ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Bachelor Thesis Economics

The influence of a country's car fleet on its gasoline taxes

Name student: Pim Swank
Student ID number: 572818
Supervisor: Otto Swank
Second Assessor: Josse Delfgaauw
Date final version: 02-07-2023

Abstract

This thesis uses three different regression methods to examine the relationship between a country's car fleet and fuel taxes. The goal is to find a possible explanation for the different gasoline taxes within the European Union. Findings indicate a relationship between the car fleet composition and fuel taxation, as evidenced by the significant coefficients in the linear regression and two-stage least square instrumental variable regression. However, due to limitations in the data, there cannot be spoken about causal effects.

Keywords: Country's car fleet, fuel taxes, regression methods

JEL Codes: C21, C23, C26, D12, H21

Table of Contents

I. Introduction ... 3
II. Theoretical model ... 5
III. Dataset ... 6
Sample ... 6
Data collection ... 6
Car fleet ... 7
Composition of the car fleet ... 10
Tax rates ... 12
Density ... 13
Scatterplot and data overview ... 13
Limitations in the data ... 16
Methodology ... 17
Linear regression ... 17
Country fixed effect regression ... 17
2sls IV regression ... 18
Assumptions and validity of the methodology ... 18
Linear regression ... 18
Country fixed effect regression ... 19
2sls IV regression ... 19
IV. Results ... 22
Linear regression ... 22
Country fixed effects regression ... 22
2sls IV regression ... 23
Robustness checks ... 24
V. Conclusion ... 26
References ... 28
Used websites ... 28
Appendix ... 29

I. Introduction

The taxation of gasoline plays a significant role in shaping government revenues and consumer behaviour within the transportation sector. Different countries employ varying levels of gasoline taxes, see Table 1. Parry and Small (2005) compared gasoline tax levels between Britain and the United States. They were able to explain thirteen per cent of the differences in gasoline taxes between these two countries based on efficiency grounds. This thesis aims to investigate another possible explanation for the remaining variations in tax policies between countries, the influence of the fuel efficiency of a country's car fleet.

Table 1. Gasoline prices, taxes and the percentages of the prices

Country	Gasoline price Euro per litre of gasoline	Gasoline tax Euro per litre of gasoline	Tax % of the price
Germany	1.76	0.92	52.27
Spain	1.52	0.72	47.37
Bulgaria	1.24	0.56	45.16
Czechia	1.48	0.77	62.10
Netherlands	1.99	1.15	57.79
Denmark	1.84	0.98	53.26
Slovenia	1.35	0.66	48.89

https://www.statista.com/statistics/937796/pump-price-and-tax-of-petroleum-per-liter-by-country-eu/

Note: the gasoline prices and taxes as of February 14, 2022.

Table 1 clearly shows the variations in gasoline taxes among European Union countries. These countries were selected due to their diverse range of tax policies and other specific characteristics, like Eastern and Western countries and small and big countries. This thesis primarily focuses on analysing the taxes rather than the prices of gasoline. From the information presented in column 3 of table 1, it is evident that the taxes imposed on gasoline make up a significant proportion of the total gasoline price. When deducting the tax from the total price, it is clear that there is a small variation in the pre-tax gasoline prices. This small variation in pre-tax gasoline prices across the European Union can partly be attributed to the fact that oil prices are determined by the global market.

Several studies have explored the economic and distributional impacts of gasoline taxes. Bento et al. (2009) and Li et al. (2014) have researched the elasticity of the gasoline market. Both studies have found that as gasoline taxes increase, consumers tend to buy more fuel-efficient vehicles. Li et al. additionally found that this relationship becomes stronger in the long run, suggesting that individuals may be more responsive to tax changes than to price fluctuations. This could be attributed to the perception that tax adjustments signify longer-term shifts, which provide stronger reactions from consumers.

However, there is still a need to fully understand the influence of the fuel efficiency of a country's car fleet on its gasoline tax policies, particularly in the context of the variations in tax levels between countries. By investigating this relationship, this thesis aims to provide insights into the factors driving differences in gasoline tax policies and their potential implications. Understanding the variations in gasoline tax policies between countries is crucial for policymakers and stakeholders.

Allcott and Wozny (2014) found that people in the United States take into account the future gasoline taxes when buying a new car. So again the tax is of importance within the gasoline price. So there is already done some research to the relationship between gasoline taxes and a country's car fleet. It is therefore important to keep in my mind that reverse causality could be a problem.

By analysing an existing model and conducting empirical analyses, this study seeks to contribute to the ongoing discourse surrounding gasoline taxation, providing policymakers with valuable insights into the potential implications of varying gasoline tax policies. The findings from this research can inform discussions around tax reform, economic efficiency, and revenue generation within the transportation sector.

II. Theoretical model

Delfgaauw and Swank (2023) have developed a model of fuel taxes that tries to contribute to the understanding of the relationship between tax policy, vehicle choice, and environmental outcomes. They provide a detailed analysis of the " Gasoline Climate Trap" - a phenomenon where high levels of gasoline consumption hinder the adoption of fuel-efficient vehicles. The authors introduce two main equilibria to explain this trap: the low tax equilibrium and the high tax equilibrium.

The low tax equilibrium proposes that lower taxes on fuel do not encourage consumers to buy larger, more fuel-efficient cars. This leads to more cars with high fuel consumption and subsequently higher carbon emissions. The model highlights the role of tax policies in consumer behaviour and its impact on the environment. Delfgaauw and Swank (2023) suggest that governments should consider raising fuel taxes to discourage the purchase and use of inefficient vehicles.

In contrast, the high tax equilibrium model suggests that high taxes on gasoline encourage people to buy fuel-efficient vehicles, so environmentally friendly choices. According to this equilibrium, higher taxes are an economic incentive to buy and use efficient fuel vehicles. By imposing higher taxes on gasoline, governments can create economic incentives for consumers to switch to fuel-efficient cars, ultimately reducing the environmental impact of transportation in a country.

Delfgaauw and Swank's study indicates that there exist two possible equilibria in a country, the low- or high-tax equilibrium. By examining research in these two contrasting equilibria they shed light on the complex interaction between taxes and consumer behaviour, regarding what car to buy. The study does not say anything about a one-sided relationship, which also highlights the concern about reverse causality.

So, Delfgaauw and Swank's paper contributes to our understanding of the relationship between tax policy, vehicle choice, and environmental outcomes. They emphasize the dependence of private investments, such as vehicle purchases, on public policies and future policy expectations. This highlights the relevance of investigating the relationship between the fuel efficiency of a country’s car fleet and its gasoline taxes.

III. Dataset

Sample

The primary objective of this study is to investigate the association between the fuel efficiency of a country's car fleet and the tax rate imposed on gasoline within that country. This investigation encompasses all the nations that constituted the European Union during the year 2022. The investigation focuses on the years 2018-2022, encompassing the period before, during, and after the COVID-19 pandemic. It is worth noting that Great Britain is excluded from the sample, despite its previous membership in the European Union before 2021.

It is important to highlight that there is a lack of data regarding the car fleet of Malta, resulting in its exclusion from the sample[1]. As a consequence, the final sample for analysis consists of 26 countries. Another limitation of the data is that access is restricted to only five years' worth of data. This limitation arises from the time constraints of this bachelor's thesis.

Data collection

This research necessitated a substantial amount of data acquisition. Specifically, data on the car fleet composition, tax rates, and population density of the countries within the sample were important. The population density was important as it could potentially serve as an instrumental variable in a two-stage least squares instrumental variable regression (2sls IV regression).

In considering alternative instrumental variables, the availability of public transport in a country seemed a suitable instrument. Barla et al. (2004) analysed already the relationship between gas emissions and various factors, including transit accessibility. Their study concluded that the increase in transit accessibility is one of the two most efficient strategies to reduce gas emissions. Kim and Kim (2004) investigated how public transport affects automobile ownership and miles driven in the USA. They conclude in their paper that increasing access to public transport ensures lower car ownership and fewer miles driven.

[1] Malta is not such an important country in the EU that the drop of this country out of the sample will probably not drastically change the results.

These two papers show the correlation between the accessibility of public transport in a country and car ownership and driving habits in this country. It is therefore reasonable that there is also a relationship between the availability of public transport and how fuel-efficient people drive. When individuals have better access to public transport, their need for larger, less fuel-efficient cars is diminished, as they rely less on private vehicle travel. So this could also be a suitable instrument.

Car fleet

The main challenge of this research was creating a decent measurement of the car fleet of all 26 European Union countries. The first step was gathering information about the top 10 car models sold in each country. Two sources were utilized for this purpose. For France and Germany, the website Carsalesstatistics (https://www.best-selling-cars.com/) was specifically employed due to its availability of complete top 10 lists for both countries, along with corresponding sales figures. Notably, both websites used in this research largely featured the same popular car models sold in these two countries.

The primary source was Bestsellingcarsblog (https://bestsellingcarsblog.com/), which provided data on the 10 most sold car models for the remaining 24 countries for almost every year. It took some time to find out the top 10 for the remaining 24 countries on this website because this site had only the information in a text and not a complete list. This also caused that the sales figures were not available for these countries, which is important in further steps of defining the car fleet. In cases where the top 10 data was incomplete for certain years, an effort was made to include as many cars as possible[2]. It was challenging to gather all the necessary information, therefore it is important to note that there could be some errors in the dataset due to possibly misplacing some models.

The subsequent step in defining the car fleet involved acquiring information on the fuel consumption of each car model. To achieve this, the ANWB website (https://www.anwb.nl/auto/tests-en-specificaties?icp=auto_algemeen-V1B_navblok_2_2.2) was utilized. This resource enabled the identification of the fuel efficiency for each car model

[2] The minimum number of car models found is four, most of the time, there was information about 6 to 10 car models.

listed in the document. Fuel efficiency was measured by the number of litres of gasoline required by the factory specification of each car model to travel 100 kilometres.

The factory specification was chosen as a reliable measure, as individuals who select a car based on fuel consumption are likely to be informed of this value by the car salesperson[3]. For almost every car there were multiple versions of the car with most of the times a different fuel consumption. In all those cases where different versions of a car model were available, the average fuel efficiency value per model was calculated. For electric cars, the fuel efficiency value was recorded as zero, which will be explained later on.

The next step in defining the car fleet for each country involved assigning weights to every position in the top 10 list of the most sold car models. This was important due to the varying degrees of significance among the cars in the list. The weight assigned to each car is determined based on its sales performance, ensuring that the most sold car carries a greater weight compared to the car ranked 10th, as it has a higher sales volume. As previously mentioned, only for France and Germany the number of cars sold was determined for all years in the dataset. This information is used to calculate the percentage of total sales for each model. This is done for both countries.

So to ensure representative weights, for both countries separated the average percentage for each position across all years was calculated, see table 2.1. Subsequently, the average was taken for pairs of positions[4], such as the most-sold car model with the second-most-sold car model, the third and fourth-most-sold car models, and so on, see table 2.2 columns 2 and 3. Finally, the average of both countries for these outcomes was computed. As a result, every pair of cars, has its own weight, see table 2.2 column 4. The first two cars on the list have the same weight, the third and fourth most sold cars have their weight, and so forth[5].

[3] The expected fuel consumption of the car models will be higher, but this will be higher for every car, so it does not make a big difference.
[4] This is done out of convenience but the differences between the positions are small enough to not make a big difference in the final results.
[5] A robustness check (see the results section) is done to validate these weights.

Table 2.1. Percentage of the total sales per car model for France and Germany

Car model position	France	Germany
1	5.06	4.63
2	4.40	2.22
3	3.90	2.05
4	3.60	1.89
5	3.27	1.79
6	2.97	1.68
7	2.74	1.61
8	2.39	1.53
9	2.06	1.48
10	2.02	1.44

Table 2.2. Weights per pair of positions

Car model position	France	Germany	Final weights given in all countries
1-2	0.473	0.343	0.408
3-4	0.375	0.197	0.286
5-6	0.312	0.174	0.243
7-8	0.257	0.157	0.207
9-10	0.204	0.146	0.175

Note: I multiplied all the weights by 10, so that the final values for the fuel efficiency are bigger than 1, this is purely out of convenience.

In the final step to value the car fleet of each country per year, each car's fuel consumption was multiplied by its corresponding weight, and the total fuel consumption was divided by the number of cars found for that year. It is worth noting that some countries included electric cars in their top ten. As mentioned before, for these electric cars, the fuel efficiency value was recorded as zero. The choice for giving these electric cars a value of zero and not counting them towards the fuel efficiency of a country is made because of the small influence these cars have on the total car fleet of a country. The small influence is due to the relatively small share of electric cars in the dataset.

The dataset now includes information on the most sold cars for each country, year by year. So this provides information into the fuel efficiency of a part of the new cars in the countries. However, it is important to note that individuals mostly use their cars for more than one year. To account for this, a variable is created representing the average fuel efficiency over the years, reflecting the composition of the car fleet in 2022. From now on this variable is called the stock fuel efficiency of 2022, assuming that cars purchased in 2018 are still part of the car fleet in 2022. This means that individuals use their car for a period of approximately four to five years, depending on the purchase date in 2018. The same variable is created for 2021.

While this is a reasonable assumption, it is crucial to ensure the robustness and consistency of the results. To address this, a robustness check is done, the results of this check are displayed in table 4 of the results section. In this robustness check, each year has got another weight, for the reason that it is a possibility that the cars bought from 2018 to 2021 are not part of the car fleet anymore in 2022.

Composition of the car fleet

Figure 1 displays two clustered column diagrams. Figure 1.1 represents the stock fuel efficiency of 2022 and 2021, considering the assumption that individuals own their cars for multiple years. The analysis reveals that there is relatively little variation between the two years, indicating a relatively stable fuel efficiency trend. The stock fuel efficiency of 2021 will have limited relevance in the analysis due to the inclusion of more recent data from 2022. As the most recent year in the dataset, the fuel efficiency values from 2022 hold greater significance and provide a more up-to-date understanding of the current state of fuel efficiency. Additionally, it is worth noting that for the stock variable of 2021, there are in the data only three years prior to the reference year.

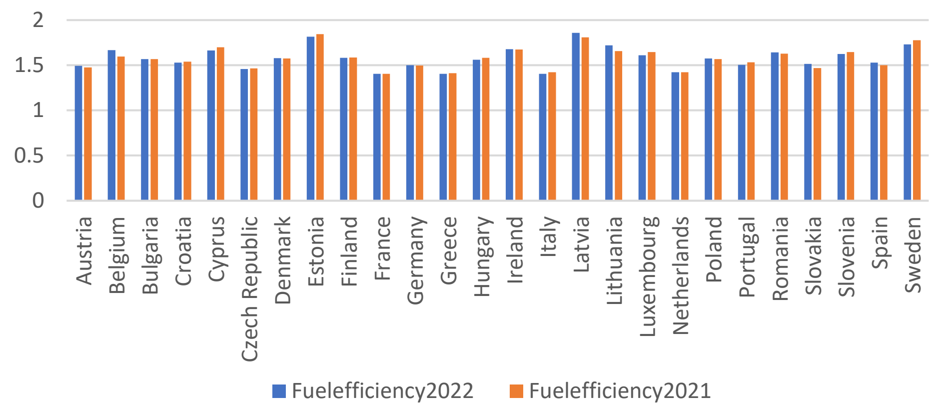

Figure 1.1. Fuel efficiency during the period 2022-2018
Stock variables

In the clustered column diagram in figure 1.2, the focus shifts to the most sold cars for each year. This representation illustrates that the fuel efficiency values show greater fluctuations when considering the specific cars that were popularly purchased each year. The contrasting patterns between the two diagrams highlight the importance of the stock variable of 2022. While the stock fuel efficiency remains relatively consistent over multiple years, the varying composition of the most sold cars leads to more pronounced fluctuations in fuel efficiency values. The stock fuel efficiency seems to be more realistic regarding the car fleet of a country, which ensures that this variable will be used the most in this research.

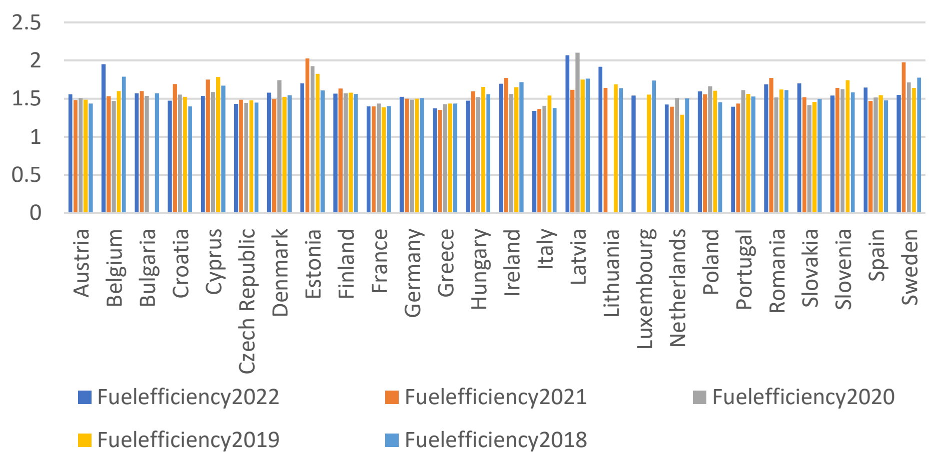

Figure 1.2. Fuel efficiency during the period 2022-2018
Flow variables

Figure 2 is presented to illustrate the variation in stock fuel efficiency in different countries, in particular, accounting for the addition of electric vehicles to provide a clear explanation as to why electric vehicles were not considered in this study. As mentioned earlier and shown in Figure 2, the impact of electric vehicles is relatively small on the overall car fleet. Only a few countries show differences in fuel efficiency when looking at electric cars. The small selection of electric car owners want probably higher tax rates as it would give them a greater advantage compared to the others and it will encourage more people to buy electric cars. However, the selection is that small, they lack the political influence to shape taxation policies, because the median voter is still the owner of a gasoline-powered car.

Following this preliminary argument, this study aims to investigate the relationship between the fuel efficiency of a country's car fleet and fuel taxes and consequently electric vehicles are not considered necessary in this particular examination of the relationship. Thus, with the research objective in mind, there are two reasons for excluding electric vehicles from this empirical study.

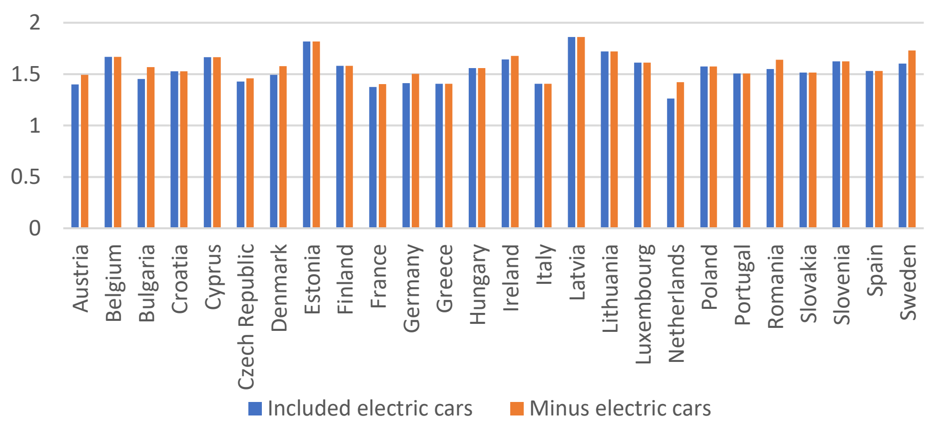

Figure 2. Stock fuel efficiency 2022 with and without electric cars

Tax rates

The second component of data acquisition involved obtaining the tax rates imposed on gasoline for each country. To accomplish this, data from ACEA (https://www.acea.auto/) was utilized. ACEA provided information on the tax rates for all the years, covering each country within the European Union. The tax rates are measured in euros per 1000 litres of fuel. As a result, the dataset contained the necessary information to generate scatterplots, enabling the investigation of potential relationships between fuel efficiency and tax rates across European Union countries.

Figure 3 presents the tax rates over the years, obtained from ACEA. The tax rates in most countries show a high degree of stability, with minimal fluctuations observed over time. The consistency in tax rates implies a relatively steady regulatory framework for taxation imposed on fuel.

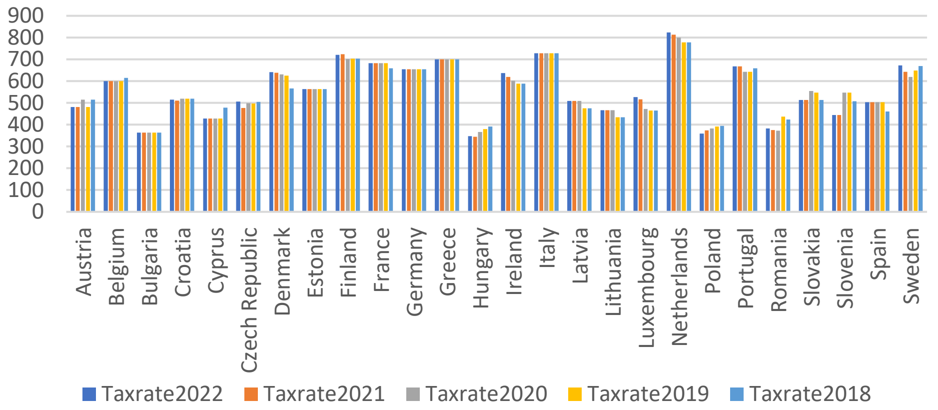

Figure 3. Tax rates during the period 2022-2018

Density

As previously mentioned, population density is a crucial variable in this research, serving as a potential instrumental variable in a 2sls IV regression. To obtain reliable data on population density, Eurostat[6] was utilized. Eurostat is a highly regarded database among researchers due to its comprehensive European information. This database provided information on population density for all five years and all 26 countries included in the research. The density is measured in the number of people per square kilometre. With the inclusion of this data, the dataset now encompasses enough information to conduct the research as effectively as possible.

Scatterplot and data overview

To examine the potential correlation between the gasoline tax rate and the fuel efficiency of a country, scatterplots were generated. Six scatterplots were created, one for each year, along with another scatterplot displaying the tax rate of 2022 and the average fuel efficiency over the years per country. In figure 4, the scatterplot shows the correlation between the stock fuel efficiency of 2022, on the x-axis, and the 2022 tax rate on the y-axis[7].

---

[6] https://ec.europa.eu/eurostat/databrowser/view/tps00003/default/table?lang=en

[7] The scatterplots per year can be found in the appendix, see figure 6.

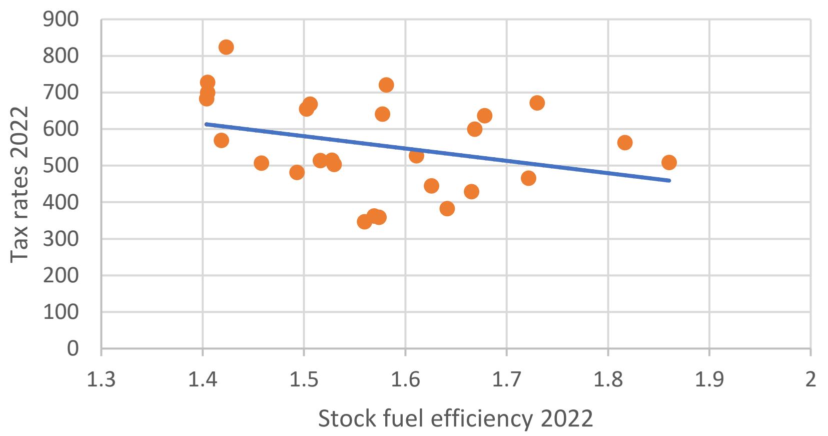

Figure 4. Correlation between the 2022 gasoline tax rate and the stock fuel efficiency of 2022

Figure 5 displays the minimum, maximum, and average values for all variables over the period 2018 to 2022, as well as for the year 2022 separated, which is the most recent and deemed the most significant within the dataset. Notably, the stock fuel efficiency is utilized as the representative variable for the fuel efficiency of 2022.

It is essential to acknowledge that the fuel efficiency values lie within the range of 1.2 and 2.2. These values are weighted to account for all the models, meaning that they do not directly imply that cars can travel 100 kilometres on 1.2 litres of gasoline. To obtain an accurate representation, these values must be multiplied by 3.477, which is derived from the product of 1.318 and 2.637. This multiplication is necessary due to the cumulative weight of the car models of 1.318. Furthermore, the top ten cars sold comprise only 26.37 per cent of the total cars sold, there has to be accounted for and therefore the value needs to be multiplied by 0.2637 as well. Due to the multiplication of 10 in the values already, this becomes 2.637. Thus, to derive an appropriate estimation for the entire car fleet, the value should be multiplied by 3.477.

To have a better understanding of how this looks like, when examining the highest and lowest stock fuel efficiency values, namely 1.860 and 1.404, obtained from Latvia and Italy correspondingly, the resultant actual fuel efficiency values are 6.467 and 4.882. Consequently, it can be deduced that in Latvia, cars achieve an average fuel consumption ratio of 1 in 15.46. In Italy, cars attain an average fuel consumption ratio of 1 in 20.48. This indicates that a distance of respectively 15.46 and 20.48 kilometres can be covered with one litre of gasoline.

Upon examination of the figure, it is evident that density and tax rates demonstrate relatively consistent values over time. The minimum, maximum, and average values observed in 2022 closely resemble those computed for the entire period spanning from 2018 to 2022. However, there exist some variations in the minimums, maximums, and averages between the stock fuel efficiency and the fuel efficiency of the period.

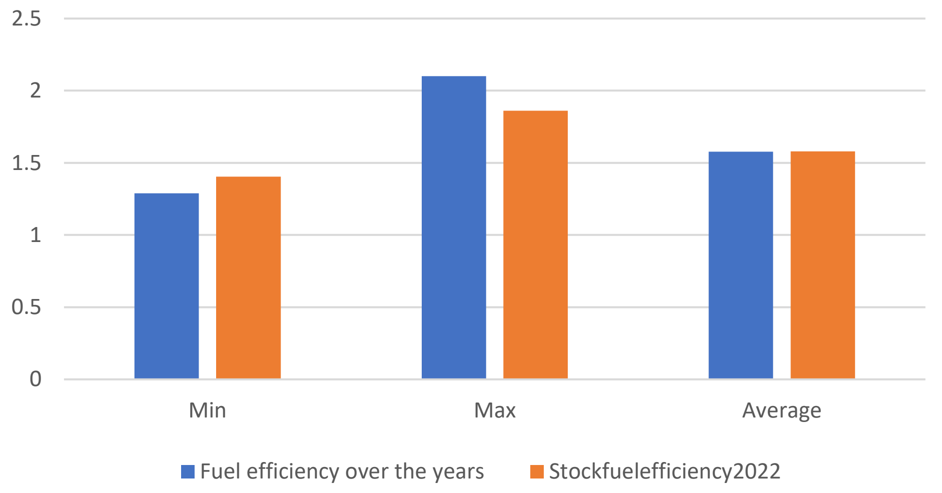

Figure 5.1. Fuel efficiency

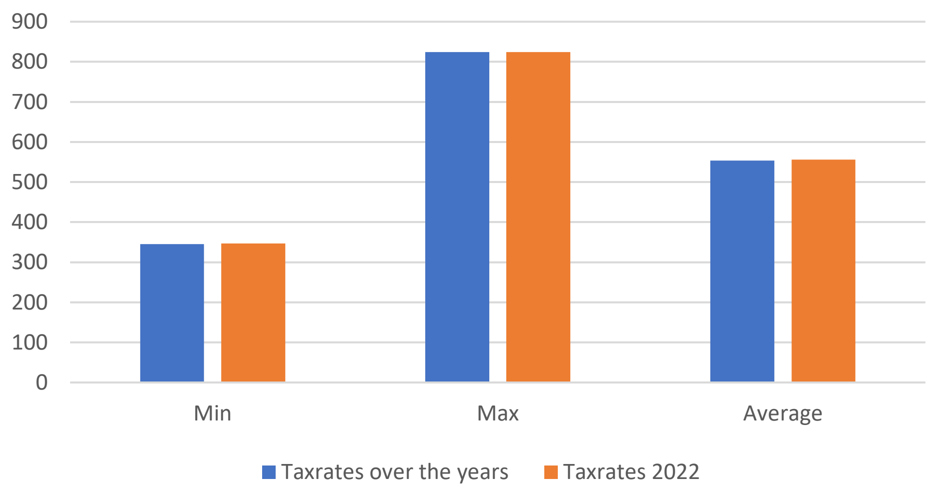

Figure 5.2. Tax rates

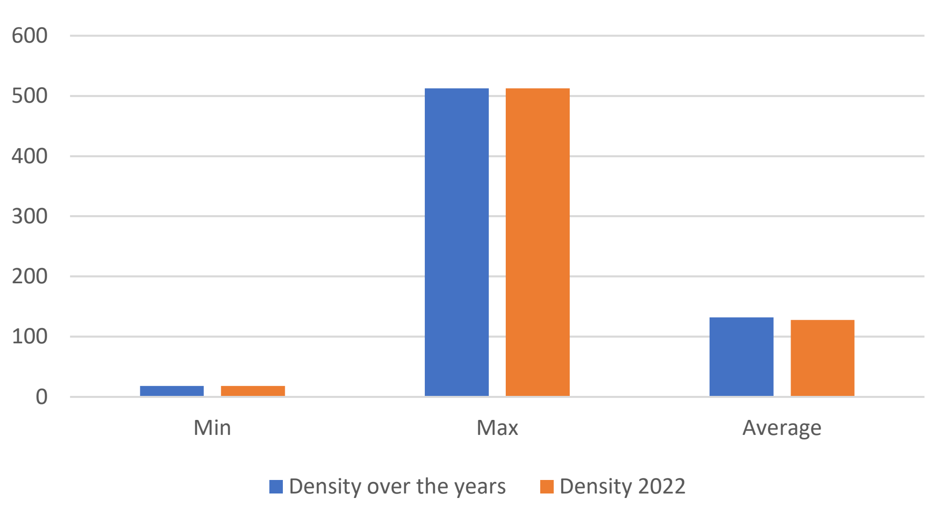

Figure 5.3. Density

Limitations in the data

The dataset has several limitations that need to be acknowledged. Firstly, as mentioned before, the dataset does not contain an equal number of most sold cars for every country. In some cases, there were even years for some countries where no cars were available in the dataset. When there are fewer than 10 cars, the total fuel efficiency is simply divided by the number of cars found, which I previously mentioned and which is not a significant issue. However, when there was for a year no list with the most sold cars in a country, the fuel efficiency for that country in that year is in the dataset recorded as zero.

Another limitation in the dataset is regarding the weights assigned to each car. As mentioned earlier, weights were assigned to pairs of cars to represent the first and second-most-sold cars, the fourth and fifth-most-sold cars, and so on. However, there is a significant difference in the representation of the car fleet between Germany and France. In Germany, the 10 most-sold cars accounted for approximately one-third of the total sales, whereas in France, this accounted for only around one-fifth. This difference of around twelve per cent indicates that the top 10 most-sold cars are for every country from different importance. Besides, the small value for the sales in France could be evidence that it may not be entirely representative of the entire country's car fleet. Despite this limitation, given the available information on the most-sold cars in European countries, this approach was the most representative in assigning weights to each car and country in the dataset[8].

[8] See the results for a robustness check for this assumption.

The third limitation in the dataset is the missing variable about the availability of public transport in the European Union countries. Due to the unavailability of data on public transport accessibility and the limited time, this variable was not included in the dataset and cannot be used as an instrumental variable. However, this provides a compelling reason for future research. If someone can find a good measurement for the availability of public transport this could possibly be a suitable instrumental variable.

Methodology

Linear regression

After the scatterplots were generated, a linear regression model was constructed to analyse the relationship between the tax rate and the fuel efficiency of the countries. The regression equation takes the form:

$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

where y_i represents the gasoline tax rate in 2022, α represents the constant term, β represents the effect of x_i (the stock fuel efficiency of 2022), and ε_i represents the error term. By fitting this regression model, we can estimate the impact of fuel efficiency on the tax rates across the countries.

Country fixed effect regression

With data collected over time, it is possible to construct another regression model that incorporates country-fixed effects. By including country-specific fixed effects, There can be accounted for time-invariant characteristics specific to each country. The regression equation for this model is as follows:

$$y_{it} = \alpha_i + \beta x_{it} + \gamma_t + \varepsilon_{it},$$

where y_{it} represents the tax rate at time t, α_i represents the country-specific fixed effect capturing all the country time-invariant characteristics per country, β represents the effect of the car fleet captured in x_{it}, γ_t represents the time effect, and ε_{it} represents the error term. By estimating this regression model, we can examine the impact of the car fleet (x_{it}) on the tax rates while accounting for country-specific fixed effects (α_i) and time effects (γ_t).

2sls IV regression

Another potential method to measure the effect of the car fleet on a country's tax rate on gasoline is a 2SLS IV regression. In this approach, the instrumental variable used is the variable density. The regression equation for this method can be expressed as follows:

$$y_i = \alpha + \beta \dot{x}_i + \varepsilon_i.$$

In this equation, α represents the constant term, ε_i represents the error term, and β represents the effect of the car fleet. However, the main difference lies in the term $\dot{x}_i$, which captures the predicted values from the regression of the car fleet on the density. This is why it is called a two-stage least squares regression, as it involves two stages: the first stage where the car fleet is regressed on the density to obtain the predicted values, and the second stage where these predicted values are used in the regression of the tax rate on the car fleet.

Assumptions and validity of the methodology

Linear regression

In the linear regression model used in this study, no control variables are included. This is due to the absence of variables in the dataset that influence both the tax rates in countries and the fuel efficiency of the car fleet. As a result, there may be omitted variables within the regression, leading to endogeneity issues. As a result of this endogeneity issue, the estimated coefficient for the effect of the car fleet on the tax rate may be biased. In other words, the coefficient obtained from this specific linear regression model may deviate from the true underlying relationship between the car fleet and the tax rate.

Furthermore, reverse causality is a concern. As mentioned in the introduction, there is research done on the effect of taxes on how fuel efficient a country's car fleet is. So while the focus of this thesis is the opposite relationship, it is plausible that gasoline taxes may also affect the fuel efficiency of a country's car fleet, which means that there is reverse causality and the fuel efficiency variable is likely endogenous. This makes speaking about a causal effect impossible.

Other assumptions of the linear regression model include the absence of autocorrelation and multicollinearity and the presence of homoscedasticity. These assumptions can be tested and there exist easy fixes. Based on the scatterplot of the residuals against the independent variable[9], there is no autocorrelation. Figure 7, shows that the residuals are lying around zero and that there is no clear pattern in the plot, which means that this assumption is met. Multicollinearity is not a concern in this model as there is only one independent variable. Lastly, a Breusch-Pagan test is conducted to assess the presence of heteroscedasticity, and the obtained p-value of 0.69 suggests that there is no heteroskedasticity, so this assumption is met as well.

Country fixed effect regression

In the country fixed effect regression, the term α_i captures all time-invariant characteristics, eliminating the need to explicitly control for these variables. This characteristic makes the estimation potentially more accurate than the linear regression analysis. However, the country-fixed effects method encounters limitations when it comes to accounting for time-varying variables or shocks that affect both the dependent and independent variable. This limitation introduces the possibility of omitted variable bias and undermines the establishment of a causal relationship.

This model is also subject to two other significant limitations. Firstly, the definition of the car fleet is limited to the top 10 new cars in each country. This narrow scope represents only a fraction of the total car fleet within a country. Secondly, to conduct a robust country fixed effects regression, a substantial amount of data is required. With only five years of data available, the sample size may not be sufficient to make conclusions in the end.

2sls IV regression

In the 2sls IV regression, a significant problem in the other regressions, namely selection bias, is addressed. This bias occurs when groups differ not only in the dependent variable but also in other unobserved factors. Furthermore, reverse causality is also addressed by the use of an instrumental variable.

When employing an instrumental variable, three crucial assumptions must be considered:

1. Strong first stage: The instrument has a clear and strong causal effect on the variable of interest.
2. Independence: The instrument is uncorrelated with the error term.

[9] See figure 7 in the appendix.

3. Exclusion restriction: The instrument does not directly affect the outcome.

To verify the first assumption, the relationship between average density and average fuel efficiency is estimated. The results are presented in table 3.

Table 3. Linear regression of the car fleet on the density

	AverageFuelefficiency
Density2022	-0.0004*
	(0.0002)
Constant	1.6300***
	(0.0371)
Number of observations	26
F-Statistic	3.77

Note: The robust standard errors are shown in parentheses. *p-value<0.1, **p-value<0.05, ***p-value<0.01.

Table 3 demonstrates a relationship between the average density and the average fuel efficiency of countries. The expectation is that higher-density areas tend to favour smaller and more fuel-efficient cars due to factors such as manoeuvrability and parking constraints. The negative coefficient for density is expected and consistent with the expectation. However, the associated f-test yields a small value. Thereby is the coefficient not as statistically significant as expected. Consequently, the explanatory power of density in accounting for fuel efficiency variance appears limited. This raises doubts about whether the first assumption is met.

The second and third assumptions cannot be directly tested. Determining the validity of the second assumption is challenging, as it is difficult to identify variables in the error term that are correlated with density. On the other hand, the third assumption may hold, as there is likely no direct relationship between density and gasoline tax rates. It is important to note that conducting a regression of average density on average tax rates would not disentangle the direct effect of density on tax rates from its indirect effect through fuel efficiency.

Since the first assumption is not convincingly met and because of the inability to test the other two assumptions, it is unlikely that density serves as a perfect instrument. As previously mentioned in the data limitations part, the availability of public transport may serve as a more suitable instrument, although no data was found for public transport access in the 26 countries of the sample.

It is important to recognize that in a two-stage least squares regression with an instrumental variable, the final estimation cannot be extrapolated to other groups. In the context of this research, this limitation is not a significant issue since the focus is on European countries. Nonetheless, it is crucial to keep in mind that the results cannot be readily applied to other populations.

IV. Results

Table 4. Results of the three regression models

	Linear regression	**Fixed effects regression**	**2sls IV regression**
	Taxrate2022	Taxrate	Taxrate2022
AverageFuelefficiency	-353.7472*	-3.788	-1119.599*
	(175.849)	(3.329)	(609.917)
Constant	1113.938***	557.180***	2323.023**
	(281.4025)	(6.713)	(948.5109)
Number of observations	26	130	26

Note: The robust standard errors are shown in parentheses. *p-value<0.1, **p-value<0.05, ***p-value<0.01.

Linear regression

The negative coefficient in the linear regression is statistically significant at a 5% significance level, indicating a negative relationship between a country's car fleet and its tax rate imposed on gasoline. It suggests that countries with more fuel-efficient car fleets on average tend to have higher tax rates on gasoline. It is important to note that this is because a more fuel-efficient car fleet, means that the value of this variable is lower. The magnitude of the increase in tax rates corresponding to an increase in fuel efficiency cannot be determined due to the bias in the coefficient. Remember, because of this biased coefficient, no causal relationship can be inferred from these results. The linear regression merely reveals a correlation between the variables.

Regarding the constant term, no meaningful interpretation can be made because it is highly unlikely for a country to have a completely fuel-efficient car fleet during the years considered in this regression. A fuel efficiency of zero would imply that all the cars in a country are electric, which may be a possibility in the future but not currently.

Country fixed effects regression

The coefficient of the stock fuel efficiency for the car fleet in the fixed effects regression is found to be statistically insignificant. Consequently, no definitive conclusions can be drawn regarding this coefficient. However, the fact that the sign of the coefficient is again negative provides further evidence of a negative relationship between fuel efficiency and gasoline tax rates.

As mentioned before, the coefficient in the fixed effects regression is expected to be less biased compared to the linear regression method since it captures the time-invariant characteristics of each country. It is a pity that the dataset does not contain control variables and just five years of data. To obtain more accurate estimates and better understand the magnitude of the effect, it would be beneficial to include additional time-varying variables that could be controlled for in the regression. It would also be very beneficial if there were more years in the dataset. This would make this regression, quite valid and this would provide more insights into the relationship between fuel efficiency and tax rates imposed on gasoline. For now, it is not possible to say something about the magnitude because of the insignificant coefficient.

2sls IV regression

The coefficient of the fuel efficiency obtained from the 2SLS IV regression analysis aligns with the previous analyses, demonstrating a negative relationship between a country's car fleet's fuel efficiency and its gasoline tax rates. This suggests that countries with more fuel-efficient car fleets tend to impose higher tax rates on gasoline. Just like in the linear regression, the interpretation of the constant should be approached with caution, as it is unrealistic to have a car fleet that is completely fuel-efficient within the years considered in this regression.

Considering the significant coefficient of the stock fuel efficiency of 2022 at a 10 per cent significance level and assuming the validity of density as an instrumental variable in our 2SLS IV regression, we can interpret the coefficient of this regression. The coefficient has a magnitude of -1119.599. So, each improvement in fuel efficiency of one litre per 100 kilometres is estimated to result in an increase of approximately 1119.599 euros per 1000 litres in the tax rate. An improvement of a country's car fleet of 0.1 litres per 100 kilometres, is associated with an estimated increase in their tax rate by approximately 11 cents per litre.

To create a better perspective, this means that when a country's car fleet decreases from 5.4 litres per 100 kilometres to 5.3 litres per 100 kilometres, their tax rate on gasoline on average increases by 11 cents for 1 litre of gasoline. When again looking at the highest and lowest value for the stock fuel efficiency in our dataset, 1.860 and 1.404, respectively from Latvia and Italy, the difference is 0.456. According to the results, the tax rate in Latvia should be 1,74 euros per litre lower than in Italy[10]. So with the calculations, it is again important to not forget to multiply the values by 3.477 because of the weighted values of the fuel efficiency.

Robustness checks

After formulating certain assumptions, it becomes essential to conduct robustness checks to assess the consistency of the results under alternative assumptions. One specific assumption pertains to assigning equal weight to each year by calculating an appropriate weight for the stock fuel efficiency in 2022. Performing robustness checks helps evaluate whether the findings remain unchanged when different assumptions are considered. Table 4 shows the results after giving different weights for every year[11]. For the country fixed effects regression, the stock fuel efficiency of 2022 is not used, so the results of this regression need no check.

Table 5. Robustness checks, same weight for every year assumption

	Linear regression	**2sls IV regression**
	Taxrate2022	Taxrate2022
AverageFuelefficiency2	-460.0224**	-1238.625*
	(213.4037)	(677.133)
Constant	1136.76***	2120.627**
	(271.7545)	(846.5797)
Number of observations	26	26

Note: The robust standard errors are shown in parentheses. *p-value<0.1, **p-value<0.05, ***p-value<0.01.

When comparing the results of the robustness checks to the original findings, it is clear that there are some differences. The coefficient of the linear regression shows slightly increased significance, and the magnitudes of the coefficients in both regressions are higher. However, the overall differences are not substantial. The constant terms remain significant as observed in the initial results, and their magnitudes are nearly unchanged as well. These findings suggest that while there are some variations in the results, the overall patterns and conclusions remain consistent.

[10] There are two possible calculations: 0.456*3.477/0.1*0.11=1.74 and (6.467-4.882)/0.1*0.11=1.74
[11] The new weights are: 2022-1; 2021-0.9; 2020-0.8; 2019-0.7; 2018-0.6

The other assumption made was the one about giving every car a representative weight. The weight that is given was the average between how much the cars were sold, compared to the total cars sold, in Germany and France. Again doing robustness checks helps to check whether the findings remain the same when giving other weights per two cars[12]. The results are shown in Table 5.

Table 6. Robustness checks, same weight for every year assumption

	Linear regression	**Fixed effects regression**	**2sls IV regression**
	Taxrate2022	Taxrate	Taxrate2022
Fuelefficiency	-294.3871**	4.228249	-952.0424*
	(142.9195)	(6.994372)	(530.203)
Constant	1127.898***	543.0264***	2406.711**
	(281.9066)	(14.96513)	(1018.278)
Number of observations	26	130	26

Note: The robust standard errors are shown in parentheses. *p-value<0.1, **p-value<0.05, ***p-value<0.01.

Upon conducting further robustness checks and controlling for validity, there remain differences between the results of these checks and the initial findings. Interestingly, this time the coefficients tend to be lower when controlling for validity. However, similar to the previous robustness checks, these differences are not substantial, and in fact, they are even smaller than the differences observed in the first round of checks. Once again, the coefficient in the linear regression exhibits increased significance compared to the initial results. Moreover, the constant terms remain significant, as seen in the original findings, and their magnitudes remain largely unchanged. These results indicate that, despite some variations, the overall patterns and conclusions remain fairly consistent.

It is worth acknowledging that this study is being conducted as part of a bachelor's thesis, and due to time constraints it was not possible to conduct more comprehensive robustness tests but based on the available results of the robustness checks that are done above, it has been reasonably argued that the underlying assumptions presented in this study are not entirely unfound.

[12] The new weights are: 1st and 2nd car-0.473; 3rd and 4th car-0.375; 5th and 6th car-0.312; 7th and 8th car-0.257; 9th and 10th car-0.204, based on the averages of France.

V. Conclusion

In this thesis, the main objective was to investigate the relationship between the car fleet of a country and the taxes imposed on gasoline. By employing various regression techniques including linear regression, country fixed effects regression, and a 2SLS regression with density as the instrumental variable, we aimed to uncover any causal effects and understand the underlying factors influencing the tax rates.

The importance of this relationship is again underlined by discussing the paper of Delfgaauw and Swank. Their model concluded a relationship between people's choice of what car to buy and public policies. Through a short literature review, it was determined that reverse causality could potentially be a confounding factor in this relationship. Therefore, it was crucial to employ empirical analysis to establish any meaningful conclusions.

The findings from the regression analysis revealed that while the coefficients of the linear regression and the 2SLS regression were statistically significant, indicating a relationship between the car fleet and tax rates, no causal effects could be established. Both regressions have namely limitations and could not meet some important assumptions. It is important to note that the coefficient of the country fixed effects regression was not significant, potentially due to limitations in the available data.

These results suggest that while there is an association between the car fleet composition and tax rates imposed on gasoline, further research with more data availability is needed to establish a potential causal relationship. It is likely that other factors not considered in this study, political and economic factors, may also contribute to the determination of tax rates. For now, it was not possible to control for such factors to measure the exact effect of the regression.

Some suggestions for further research include extending the data collection period to encompass more years, enabling the inclusion of stock fuel efficiency data over a longer time frame and potentially leading to significant results in a country fixed effects regression. Additionally, gathering information on the availability of public transport and employing it as an instrumental variable could provide valuable insights. Furthermore, enhancing the validity of the linear regression model could be achieved by identifying and incorporating relevant control variables. The inclusion of such variables would strengthen the model's robustness and improve the overall validity of the findings.

In conclusion, this thesis provides valuable insights into the complex relationship between the car fleet of a country and the taxes imposed on gasoline. While no causal effects were found, the significant coefficients in the linear regression and 2SLS regression indicate the presence of an association. Future studies should aim to address the limitations of this research and explore additional factors to gain a more comprehensive understanding of this relationship.

References

Allcott, H., & Wozny, N. (2014). Gasoline prices, fuel economy, and the energy paradox. *Review of Economics and Statistics*, *96*(5), 779-795.

Barla, P., Harding, C., Miranda-Moreno, L., Patterson, Z., & Zahabi, S. A. H., (2012). Transportation greenhouse gas emissions and its relationship with urban form, transit accessibility and emerging green technologies: A Montreal case study. *Procedia-Social and Behavioral Sciences*, *54*, 966-978.

Bento, A. M., Goulder, L. H., Jacobsen, M. R., & Von Haefen, R. H. (2009). Distributional and efficiency impacts of increased US gasoline taxes. *American Economic Review*, *99*(3), 667-699.

Delfgaauw and Swank (2023). The gasoline climate trap. *Tinbergen Institute.*

Kim, E., & Kim, H. S. (2004). Effects of public transit on automobile ownership and use in households of the USA. *In review of urban & regional development studies: journal of the applied regional science conference*, *16*(3), 245-262.

Gannon, G., Gearheard, S. F., Gerlach, S. C., Hamilton, L. C., Huntington, H. P., & Loring, P. A. (2018). Staying in place during times of change in Arctic Alaska: the implications of attachment, alternatives, and buffering. *Regional Environmental Change*, *18*, 489-499.

Li, S., Linn, J., & Muehlegger, E. (2014). Gasoline taxes and consumer behavior. *American Economic Journal: Economic Policy*, *6*(4), 302-342.

Parry, I. W. and Small, K. A. (2005). Does Britain or the United States have the right gasoline tax? *American Economic Review, 95*(4), 1276–1289.

Used websites

https://www.anwb.nl/auto/tests-en-specificaties?icp=auto_algemeen-V1B_navblok_2_2.2

https://bestsellingcarsblog.com/

https://www.best-selling-cars.com/

https://www.statista.com/statistics/937796/pump-price-and-tax-of-petroleum-per-liter-by-country-eu/

Appendix

Figure 6 shows the correlation between the gasoline taxes in euros/1000 litres and fuel efficiency in litres/100 kilometres over the years 2022-2018.

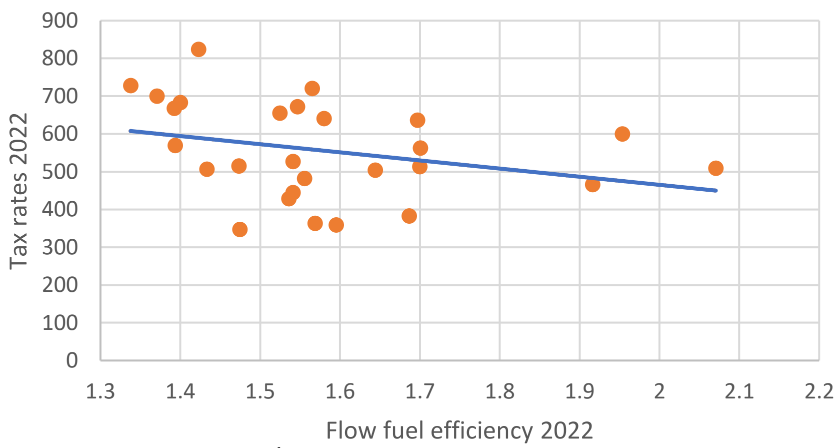

Figure 6.1. Scatterplot 2022

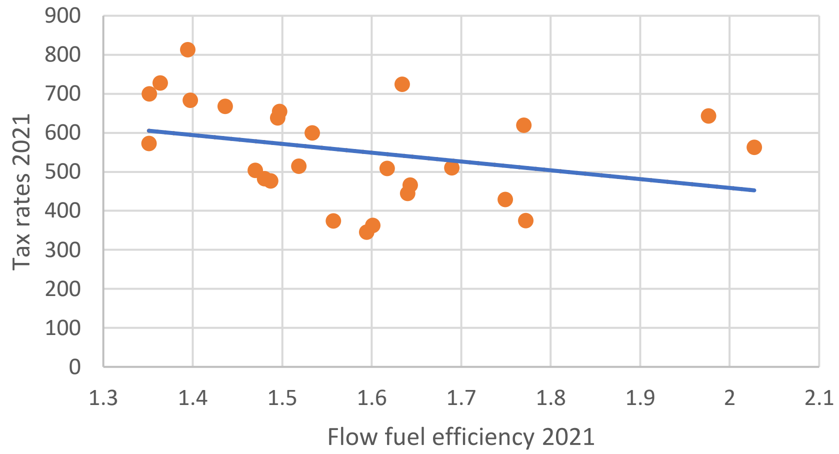

Figure 6.2. Scatterplot 2021

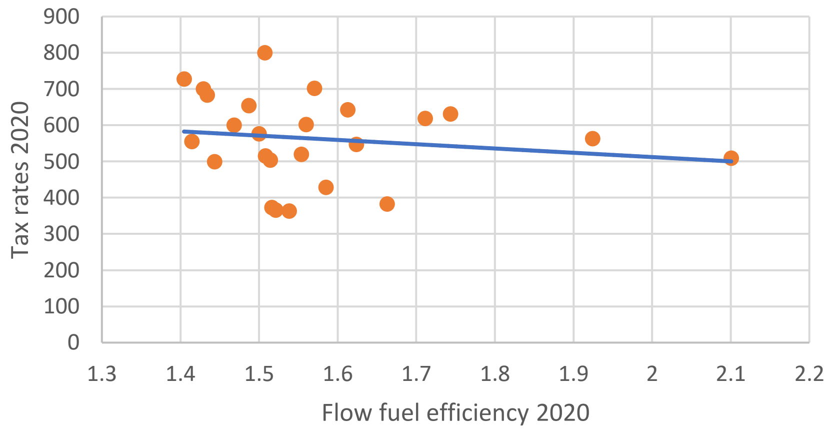

Figure 6.3. Scatterplot 2020

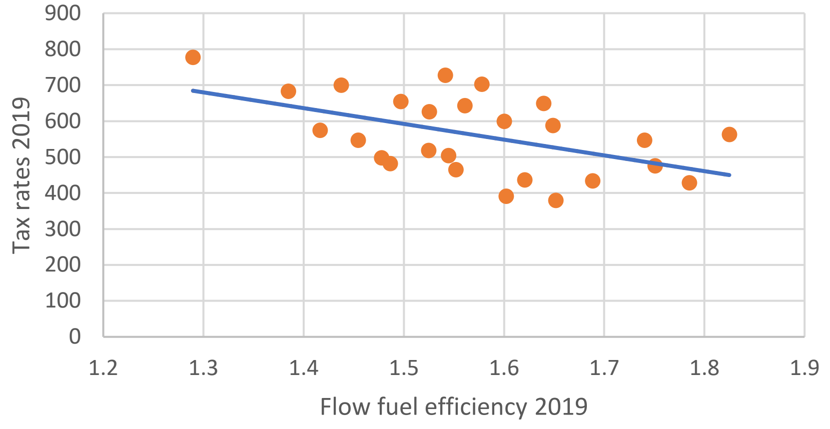

Figure 6.4. Scatterplot 2019

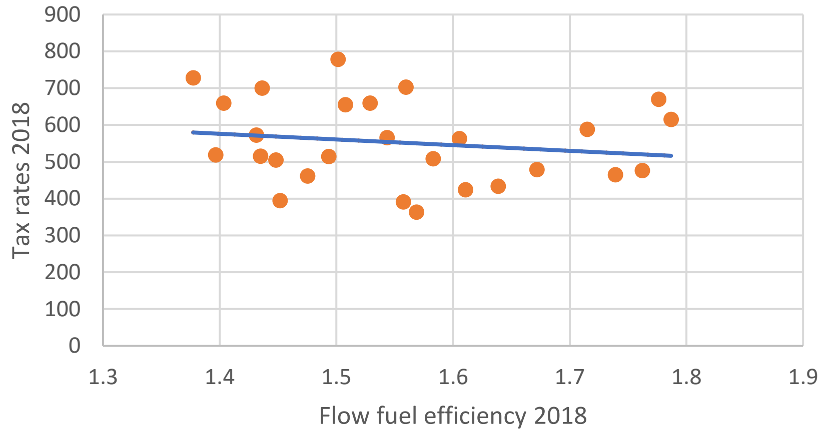

Figure 6.5. Scatterplot 2018

Figure 7 shows a residual plot, a scatterplot of the residuals against the independent variable.

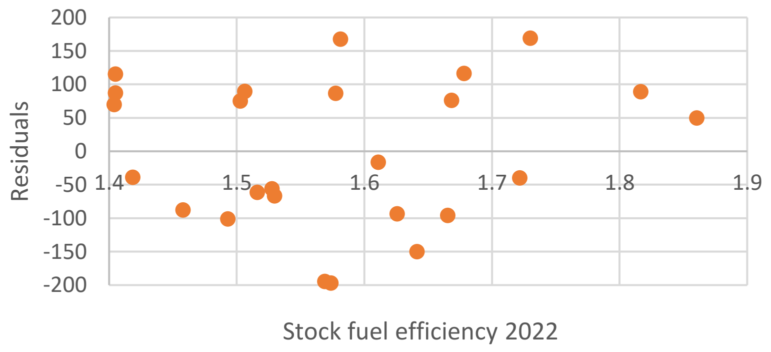

Figure 7. Residual Plot