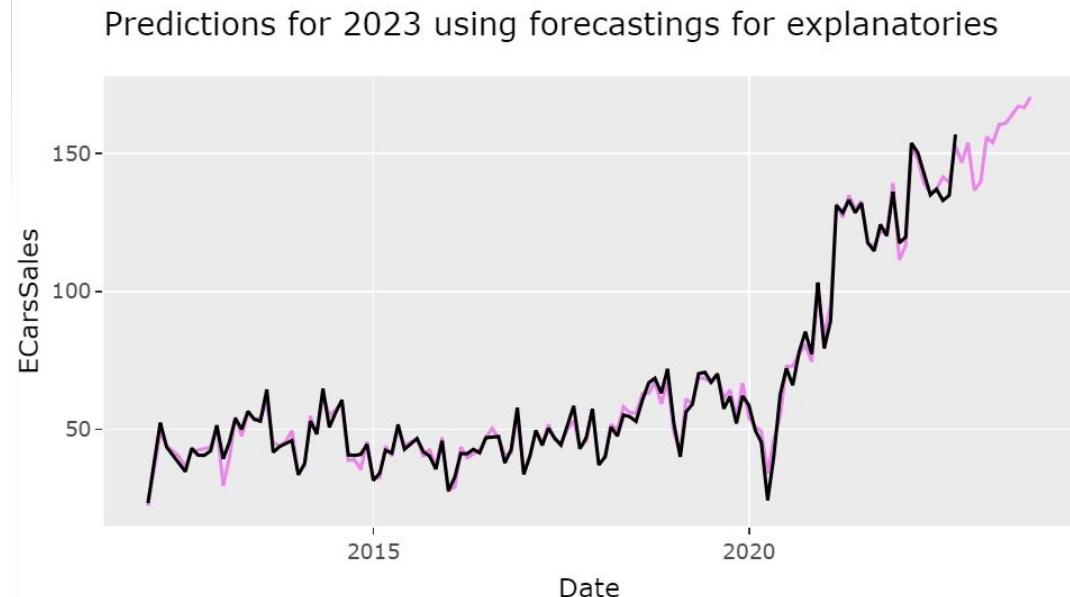




Electric car sales FORECASTING

Predict 2023 e-car sales by training models on data up to and including 2022.

Problem: ARIMA models have to be used to predict data of 2023.





Conclusions

Explanatory variables are here ordered by the influence they have in achieving good models for non-electric or electric time series.

Non-electric car sales

1. **unemployment rate**
2. **inflation rate**
3. **oil prices**
4. electricity prices
5. covid cases
6. GDP

Electric car sales

1. **oil prices**
2. **GDP**
3. **inflation rate**
4. electricity prices
5. unemployment rate
6. covid cases