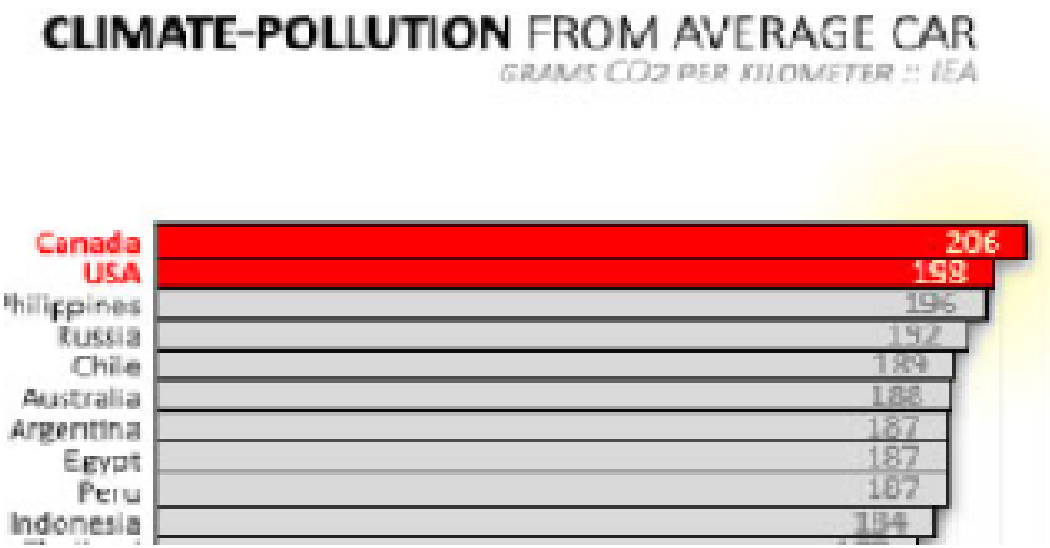
CO2 Emission emit by Vehicles ?

Thomas
Mohammadyoosef
Afsaneh

ENAC 2020

PREV Project - Defense - **April 25, 2022**

The Goal of this Project ?



As you can see, Canadians top the chart with cars that emit an average 206 grams of CO2 (gCO2) per kilometre.Our aim was to know which factors effect on this situation ?!
(Source: Canada's National Observer)

Data collection

This dataset captures the details of how CO2 emissions by a vehicle can vary with the different features. The dataset has been taken from Canada Government official open data website.
Website Kaggle

- n = 7126
- Categorical of data :
  Vehicle class = 16 types of cars
  type of fuel :
  x = Regular gasoline z = Premium gasoline d = Diesel e = Ethanol (E85) n = Natural gas
- Numerical : CO2 Emission(g/km), Engine Size(L), Cylinders, Fuel consumption in city roads (L/100 km), Fuel consumption in highways (L/100 km), The combined fuel consumption is shown in L/100 km

Clean Data

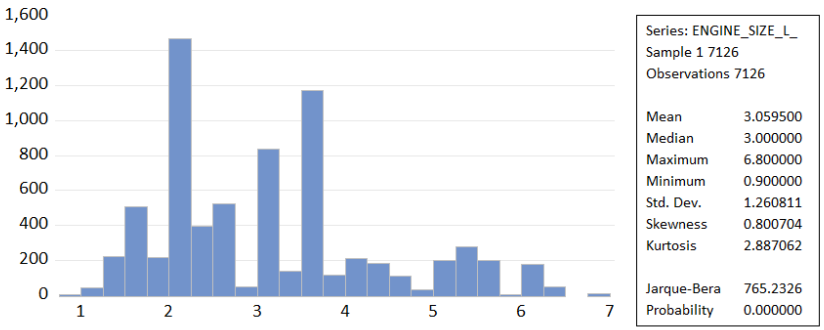
```
2 data1.describe(include='all')
```

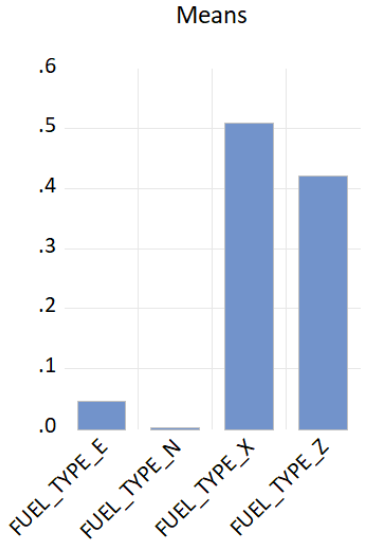
	Make	Model	Vehicle Class	Engine Size(L)	Cylinders	Transmission	Fu Ty
count	7385	7385	7385	7385.000000	7385.000000	7385	73
unique	42	2053	16	NaN	NaN	27	
top	FORD	F-150 FFV 4X4	SUV - SMALL	NaN	NaN	AS6	
freq	628	32	1217	NaN	NaN	1324	36
mean	NaN	NaN	NaN	3.160068	5.615030	NaN	Na
std	NaN	NaN	NaN	1.354170	1.828307	NaN	Na
min	NaN	NaN	NaN	0.900000	3.000000	NaN	Na
25%	NaN	NaN	NaN	2.000000	4.000000	NaN	Na
50%	NaN	NaN	NaN	3.000000	6.000000	NaN	Na
75%	NaN	NaN	NaN	3.700000	6.000000	NaN	Na
max	NaN	NaN	NaN	8.400000	16.000000	NaN	Na

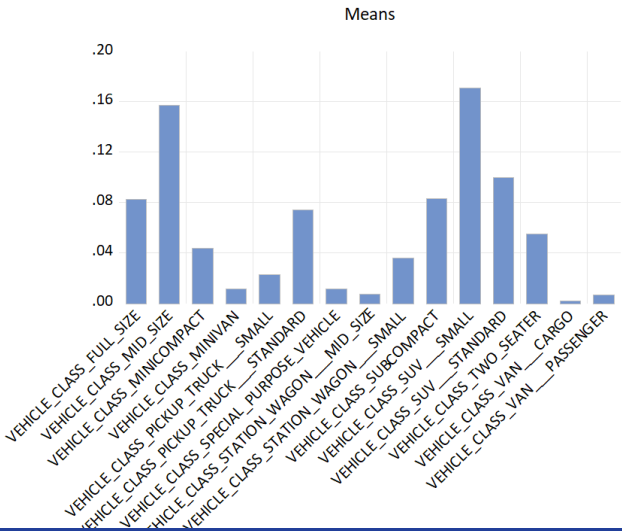
```
1 sns.kdeplot(raw_data['Fuel Consumption Comb (L/100 km)'], shade=True)
```

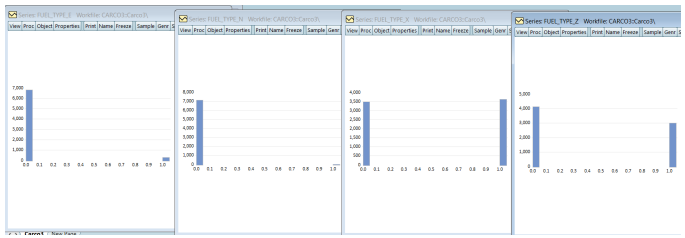
```
1 data_with_dummy = pd.get_dummies(data_V3, drop_first = True)
2
3 data_with_dummy
```

Histograms of variables used

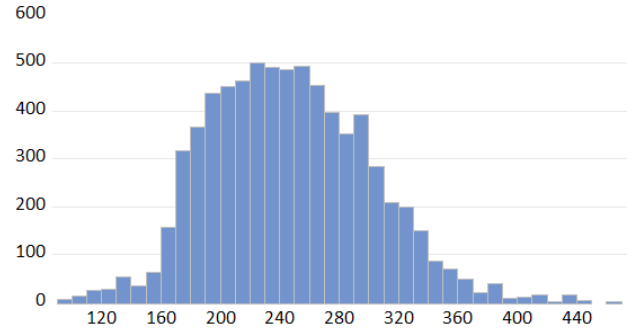




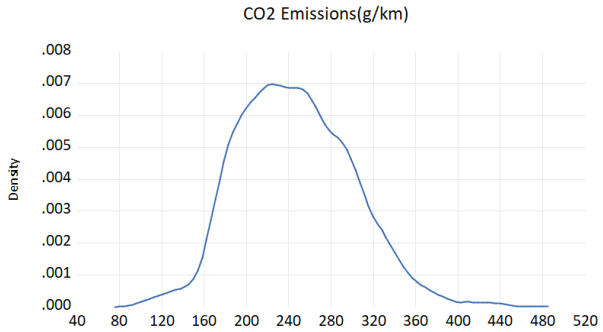


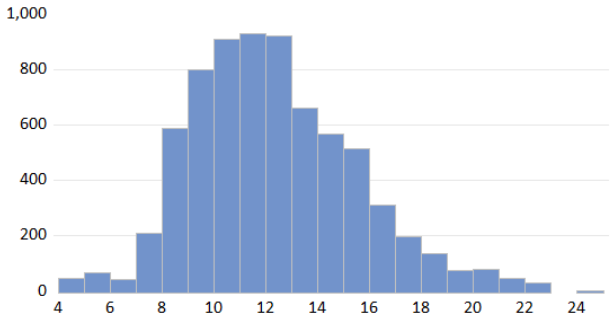


Histograms of variables used

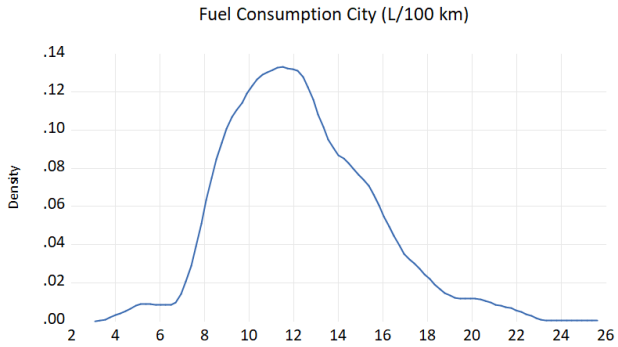


Series: CO2_EMISSIONS_G_KM_	
Sample 1 7126	
Observations 7126	
Mean	246.0841
Median	243.0000
Maximum	465.0000
Minimum	96.00000
Std. Dev.	53.91308
Skewness	0.336100
Kurtosis	3.122245
Jarque-Bera	138.5994
Probability	0.000000

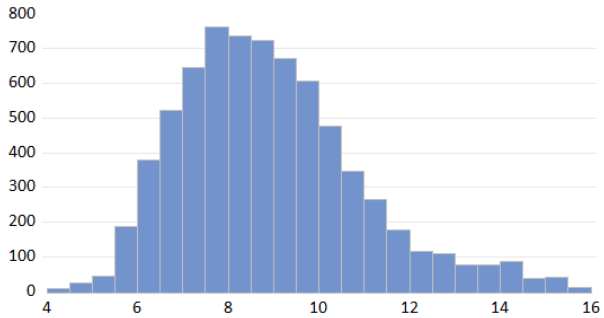




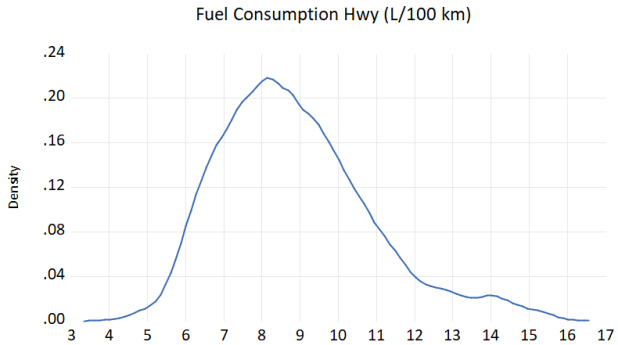
Series: FUEL_CONSUMPTION_CITY__L	
Sample 1 7126	
Observations 7126	
Mean	12.26618
Median	11.90000
Maximum	24.50000
Minimum	4.200000
Std. Dev.	3.150713
Skewness	0.515003
Kurtosis	3.297689
Jarque-Bera	341.3148
Probability	0.000000

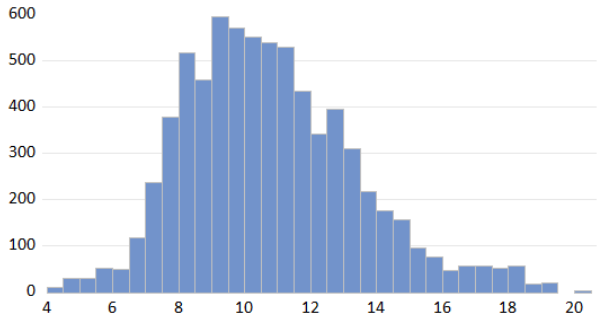


Histograms of variables used

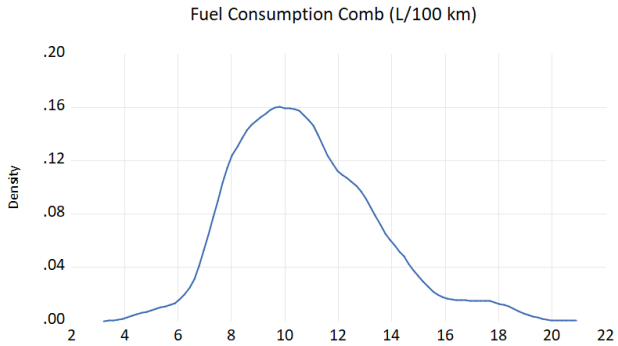


Series: FUEL_CONSUMPTION_HWY__L_	
Sample 1 7126	
Observations 7126	
Mean	8.872074
Median	8.600000
Maximum	15.90000
Minimum	4.000000
Std. Dev.	2.007780
Skewness	0.749113
Kurtosis	3.542619
Jarque-Bera	753.9053
Probability	0.000000





Series: FUEL_CONSUMPTION_COMB__L	
Sample 1 7126	
Observations 7126	
Mean	10.73908
Median	10.40000
Maximum	20.00000
Minimum	4.100000
Std. Dev.	2.603955
Skewness	0.594992
Kurtosis	3.347642
Jarque-Bera	456.3362
Probability	0.000000



First Model: Highlighted are Insignificant Variables

Dependent Variable: CO2 EMISSIONS G KM
Method: Least Squares
Date: 04/24/22 Time: 20:23
Sample: 1 7126
Included observations: 7126

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CYLINDERS	0.133607	0.098551	1.355714	0.1752
ENGINE SIZE L	-0.012011	0.121181	-0.099113	0.9211
FUEL_CONSUMPTION_CITY__L_100_...	6.255880	0.619621	10.09630	0.0000
FUEL_CONSUMPTION_COMB__L_10...	11.62904	1.121061	10.37324	0.0000
FUEL_CONSUMPTION_HWY__L_100...	5.037777	0.514008	9.800964	0.0000
FUEL_TYPE_E	-140.7370	0.471796	-298.3007	0.0000
FUEL_TYPE_N	-112.1939	4.068563	-27.57580	0.0000
FUEL TYPE X	-30.98981	0.329351	-94.09367	0.0000
FUEL TYPE Z	-30.71639	0.349815	-87.80754	0.0000
VEHICLE CLASS FULL SIZE	1.329591	0.219525	6.056677	0.0000
VEHICLE_CLASS_MID_SIZE	0.239364	0.178784	1.338844	0.1807
VEHICLE_CLASS_MINICOMPACT	-0.212919	0.267374	-0.796333	0.4259
VEHICLE_CLASS_MINIVAN	0.359333	0.483989	0.742442	0.4578
VEHICLE_CLASS_PICKUP_TRUCK__...	2.401161	0.394915	6.080198	0.0000
VEHICLE_CLASS_PICKUP_TRUCK__...	0.878794	0.291492	3.014814	0.0026
VEHICLE CLASS SPECIAL PURPOS...	3.203657	0.496087	6.457856	0.0000
VEHICLE CLASS STATION WAGON ...	0.564182	0.589000	0.957864	0.3382
VEHICLE CLASS STATION WAGON ...	-0.174020	0.289340	-0.601438	0.5476
VEHICLE_CLASS_SUBCOMPACT	0.311846	0.213566	1.460187	0.1443
VEHICLE_CLASS_SUV___SMALL	0.811403	0.194002	4.182455	0.0000
VEHICLE_CLASS_SUV___STANDARD	0.455899	0.248837	1.832116	0.0670
VEHICLE_CLASS_TWO_SEATER	0.179986	0.249485	0.721431	0.4707
VEHICLE_CLASS_VAN___PASSENGER	0.406408	0.700938	0.579805	0.5621
VEHICLE CLASS VAN CARGO	-2.093510	1.218441	-1.718187	0.0858
C	33.55902	0.442883	75.77408	0.0000

R-squared	0.994376	Mean dependent var	246.0841
Adjusted R-squared	0.994357	S.D. dependent var	53.91308
S.E. of regression	4.049999	Akaike info criterion	5.638813
Sum squared resid	116474.1	Schwarz criterion	5.662920
Log likelihood	-20066.09	Hannan-Quinn criter.	5.647112
F-statistic	52312.15	Durbin-Watson stat	1.447598
Prob(F-statistic)	0.000000		

Second Model: All significant Variables

Dependent Variable: CO2 EMISSIONS G KM
Method: Least Squares
Date: 04/24/22 Time: 21:24
Sample: 1 7126
Included observations: 7126

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FUEL_TYPE_X	41.82072	0.809271	51.67700	0.0000
FUEL_TYPE_Z	48.44913	0.821957	58.94363	0.0000
VEHICLE_CLASS_FULL_SIZE	1.926563	0.718862	2.680017	0.0074
VEHICLE_CLASS_MINIVAN	6.571629	1.752588	3.749672	0.0002
VEHICLE_CLASS_PICKUP_TRUCK__...	32.94472	1.303723	25.26973	0.0000
VEHICLE_CLASS_PICKUP_TRUCK_...	12.99163	0.878826	14.78295	0.0000
VEHICLE_CLASS_SPECIAL_PURPOS...	10.94073	1.781919	6.139855	0.0000
VEHICLE_CLASS_STATION_WAGON_...	3.074716	1.009141	3.046864	0.0023
VEHICLE_CLASS_SUBCOMPACT	2.565145	0.701008	3.659223	0.0003
VEHICLE_CLASS_SUV_SMALL	9.926496	0.573273	17.31549	0.0000
VEHICLE_CLASS_TWO_SEATER	3.053325	0.845262	3.612285	0.0003
VEHICLE_CLASS_SUV__STANDARD	13.38495	0.749623	17.85558	0.0000
VEHICLE_CLASS_VAN_CARGO	31.76011	4.431397	7.167064	0.0000
VEHICLE_CLASS_VAN__PASSENGER	44.32889	2.468038	17.96119	0.0000
CYLINDERS	5.586153	0.360149	15.51065	0.0000
ENGINE_SIZE_L	2.268508	0.444317	5.105613	0.0000
FUEL_CONSUMPTION_COMB_L_10...	15.34128	0.144055	106.4962	0.0000
C	-3.551581	1.491451	-2.381292	0.0173

R-squared	0.921986	Mean dependent var	246.0841
Adjusted R-squared	0.921800	S.D. dependent var	53.91308
S.E. of regression	15.07643	Akaike info criterion	8.266665
Sum squared resid	1615639.	Schwarz criterion	8.284022
Log likelihood	-29436.13	Hannan-Quinn criter.	8.272640
F-statistic	4941.419	Durbin-Watson stat	1.666257
Prob(F-statistic)	0.000000		

Third Model

Dependent Variable: LOG(CO2)
Method: Least Squares
Date: 04/25/22 Time: 09:48
Sample: 1 7126
Included observations: 7126

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FUEL TYPE X	0.140828	0.002934	47.99054	0.0000
FUEL TYPE Z	0.164890	0.003012	54.74987	0.0000
VEHICLE CLASS FULL SIZE	0.007105	0.002703	2.628696	0.0086
VEHICLE CLASS MINIVAN	0.022737	0.006624	3.432341	0.0006
VEHICLE CLASS PICKUP TRUCK ...	0.101424	0.004909	20.65957	0.0000
VEHICLE CLASS PICKUP TRUCK ...	0.039517	0.003308	11.94534	0.0000
VEHICLE CLASS SPECIAL PURPOS...	0.018145	0.006710	2.704203	0.0069
VEHICLE CLASS STATION WAGON ...	0.009175	0.003793	2.419230	0.0156
VEHICLE CLASS SUBCOMPACT	0.008996	0.002641	3.406403	0.0007
VEHICLE CLASS SUV SMALL	0.028837	0.002189	13.17146	0.0000
VEHICLE CLASS TWO SEATER	0.007496	0.003184	2.354493	0.0186
VEHICLE CLASS SUV STANDARD	0.038808	0.002807	13.82733	0.0000
VEHICLE CLASS VAN CARGO	0.070506	0.016647	4.235292	0.0000
VEHICLE CLASS VAN PASSENGER	0.091958	0.009101	10.10362	0.0000
CYLINDERS*ENGINE SIZE L	0.001774	9.23E-05	19.22146	0.0000
LOG(FUEL CONSUMPTION COMB ...	0.795607	0.005211	152.6900	0.0000
C	3.424501	0.011620	294.7027	0.0000

R-squared	0.936145	Mean dependent var	5.481097
Adjusted R-squared	0.936001	S.D. dependent var	0.224529
S.E. of regression	0.056801	Akaike info criterion	-2.896131
Sum squared resid	22.93641	Schwarz criterion	-2.879738
Log likelihood	10335.92	Hannan-Quinn criter.	-2.890488
F-statistic	6513.826	Durbin-Watson stat	1.721528
Prob(F-statistic)	0.000000		

Third Model

LOG(CO2) = 0.140827901069 * FUEL TYPE X + 0.164890142074 * FUEL TYPE Z + 0.0071048001467 * VEHICLE CLASS FULL SIZE + 0.0227367574717 * MINIVAN + 0.101424052436 * PICKUP TRUCK SMALL + 0.0395169155485 * PICKUP TRUCK STANDARD + 0.0181454457459 * SPECIAL PURPOSE VEHICLE + 0.00917503090167 * STATION WAGON SMALL + 0.00899598318004 * SUBCOMPACT + 0.0288373328436 * SUV SMALL + 0.00749554160184 * TWO SEATER + 0.0388076703373 * SUV STANDARD + 0.0705064281763 * VAN CARGO + 0.0919577333909 * VAN PASSENGER + 0.00177359990506 * CYLINDERS * ENGINE SIZE L + 0.795606884999 * LOG(FUEL CONSUMPTION COMB L/100KM) + 3.42450069483

Breusch-Pagan-Godfrey Test

Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic	858.3398	Prob. F(17,7108)	0.0000
Obs*R-squared	4791.800	Prob. Chi-Square(17)	0.0000
Scaled explained SS	21540.92	Prob. Chi-Square(17)	0.0000

The Prob.Chi-Square is less than 0.05, so our dataset has Heteroscedasticity

VIF (Variance Inflation Factor)

Variance Inflation Factors
Date: 04/25/22 Time: 09:51
Sample: 1 7126
Included observations: 7126

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
FUEL TYPE X	8.61E-06	9.699157	4.752941
FUEL TYPE Z	9.07E-06	8.433880	4.883271
VEHICLE CLASS ...	7.31E-06	1.331323	1.221469
VEHICLE CLASS M...	4.39E-05	1.088051	1.075836
VEHICLE CLASS P...	2.41E-05	1.187735	1.161234
VEHICLE CLASS P...	1.09E-05	1.780784	1.649587
VEHICLE CLASS ...	4.50E-05	1.074559	1.062948
VEHICLE CLASS ...	1.44E-05	1.123426	1.083698
VEHICLE CLASS ...	6.97E-06	1.279705	1.173392
VEHICLE CLASS ...	4.79E-06	1.808072	1.499284
VEHICLE CLASS ...	1.01E-05	1.225064	1.158017
VEHICLE CLASS ...	7.88E-06	1.733404	1.560696
VEHICLE CLASS ...	0.000277	1.030754	1.029018
VEHICLE CLASS ...	8.28E-05	1.129687	1.122711
CYLINDERS*ENGI...	8.51E-09	9.380809	2.995960
LOG(FUEL CONS...	2.72E-05	333.2179	3.530600
C	0.000135	298.2326	NA

$1/(1\text{-}R2\ ) = 12.81$

Wald Test

Wald Test:
Equation: Untitled

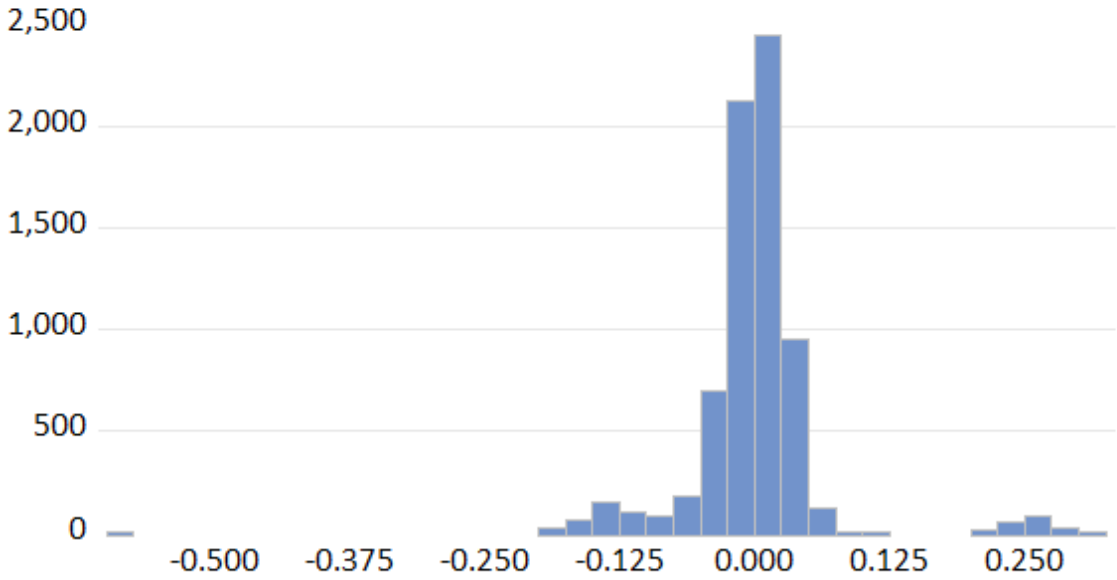
Test Statistic	Value	df	Probability
t-statistic	19.22146	7109	0.0000
F-statistic	369.4647	(1, 7109)	0.0000
Chi-square	369.4647	1	0.0000

Null Hypothesis: C(15)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(15)	0.001774	9.23E-05

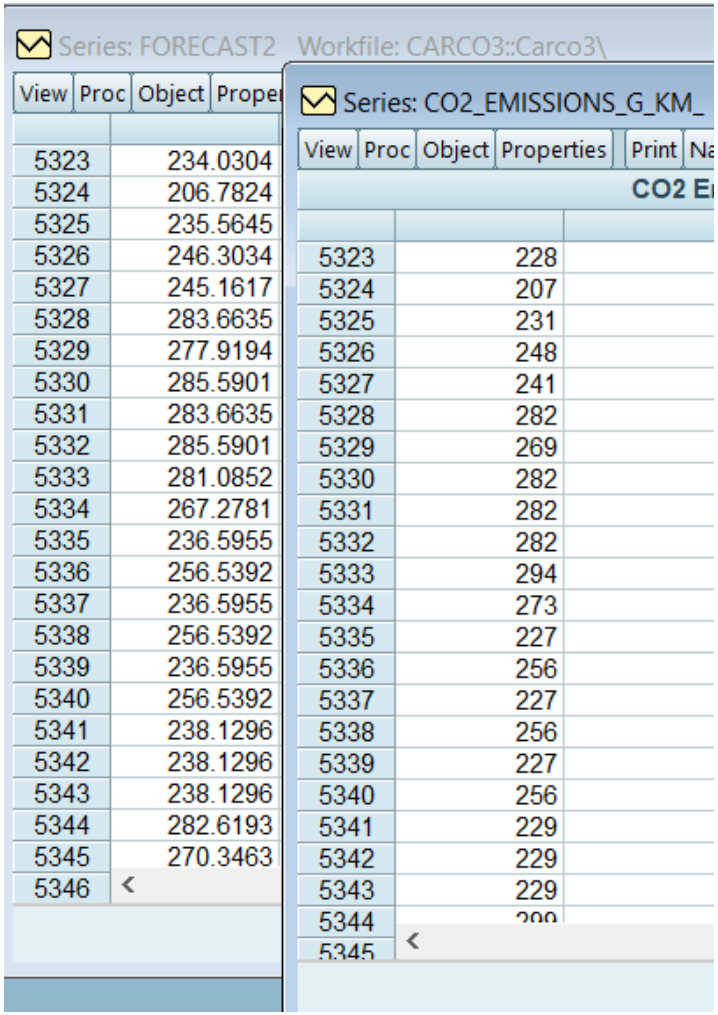
Restrictions are linear in coefficients.

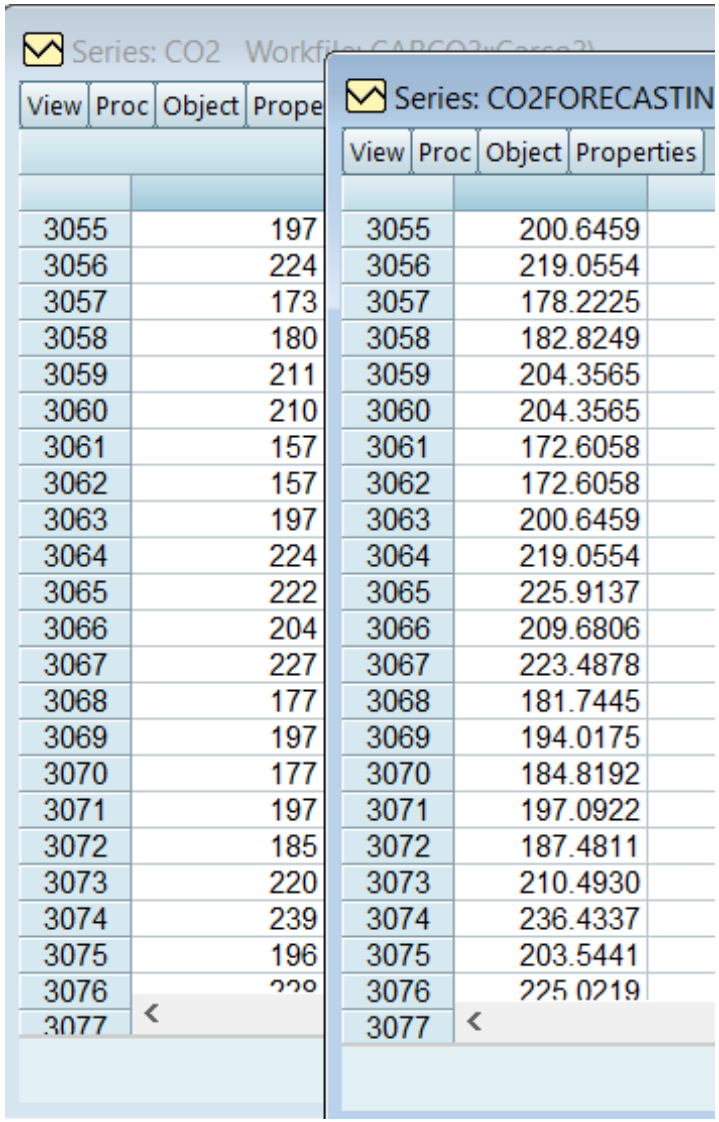
Residual



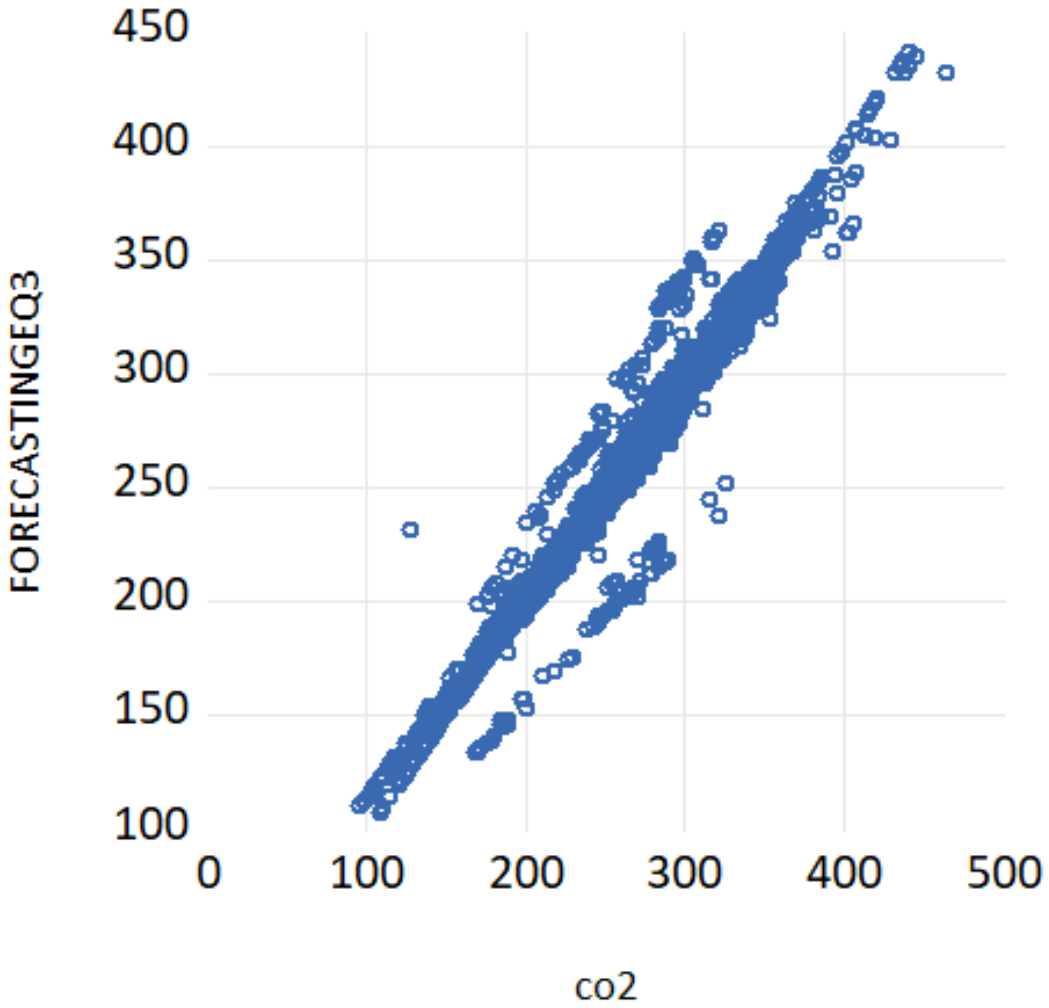
Series: Residuals	
Sample 1 7126	
Observations 7126	
Mean	-1.24e-15
Median	0.001013
Maximum	0.308502
Minimum	-0.592505
Std. Dev.	0.056738
Skewness	1.271283
Kurtosis	13.81512
Jarque-Bera	36648.84
Probability	0.000000

Forecasting





Forecasting



Conclusion

Regarding the coefficient of EQ3, it is obvious that the type of fuel and fuel consumption could be important for this problem.

End

Thank you for your attention