Institut für Thermische Strömungsmaschinen
Prof. Dr.-Ing. H.-J. Bauer, Ord.

Modelling of the development of the European passenger car market

Masterthesis
von
Gabriel Möring Martínez

Betreuer: **Cihan Ates, Dr.-Ing.**
Patrick Jochem, Dr. rer. pol.
Uwe Langenmayr
Murat Senzeybek

September 2022

Karlsruher Institut für Technologie

Ich versichere, die Arbeit selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel verwendet zu haben. Die wörtlich oder inhaltlich übernommenen Stellen sind als solche kenntlich gemacht. Die Satzung des Karlsruher Instituts für Technologie (KIT) zur Sicherung guter wissenschaftlicher Praxis wurde von mir in der aktuellen Fassung beachtet.

Karlsruhe, 30. September 2022

Acknowledgement

I would first like to thank my thesis supervisor Murat Senzeybek in DLR. He has oriented me during all the thesis and has made me become a better researcher. He was always willing to help and has dedicated many hours to help me improve this master's thesis.

I would also like to thank Prof. Dr. Jochem in IIP, who offered me the possibility to apply for this thesis. Prof. Dr. Jochem and Uwe Langenmayr have helped me through this thesis.

Additionally, I would like to thank Dr. Ates, who was always willing to help. He dedicated the time that I needed and helped me to improve the clustering model, among other things.

I would also like to thank all "Institut für Fahrzeugkonzepte" in DLR. Many colleagues helped me when I was stuck at certain points of the thesis and have made my job easier.

Finally, I must express my very profound gratitude to my parents, Maria Luisa and Cristóbal, my grandparents, and my girlfriend for providing me with unfailing support and continuous encouragement throughout the process of researching and writing this thesis. Living in Karlsruhe and working in Stuttgart was not easy. You kept me going during these six months. This accomplishment would not have been possible without them.

Contents

List of Figures

List of Tables

1 Background

The directives of the European Union are leading the transition toward reducing CO2 emissions. The field being addressed is road transport, specifically passenger car transport. In January 2020, vehicle manufacturers had to meet an average CO2 emission standard of 95 grams of CO2 per km traveled in New European Driving Cycle (NEDC) tests for passenger cars. In 2020, this only applied for 95% of the new car registrations, but since 2021 it applies to 100% [30] and Worldwide Harmonised Light Vehicles Test Procedure (WLTP) has been introduced. This threshold would be reduced successively to reach zero emissions in the future.

In order to do that, more efficient cars in terms of emissions are being sold as hybrid and in 2020 there has been a huge intake of plug-in hybrid electric vehicles (PHEV) and battery electric vehicles (BEV).

Other technologies are being developed, like alternative fuels such as bio-diesels or fuel-cell powered vehicles. Depending on the technological progress and the investment produced, they will have a more or less considerable influence on the future automotive fleets.

Once this brief introduction has been done, European Union sets the directives and directly affects car manufacturers (supply-side) and how the fleets are composed. The policy selection to achieve the targets is not specified, and national countries will significantly influence this transition. They can set different targets from the EU, directly affect the supply and demand side, develop the necessary infrastructure, create social awareness...

It wants to be studied how different demand-side policies and external factors can affect the evolvement of European vehicle fleets at a national level under a reference scenario. There are 27 European Union (EU) countries, which would make analyzing each country separately not only an exhaustive task, but also it may be possible there is a lack of data of certain countries, making the task impossible. Then, this thesis is split into two main modules:

- Grouping of European countries by automotive fleet market characteristics. (See Chapter 2)
- Modelling of the future development of passenger car sales up to 2035. Selected EU countries and the EU as a whole are modelled. (See chapters 3, 4 and 5)

2 Grouping of European countries by automotive fleet market characteristics

2.1 Abstract of grouping European countries

2.1.1 Purpose of grouping countries

European Union sets directives, and as explained in 1, different countries apply different policies for achieving the passenger car CO2 sales targets. Experience gained in previous policies and measures taken by other countries can be a significant source of knowledge for applying policies efficiently and reducing inefficiencies.

The purpose of grouping countries is to model the EU passenger car sales up to 2035. If countries are not grouped initially, an exhaustive amount of data should be collected for 27 countries. Grouping allows to have differentiated passenger car markets and assumes the development of the passenger car market share will be the same inside these groups. Additionally, finding similar automotive markets facilitates learning from other countries and knowing which policies can be more suitable for a particular country inside each country group.

European Union (27 countries) data was collected, plus Norway and the United Kingdom. Norway is included due to the leading role in electric vehicle (BEV+PHEV) intake; 75% of the new car sales were electric, followed by Sweden, where 32% were electric. Great Britain is added due to its proximity and market size.

2.1.2 Design/methodology/approach

An extensive literature review identified country factors that can affect the situation of national markets and their evolvement. The factors can be used to assess the similarity of the national markets. Cluster analysis was performed using hierarchical clustering and Principal Component Analysis (Dimensionality reduction tool).

2.1.3 Findings

Seven country clusters were obtained and validated using Silhouette Score. All clusters had at least one principal component mean, which was statistically significant from the rest of the cluster means (95% confidence level). Clusters 1 and 2 did not have statistically significant different means, but they have Silhouette scores that validate each cluster.

The clusters may facilitate learning from other national policies, such as public infrastructure, social awareness campaigns, and supply and demand-side policies. Countries from the same cluster should have similar outcomes when applying the same policies (This has not been verified yet).

2.1.4 Research limitations and implications

It is a data-driven approach. The finding depends on the factors selected to be relevant for describing the evolvement of the automotive fleet. Literature has been studied, and researchers' experience in DLR has been used.

Another limitation is that this analysis is cross-sectional. Although the findings cannot be generalized and clusters are not static, it is believed that the methodology used in this research is repeatable for different countries considering different time frames. Variables in relative terms (for example, the evolution of RES share in the last ten years) can help have a model that better captures the market's dynamics.

The last limitation can be finding data if the model wants to expand internationally. Extensive data collection work would be required, and sometimes it may not be possible to get specific data. If the amount of the data can not be obtained only for a small number of countries, there are different tools for dealing with missing data in clustering models.

2.1.5 Originality

Theoretical contributions include identifying factors that can be used for similarity assessment of countries for transferring lessons learned as well as a methodology for clustering countries regarding the similarities of their vehicle fleets and the reduction of CO_2 emissions transition. Findings may also have practical value for European policymakers.

2.2 Scope of grouping European countries

Clustering of national automotive fleet markets serves to identify groups of countries with some degree of similarity and explain what factors account for these similarities. The objective of doing this is to reduce the complexity of analyzing each automotive market separately, reducing the exhaustive task of obtaining information for all the countries. The results of one country of a cluster should be similar to the rest of the countries inside the cluster.

Empirically, it allows us to determine also what drives national automotive fleet development and what causes countries to be considered similar, and as a result, develop implications for country policymakers.

Policymakers could examine cluster membership to determine which could be the most optimal policy decisions.

2.3 Review of selected literature

Much research explains what defines the current vehicle fleet and what can define the automotive fleet evolvement. There are bigger automotive sectors that have higher registrations of passenger vehicles every year. These markets can evolve faster than others if they adopt new technologies. Factors explaining the market dynamics are mainly macroeconomic, like GDP per capita, and other factors, like the size of the market (for example, car ownership per person) and new registrations.

Adopting new technologies requires different conditions to be fulfilled. Firstly, economics as has been explained in the dynamic of the market. Alternative fuel vehicles include various types of technologies. Generally, they have a high acquisition cost [39], or if not, they present other barriers, which limits a considerable market penetration. The electric car has reduced its price (mainly because of the reduction of the cost of battery cells), but affordability remains a major barrier to consumers. In 2020, the market uptake of electrically chargeable vehicles (ECVs) is directly correlated to a country's GDP per capita. [2]

Additionally, another principal barrier is linked to the refueling infrastructure scarcity [39]. Although particular technologies determine environmental impacts, Grübler et al. show that the selection of technologies is partially governed by system attributes, such as the availability of necessary infrastructures [61]. Apart from technical (development of new technologies and/or implementation of the necessary infrastructure) and economic development, an institutional factor is crucial to add. These are regulations (supply-side policies and targets) and economic incentives (demand-side policies). In the vehicle market, decarbonization is a strong reason for forceful policy interventionism [77].

Once the economic and market conditions are defined, according to Yan [145], environmental awareness is essential when selecting a BEV. At a national level, this is included inside effort sharing regulation which translates the environmental commitment into binding annual greenhouse gas emission targets for each Member State for the period 2021–2030, based on the principles of fairness, cost-effectiveness, and environmental integrity. Apart from that, it remarks on customer demands as travel demand.

2.4 Selection of features

Different studies have shown different barriers and reasons which limit the intake of alternative fuel vehicles. Each country in the EU has its macroeconomic situation, vehicle market, infrastructure, future targets, and social awareness. EU wants to be climate neutral by 2050 [3], and in the short term, there are other objectives. For example, by 2030, there is a target 30% emissions cut in non-ETS sectors compared to 2005 levels [48]. Effort-sharing legislation or regulation sets the national target for different sectors, which account for more than 60% of total EU emissions. The main sectors covered are transport, energy consumption in buildings, agriculture, small industry installations, and waste management. The national targets differ and are not aligned with the EU. The target set shows the effort a country is willing to take and at which point in the decarbonization phase expects to be in 2030. Regarding vehicle market evolvement, the targets show the compromise each country has for applying different policies to decarbonize their own automotive national market.

Once the relevant features have been defined, there are classified into three main groups (There is a different group that will not be used for clustering):

- **Macroeconomic indicators**: An example is GDP per capita
- **Current vehicle market situation, targets for national automotive fleet and infrastucture**: An example is the number of public chargers or the current market vehicle distribution.
- **Social situation**: An example is electric vehicle awareness programs
- **Supply-side and demand-side policies**: An example is purchase incentives like subsidies or tax discounts. Taxation systems will be studied afterward when the countries have already been grouped

Data collection starts once the type of features that are required are defined. Macroeconomic data such as GDP per capita can be found in International Monetary Fund, targets, vehicle market situation and infrastructure can be found in European Automobile Manufacturers' Association (ACEA), European Environment Agency, and other sources.

There are 29 countries in the clustering analysis, and the main groups selected, and their features are described. Sources used can be consulted in Appendix 7.1.

2.4.1 Macroeconomic data

Initially, it was considered to include GDP per capita, public debt, and long-term interest rates when including macroeconomic data features. It had initially been considered public debt to be a factor that limits public investment. Indeed, in [106] using the Generalized Method of Moments from Arellano and Bond [12], it was obtained that a 1% increase in public debt reduces public investment in the European Union by 0.03%. However, it is obtained that the negative impact of debt on investment is slightly smaller in the Eurozone than in the entire EU. This distorts the indicator because the countries outside of the Eurozone need to have smaller debt overhangs to get credit at the same interest rate in financial markets.

Observing this, instead of using gross debt as a feature, it can use long-term interest rates (government bonds). Interest rates are affected by the currency, and it is also correlated with public debt because if financial markets start questioning the sustainability of a country, they will demand higher interest rates to compensate for the increased default risk. [37] To conclude, higher interest rates make credit more expensive and, thus, limit public investment.

The zero-bound limit of long-term interest rates will limit a bit the efficiency of this indicator. The bound can be explained by the secular stagnation theory reintroduced by Larry Summers [129]. Near zero long-term interest rates are favorable for investment, and Next Generation EU was approved under these conditions. Next Generation EU will boost the greener transition [113].

The variables selected for describing the macroeconomic situation of the countries are:

- **GDP per capita**: GDP per capita in 2021 at purchase power parity at current international US $. GDP per capita is a good indicator for showing the wealth of a country. A higher GDP per capita means higher resources for a country if the taxation is equal. It allows the country to focus on innovations and develop future infrastructures, which will benefit the long term. Conversely, for a country with lower GDP per capita, in the short term, the resources need to be concentrated on other problems which are more urgent or at least look more urgent to the population.

- **long-term interest rate**: 10 year governement bonds interest rates in February 2022. Expressed in %. This shows investors' reliability in a specific country and the interest rate they demand to invest. This parameter is heavily influenced by the level of indebtedness and the currency's strength. The euro area allows countries to have a higher level of indebtedness and still have a low-interest rate. An example is the case of Slovenia (0.46%) compared with Hungary (2.82%), presenting a very similar level of indebtedness or Ireland (0.58%) and Poland (2.53%). Hungary and Poland do not have the euro as a currency which will allow them to devaluate their currency more easily, generating more uncertainty in investors.

2.4.2 Current automotive market situation, targets for national automotive fleet and infrastucture

The decarbonization of the vehicle market will rely on a diverse portfolio of technical solutions [100]. Different factors will determine which technology establishes, mainly technical, economic, and institutional [77].

A scenario based on hydrogen can not be discarded in the medium-long term. However, the current national markets are developing specific targets for alternative vehicles composed mainly of biofuels and rising uptake of electric vehicles.

It has been considered a good way of describing how advanced the vehicle markets are in their transition to zero-emission the following features:

- **greenhouse gas (GHG) intensity reduction of road transport - % of decrease between 2010 and 2019**: This indicator measures the reduction of GHG emission intensity of fuels sold by road transport between 2010 and 2020. For better understanding, an example could be combining gasoline with bio-fuels as, for example, E10, which is the gasoline that has a 10% of bio-ethanol. Bio-ethanol does not produce greenhouse gases considering its full life cycle. This feature is relevant because bio-fuels have been the main responsible for the use of a 10% of renewable energy for transport in Europe by 2020, and electric vehicles still play a small role by 2020. [138]

  The EU had a target to reduce the GHG intensity of fuels to 6% in 2020 with respect to levels of 2010. This target was set in RED and, in 2019, adjusted for 2030 with a more ambitious target [17]. In 2019, the levels decreased by 4.3%. The reduction is mainly due to the use of biofuels. Finland and Sweden have only accomplished the intensity reduction target of 6%. Outside the EU, Norway has accomplished as well the EU target. It is important to remark that if indirect land use change is accounted for, most countries have smaller reductions, and the average intensity reduction in UE is 3.4% instead of

4.3%. The values used include indirect land use change because this is the relevant value to know how much the CO_2 emissions have been reduced.

- **Effort Sharing Regulation target for 2030**: Effort sharing legislation or regulation sets the national target for different sectors, which account for more than 60% of total EU emissions. [46] The target is set for the year 2030. Expressed in %. The main sectors covered are transport, energy consumption in buildings, agriculture, small industry installations, and waste management. The target for 2020 at the EU level was to have a reduction of 10% with respect to 2005, and a reduction of 15% was achieved. For 2030, the aim is a reduction of 30% with respect to 2005 levels. The national targets differ and are not aligned with the EU. This shows the effort a country is willing to take and at which point the decarbonization phase expects to be in 2030. In 2020, 21 out of 27 accomplished their targets. 3 out of the 6 remaining were very close to the target without applying flexible mechanisms. This shows that the national targets are good indicators of the effort a country is willing to take.

- **Renewable Energy Source share in 2020**: This indicator shows the share of energy from renewables used for transport in 2020. Expressed in %. The indicator is relevant because decarbonization at a considerable level is only possible if it is invested in renewable energy sources. Biofuels, electric, and hydrogen vehicles can be zero-emission only if renewable energy is used.

  Pursuing the target of 10% renewable energy share for transport in the EU [117], the share increased from under 2% in 2005 to 10.2% in 2020. As it has been said above, the increase of last years in all EU was explained mainly by the high intake of biofuels. The electrification of fleets still plays a small role in this share. The target for 2020 was achieved because it was 10%. However, there is a very diverse share achieved by countries in 2020. The target defined in RED II (European directive) is to increase to 14% the renewable energy used in transport by the year 2030.

- **New registrations of electric cars**: Electric cars consider battery electric vehicles (BEV) and plug-in hybrid electric (PHEV). This indicator shows the current share of new electric car registrations in 2020. Expressed in %. Passenger cars are the alternative technology with faster development in recent years, and a huge intake is expected in the coming years. Electric cars have had a massive intake in the EU by 2020. It increased the share from 3.5% to 11% in just one year. (from 550,000 units in 2019 to 1,325,000 units in 2020). BEV accounted for the majority of electric vehicles over PHEV.

  In the past, the total cost of ownership (TCO) has been considered a significant barrier to adopting electric vehicles. According to [50], the economic competitiveness of BEVs is the most crucial element of the consumer adoption decision, and the upfront costs weigh heavier in the decision than savings when using the vehicle [62]. BEV had a much higher retail cost than ICE vehicles due to the cost of batteries. In recent years, the TCO has decreased due to the exponential decrease in the cost of batteries. It is expected to have a breakeven point with ICE vehicles in the second half of 2020. [60]. This will allow the intake of electric cars to keep growing. Other factors such as range and infrastructure will still be disadvantages for customers when selecting the type of drive train and will act in the opposite direction.

- **Average of vehicle age**: This indicator shows the average age of passenger cars in 2020. It shows how dynamic each national market is. This is relevant because a transition could be faster in a country where vehicle age is lower. There are considerable differences among countries: EU cars are now, on average, 11.8 years old. Lithuania and Romania have the oldest car fleets, with vehicles almost 17 years old. On the other hand, the newest passenger cars with only 6.7 years can be found in Luxembourg.
- **Automotive industry size**: This indicator is the employment share the automotive industry represented in 2019. Internal combustion engine (ICE) production gives work to a large population in certain countries. The transition to electric engines will considerably reduce the number of workers in automotive workers because the complexity is much lower with respect to ICE. [87] Then, countries with a higher share of workers in the automotive industry will have lobbies to defend the workers. Initially, it was considered to use the automotive industry's GDP as well. However, they had a high correlation, and if both features are combined, the results are too much influenced by the industry's size. Thus, only one feature has been selected.
- **Public electric chargers infrastructure**: Public electric chargers infrastructure is measured with two indicators:
  - **geographical electric public charger density**: Charging points per 100 km
  - **electric public charger availability**: Electric vehicles per charging point

  Two indicators are selected due to the relevance of the infrastructure in the intake of new technology. As explained in 2.3, infrastructure is a major barrier [39]. Bio-fuels do not need infrastructure, but their growth is moderate compared with electric vehicles. Electrification of the roads requires a huge investment in public chargers.

  Public electric chargers are required if the uprising growth wants to be sustained. The massive growth of registrations has made the electric vehicle per charging point grow from 2 cars per charging point in 2010 to 11 in 2020. This growth occurred even with infrastructure built at a robust pace. Based on European Commission calculations, a further decrease of $CO2$ of 50% based on 1990 levels would require 6 million publicly available infrastructure points, translated into 27-fold in less than a decade. [2]

  Regarding the second indicator, inhabitants per charging point, the growth in the different electric vehicle markets have been very uneven. 73% of all European sales are concentrated in 4 western countries with some of the highest GDPs per capita, and three countries (Germany, France, and the Netherlands) concentrate 70% of all public chargers in Europe. These countries, especially the Netherlands, present an excellent infrastructure. Outside the EU, it is worth mentioning Norway, which has had an enormous EV uptake, and the infrastructure is also at the level of the leading European Union countries mentioned before. Finally, the last indicator is important because charging points per 100 km show the possibilities to connect among European countries. The low public chargers in certain countries show that the difficulties of traveling with EVs across Europe will still be complex. Finally, it is important to remark that public chargers have not been separated between fast chargers and normal chargers. This affects countries like the UK, which has a large number of fast chargers compared to other countries.

2.4.3 Social awareness

According to Wappelhorst et al. [142], key barriers to alternative vehicles are affordability, convenience, and awareness. Vehicle affordability (total cost of ownership) and convenience (range, infrastructure) have already been considered in other features. At a macro level, awareness is partly fixed by the different targets each country is setting. However, at a micro level, this factor has not yet been considered. This can be done through information campaigns to raise awareness and increase visibility regarding available alternative fuel vehicles and their benefits.

Each country has a different level of awareness, and it would be perfect to have a specific indicator that measures the awareness regarding ICE pollution and how prone they are to use alternative fuel vehicles under certain circumstances. Due to the lack of data and the lack of resources for performing a 29-country interview for vehicle customers, it has been selected an interview done at a European level that measures the level of awareness regarding climate change.

The indicator used has been the following:

- **climate change awareness indicator**: It measures environmental awareness, and it is expressed in %. Social awareness is a very important factor in the transition to zero-carbon emission. If citizens consider climate change a major problem, they find it legitimate that the investment made in this area is huge. This area includes transport, electricity generation and other fields. Investments for developing infrastructure, giving subsidies, and making tax reductions. Citizens of a country with a lower GDP per capita or a smaller consciousness of global warming will be less prone to see adequate to promote green technologies and measures. In order to measure this, a poll done by Eurobarometer in October 2021 is used for European Union countries. The question is: *which of the following do you think are the main challenges for the EU?* They can choose maximum 3 answers and some of them are *ageing of population*, *unemployment*, *social inequalities...* The answer which is relevant for us is *environmental issues and climate change*. In many countries *environmental issues and climate change* is the answer which is most voted to show awareness for climate change.

All variables have been considered to have the same relative importance. Manual weighting has not been performed in these variables.

2.5 Methodology

Once the data is selected, the methodology consists of exploring the data and analyzing the type of problem that is being solved.

The objective is to define a certain amount of characteristic groups (clusters) that group countries by their similarities in the automotive market. A clustering problem is being solved because these countries have not been labeled, and results cannot be tested as it is done in classification or regression problems.

The labels will be defined once the clusters have been formed. The objective is to form 5 or 6 clusters. For labelling, an analogy to Rogers curve can be used depending on how new alternative technologies have penetrated the different countries (Further explanation of Rogers curve in Section 2.5.2.4.2).

The process followed for solving the problem can be divided into:

1. **Data exploration and preparation**
    - Correlation matrix
    - Scatter plots
    - Outliers
    - Standardisation
    - Variance inflation factor
2. **Selection of the baseline model and approach to the problem**
    - Baseline models and first results
    - Selection of the baseline model
    - Multicollinearity reduction (if required)
    - Interpreting new orthogonal space (if dimensionality reduction techniques are applied)
3. **Development of the baseline model selected and analysis of the results**
    - Hyperparameters tuning
    - Final model selection
    - Significance of results
    - Interpretation of results and labelling

Once the process has been shown, different aspects of each part that are relevant are going to be briefly described.

2.5.1 Data exploration and preparation

The data is collected in a CSV file, and the problem is solved using Python, more specifically using a Notebook. There are no missing values, and all values are numerical.

Once this is checked, the statistical values of variables are analyzed. Means, standard deviation, and the different quartiles can be seen. This helps to see how equally or unequally distributed data is for each feature.

2.5.1.1 Correlation matrix

After observing the basic statistical values, it is analyzed how correlated one feature is with the other. This is done by plotting the correlation matrix (Figure 2.1).

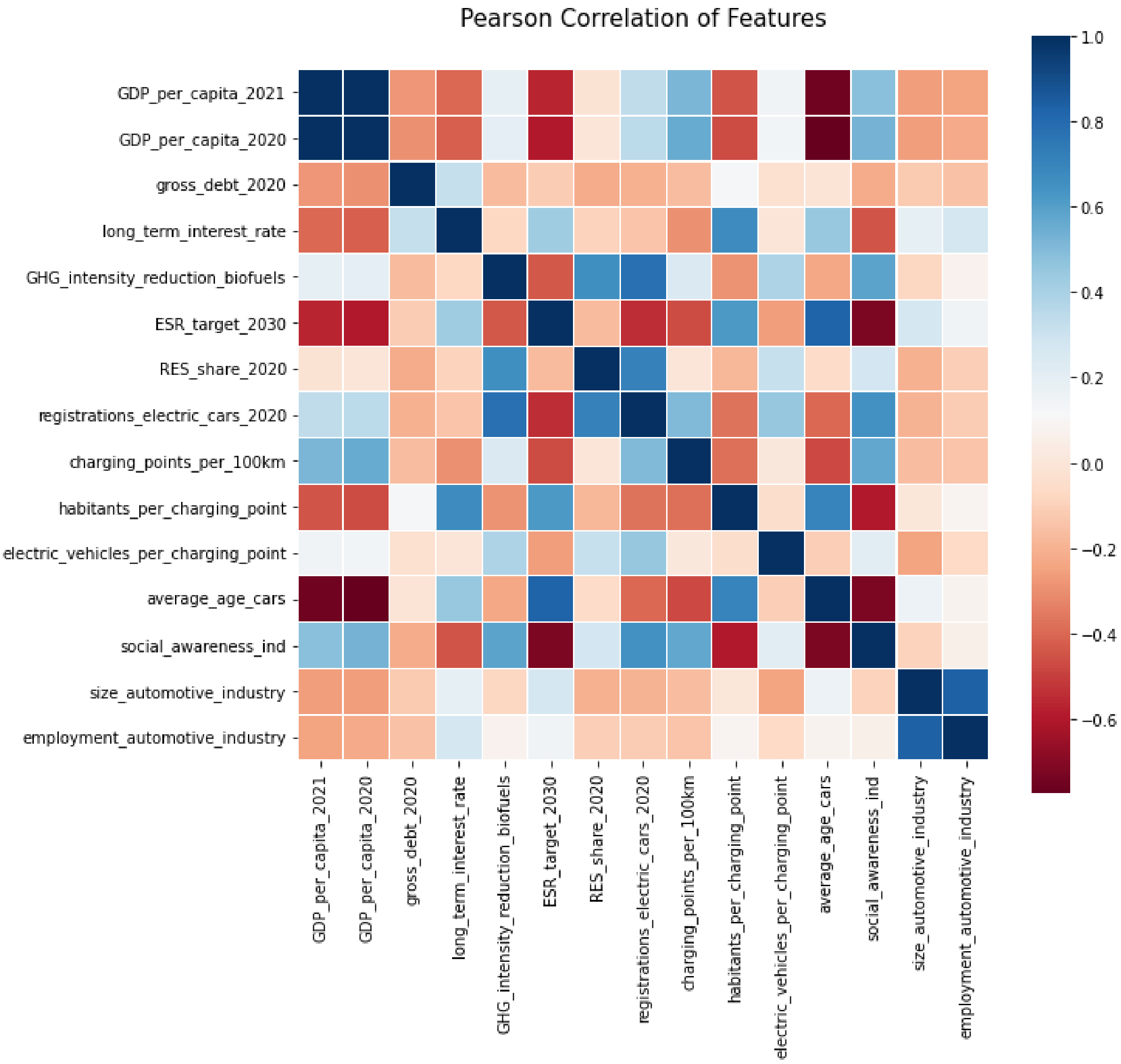


Figure 2.1: Correlation matrix of the features selected

Graphically, it can be observed that there are relatively high correlations between many features higher than 0.6 and lower than -0.6. After that, the highest correlation coefficients are shown in Table 2.1:

feature 1	**feature 2**	**Pearson correlation coefficient**
RES_share_2020	**registrations_electric_cars_2020**	**0.713945**
social_awareness_ind	**average_age_cars**	**0.724123**
average_age_cars	**GDP_per_capita_2021**	**0.753348**
average_age_cars	**GDP_per_capita_2020**	**0.776415**
registrations_electric_cars_2020	**GHG_intensity_reduction_biofuels**	**0.780945**
ESR_target_2030	**average_age_cars**	**0.831014**
employment_automotive_industry	**size_automotive_industry**	**0.835934**
GDP_per_capita_2021	**GDP_per_capita_2020**	**0.994621**

Table 2.1: Highest correlation coefficients of the selected features

It can be observed there are different correlations observed among different terms, but over 0.90 it can only be observed in the GDP per capita in 2021 and in the GDP per capita in 2020. It has a correlation of 0.995 which means that GDP per capita in 2020 should not be in the model. It was deleted because it was an additional dimension without giving extra information for classifying countries. The size of the automotive industry (which is the share of GDP of the automotive industry) and employment of the automotive industry have a high correlation, 0.836, and both are trying to measure the automotive industry size. Then, it is decided to delete the size of the automotive industry and keep the employment in the automotive industry.

Additionally, it can be observed a correlation above 0.8, which is the ESR target of 2030 and the average age of cars. This is a very high value, and regarding the relevance of both terms, it will be decided later if they should be deleted.

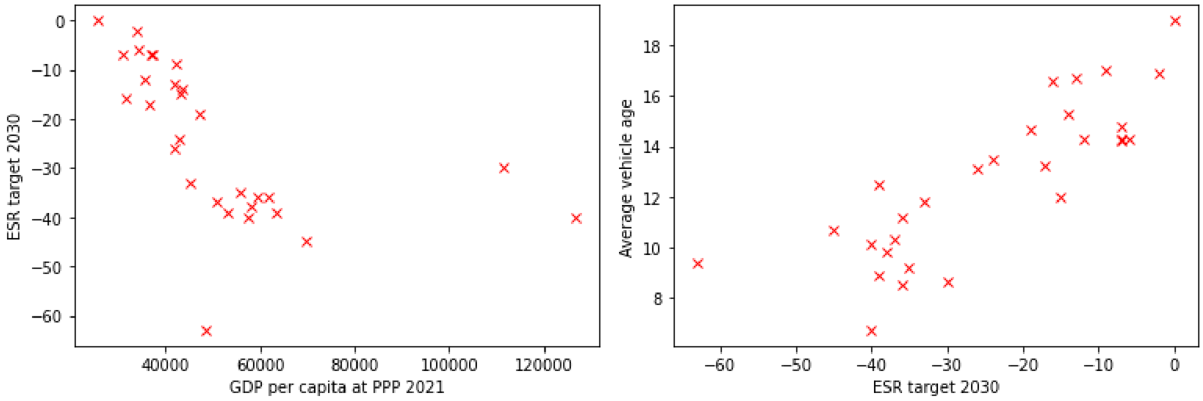
High correlations can be observed in the model, and this may cause problems. It will be analyzed if features that apparently represent different characteristics of the automotive market should be deleted or how to deal with this.

Note: Gross debt and habitants per charging points are parameters that are not included in the model. Initially, they were considered, but after revising the literature, gross debt was dropped (See in Section 2.3), and habitants per charging point feature was dropped because using three terms related to infrastructure was giving too much importance to infrastructure.

2.5.1.2 Scatter plots

Scatter plots are useful to observe visually how the data distribution looks between two features. The correlation between features can be observed visually. Two scatter plots are shown in Figure 2.2 where ESR target 2030 respect to GDP per capita is represented in Subfigure 2.2a and the average of vehicle age respect to ESR target 2030 is represented in Subfigure 2.2b.

From scatter plots, very interesting trends can be observed. In the first case (Figure 2.2a), a very clear negative correlation is observed. The higher the GDP per capita, the higher the national reduction target of greenhouse gases for the year 2030. The correlation between these two features is 0.72. It can be observed there are three outliers. Two, because they have higher GDPs per capita, and a third one which has a very high ESR reduction target.



(a) GDP per capita vs. ESR target 2030 (b) ESR target 2030 vs. average vehicle age

Figure 2.2: Scatter plot of selected features

In the second case (Figure 2.2b) it can be observed a positive correlation. The older the average age of the vehicle fleet, the lower the reduction target. This plot looks more spread, but it has a smaller amount of outliers. Because of that, it ends up having a higher Pearson correlation coefficient.

From Figure 2.2b it can be extracted that countries with a more dynamic market expect to reduce more greenhouse gases which is logical because they change cars much faster, and this allows to have a major intake of alternative vehicles where the share at the moment is quite low.

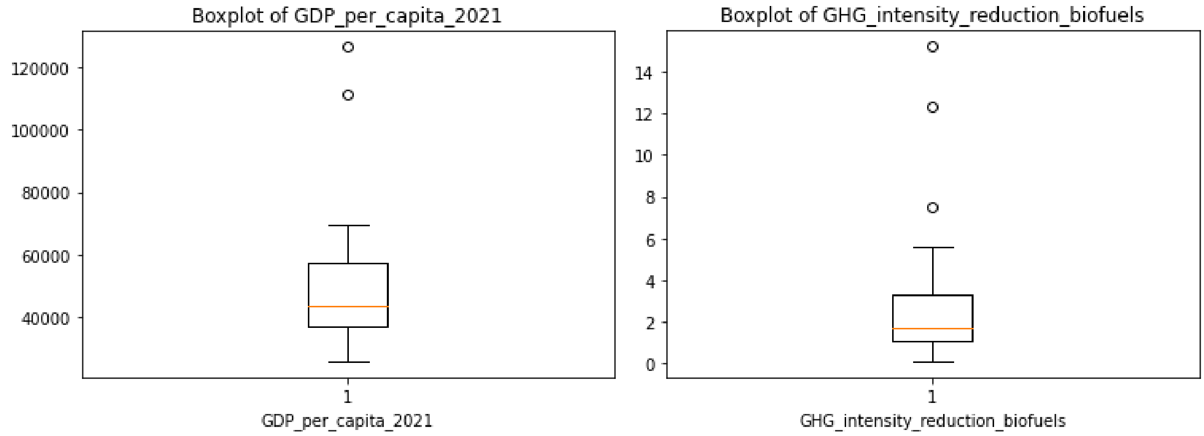
2.5.1.3 Outliers

Outliers have been observed. Luxembourg and Ireland are clear outliers when considering their GDP per capita with respect to other countries. Both have very attractive tax policies for corporations, and this is one of the reasons they have high GDPs per capita. However, the low taxation makes the development of greener technologies slower than what would be expected (compared to other countries and following the trend line). Transitioning to clean transport requires a big amount of public budget. On the other hand, Norway has very specific characteristics and, in many features, looks like an outlier. These conditions can make it difficult to group these countries with others. However, it is not something it can be concluded yet.

For making a better analysis of the outliers, box plots were used. Box plots are used to interpret graphically how each feature of the data is distributed and if there are outliers. The orange line is the median, the box is the interquartile range (defined by the first and third quartile), and finally, the T-shape whiskers go to the last point, which is still within 1.5 times the interquartile range. Points that are further are outliers and represented as points.

Each feature has been analyzed, and an example of two box plots for two different features is shown in Figure 2.3.

In Figure 2.3a, it can be observed that the outliers are Ireland and Luxembourg and far away from the rest of the countries. In figure 2.3b, it can be observed that the outliers are Norway and



(a) GDP per capita (b) GHG intensity reduction of biofuels

Figure 2.3: Box plot of selected features

Sweden and, finally, Finland, which is closer to the T-shape whisker.

After analyzing all box plots separately, it can be concluded that there are many countries that have been at least in one feature outliers. This will make the clustering a bit more difficult, and the clusters will definitively be less dense. Sweden and Norway are systematic outliers, and the Netherlands is as well. Finland is an outlier in two features and Portugal and Luxembourg are too. There are other countries that are outliers just once.

2.5.1.4 Standardization

It is important to remember that most machine learning algorithms work better if the data is normalized around zero; It has a mean value of zero with a standard deviation of one. Principal Component Analysis and other clustering methods are sensitive to data magnitudes. Here, feature standardization is shown in Figure 2.4.

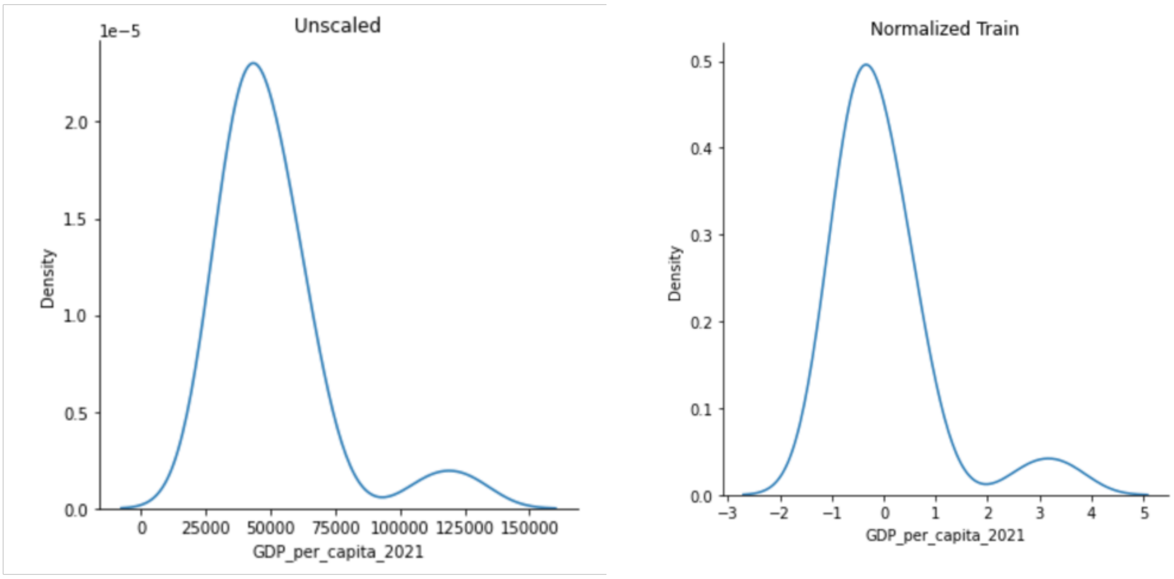
In Figure 2.4, it is observed how the GDP per capita has been standardised. Once it is scaled (Figure 2.4b), it can be observed the mean is around 0 and a standard deviation of 1. Standardization has been performed for all features.

In Figure 2.4, it can be observed something similar to a Gaussian with a small decay on the right for the case of GDP per capita 2021. The data set is small, and each of the features presents different distributions, which in most cases do not follow a Gaussian distribution.

2.5.1.5 Variance inflation factor

Before analyzing the baseline models, it wants to be analyzed, which is the level of multicollinearity among the features selected.

High multicollinearity is a problem because it undermines the statistical significance of an independent variable. [5] It exists whenever an independent variable is highly correlated with



(a) unscaled data

(b) scaled data

Figure 2.4: Standarisation of GDP per capita

one or more of the other independent variables in a multiple regression equation. This multiple regression equation is used as a theoretical base for the Variance Influence Factor (VIF).

VIF is used for analyzing multicollinearity of the model. In simple words, VIF is a metric that explains how well a feature can be explained by all the other features of the model.

It basically consists in fitting a regression model (for example, $x_1 = \beta_0 + \beta_1 x_2 + \beta_2 x_3 + \ldots$) to one of the features which is considered independent. An R^2 value is obtained for the model adjusted for a certain feature, and the formula of VIF is $\frac{1}{1-R^2}$. This is done for all the features. A possible classification can be:

- VIF = 1 $\Rightarrow$ R^2=0: No multicollinearity
- VIF between 1 and 5 $\Rightarrow$ R^2 between 0 and 0.894: Moderate multicollinearity
- VIF higher than 5$\Rightarrow$ R^2 higher than 0.894: High multicollinearity

The VIF obtained considering all features previously selected in Section 2.4 are presented in Table 2.2

A very high level of multicollinearity is obtained. 4 terms have VIF over 5. *average_age_cars* has a VIF of 10.03, which is equivalent to a R^2 of 0.95. The next three highest terms have correlations over 0.90.

This indicates that the clustering may be affected by multicollinearity. Reducing multicollinearity requires deleting features manually until VIF is reduced or applying tools such as Principal Component Analysis (PCA) which allows projecting the features in an orthogonal space and defining new principal components. Independently of the process followed, information is lost. If features are deleted manually, available features are no longer in the model, and if PCA is used, new principal components combine previous features. The new space is less understandable; usually, principal components containing a low variance (a low amount of information)

feature	**VIF**
GDP_per_capita_2021	3,79
long_term_interest_rate	1,88
GHG_intensity_reduction_biofuels	4,56
ESR_target_2030	5,95
RES_share_2020	4,87
registrations_electric_cars_2020	9,53
charging_points_per_100km	2,98
electric_vehicles_per_charging_point	1,54
average_age_cars	10,03
social_awareness_ind	5,59
employment_automotive_industry	1,46

Table 2.2: Variance Influence Factor of selected features

are deleted. Before reducing multicollinearity, baseline models are evaluated. Sometimes it is possible to obtain successful results without dealing with multicollinearity, and above all, without losing any information.

2.5.2 Selection of the baseline model and approach to the problem

2.5.2.1 Baseline models and first results

Kmeans

A good approach for clustering is applying Kmeans as a baseline model and observing which Silhouette score is obtained. Kmeans is quite straightforward: It is specified beforehand the expected number of clusters k, and the algorithm will assign each instance to one of these k clusters such that the sum of the within-cluster variations across all clusters is minimized.

Kmeans relies on geometric distance, and if the Kmeans Silhouette score is higher than 0.3, it can be considered that clusters are relatively dense and well delimited. If the model is above this threshold, it will allow to keep all features in the model and avoid using any technique for reducing multicollinearity where information of the variables is lost.

As a first approach to determine the number of clusters when it is not known, an inertia plot can be calculated. The algorithm tries to minimize the sum of the within-cluster variations across all clusters. Each time a new cluster is added, the total variation within each cluster is smaller than before until there is one cluster per point, and the inertia (variation) is 0. However, if the variation is plotted, it may be observable in a certain number of clusters where, when a new cluster is added, the marginal variation is much smaller than before, and this can be observed with an elbow plot.

The inertia can be defined as:

$$\sum_{i=0}^{n} \min_{\mu_j \in C}(||x_i - \mu_j||^2) \tag{2.1}$$

where μ_j is the mean of the samples in the cluster. Once it has been defined, the inertia plot is shown in Figure 2.5.

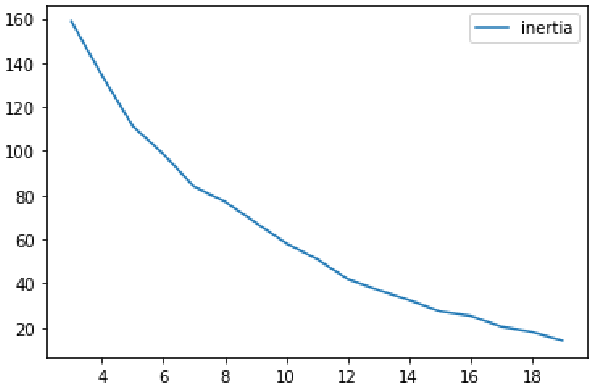


Figure 2.5: Inertia plot of K-means

It is hard to observe a very defined elbow. Thus, it is difficult to determine an optimum number of clusters. It is observed a first change of slope at 6-7 clusters, but very slightly, which can suggest that native space (the original coordinate system) is not easy to interpret. Once this has been said, the Silhouette score is evaluated for the different number of clusters to observe if it is high.

number of clusters	Silhouette average score
3	0,30
4	0,27
5	0,23
6	0,21
7	0,22
8	0,20
9	0,15
10	0,18

Table 2.3: Silhouette average score for different number of clusters using Kmeans

The objective defined is finding around 5 or 6 clusters. It has been observed in Figure 2.5, a change of slope around 7, but has obtained a higher Silhouette score for 5 clusters. However, the Silhouette score obtained for 5 clusters is 0.23, which is considerably under 0.3, and for 6 clusters, the Silhouette score is 0.21, which is even lower. Both values are clearly under the threshold. Having this Silhouette score value suggests the space is quite sparse, and multicollinearity is affecting results. Reducing multicollinearity and the number of dimensions will allow for more defined and solid groups. Silhouette score should increase.

DBSCAN

Another method that has been considered is DBSCAN which is another type of clustering method which has the characteristic feature of defining an outlier group. The maximum amount of groups that DBSCAN is able to create is three and one outlier group. The group of outliers is 9 Countries: Norway, Sweden; Netherlands, Luxembourg; Malta, Portugal, Finland; Ireland, United Kingdom; Romania. All these countries have already been observed in Section 2.5.1.3 as possible outliers. These countries have characteristics that make them different from the rest.

The relative similarities between the countries for DBSCAN (Silhouette scores) show that samples are not segregated from one another as dense regions. For any epsilon (minimum distance to group 2 countries in a country), the Silhouette score is very low, and there is a very high number of outliers. DBSCAN and Kmeans could benefit from dimensionality reduction because a more dense data space will be created.

2.5.2.2 Selection of the baseline model

The last model considered and the one used in the following sections is hierarchical clustering. In Section 2.5.3, there is a more detailed explanation of the model and why it has been selected. The Silhouette score using Kmeans obtained for 5 clusters is 0.23. This is the score to beat. The objective is to have a Silhouette score higher than 0.3 using hierarchical clustering.

In order to do that, reducing multicollinearity is the next step that needs to be done.

2.5.2.3 Multicollinearity reduction

In the model, there are only 11 features that have been selected carefully. Initially, it was tried to delete features manually. However, depending on the path followed, final different results were obtained. This shows that even with high levels of multicollinearity, all features are meaningful for the model. In order to delete features manually without noticing much change, a higher number of features in the model will be required.

A model that uses principal component analysis (PCA) has been selected to eliminate multicollinearity and examine the relationships between variables. The downside of following this path is that old variables are transformed into new principal components, which are harder to interpret.

2.5.2.3.1 Principal component Analysis (PCA)

Combining clustering with dimensionality reduction techniques is a general approach. A very used dimensional reduction technique is PCA which is afterward combined with a clustering technique like Kmeans or hierarchical clustering. This allows for obtaining an optimum cluster solution. Additionally, approximations using two or three PCs are helpful because the data can be summarized in a scatter plot [82]. Different studies combine PCA with hierarchical clustering for extracting different patterns.

Firstly, in [59], a country clustering is done considering economic, technological, cultural, demographic, and quality of life variables to observe if corruption has a real effect on the growth and evolution of countries or if it is just one more factor. Secondly, a paper published in Nature [85] following the same methodology tries to analyze if there are similar epidemic patterns of the SARS-CoV-2 in different regions of Italy. The purpose of the analysis is to provide policymakers with a snapshot of the epidemic in Italy, which might help guide the adoption of countermeasures to the situation at a regional level.

Both papers cluster a tiny dataset like the dataset of this study. In [59], 39 countries and in [85] 21 regions. A wide number of features tend to have high correlations in both papers and this study. Thus, it is necessary to reduce the dimensionality and project features in a non-correlated space to obtain optimal clustering results.

Once the literature has been shown, a theoretical background of principal component analysis is presented. PCA is a way to reduce the number of variables while maintaining most of the vital information. It transforms a number of variables that may be correlated into the same or a smaller number of uncorrelated variables, known as principal components (PCs). The principal components are linear combinations of the original variables weighted by their variances (or eigenvalues) in a particular orthogonal dimension. According to Nature, *PCA reduces data space by geometrically projecting them onto lower orthogonal dimensions called principal components (PCs), with the goal of finding the best summary of the data using a limited number of PCs.* [82].

Using PCA completely deletes multicollinearity, and principal components are ordered following variance criteria. The first PC is chosen to minimize the total distance between the data and their projection onto the PC. By minimizing this distance, it is also maximized the variance of the projected points in the new principal component. Once PC1 has been obtained, the same procedure is done for obtaining the following principal components. Thus, it can be seen that the first principal component (PC1) contains more variance than the second (PC2), PC2 more than PC3, and successively.

The variance ratio of the dataset studied is calculated, which explains how much variance, or in other words, the information each feature contains.

PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**	**PC9**	**PC10**	**PC11**
42,6%	18,7%	10,4%	7,5%	6,9%	4,8%	3,7%	2,5%	1,6%	0,7%	0,4%

Table 2.4: Variance ratio in % of principal components

It is logical to consider deleting features that do not give almost information. PC1 explains 42.6% of the variance of all features while PC11 only 0.4%. Reducing dimensions helps to deal with a smaller hyperspace which can help similarity-based machine learning approaches to obtain better results [125]. A 29-country data set is a very small dataset and papers already mentioned ([59] [85]) with similar data sets that presented high levels of multicollinearity. PCA was used for reduced dimensionality, deleting multicollinearity, and obtaining optimal cluster results. At this point, it is important to define criteria for selecting the optimum amount of principal components to keep.

For a deeper explanation of how the model becomes more meaningful when reducing dimensions that present low variance, consult Chapter of Appendix 7.3. A very detailed analysis of the improvement of the meaning of distances between countries when reducing the dimensions of the model is performed.

Selecting the number of components

After analyzing if the characteristic distance (average distance between countries) is meaningful in the original space (See Chapter of Appendix 7.3), it is confirmed that dimensionality reduction is required in this model.

Selecting an appropriate number of principal components is relevant for the final results. The criteria depend mainly on the additional information obtained by adding an additional principal component. A higher number of dimensions makes the average distance between two random points bigger. Thus, if the variance added to the model by adding a principal component is small, the gain of adding a principal component can be none because data will be more sparse compared to the previous situation.

Different criteria can be defined for selecting the number of components. Here, three criteria, which are the most common, are selected:

- Select the number of components that hold 80-90% of the total variance together.

  If this criteria is used, variance ratio from Table 2.4 is analyzed. Between 5 and 6 principal components would be selected.

- Use a scree plot: A scree plot is a plot of the variance. It is shown in Figure 2.6. The suggested number of components to keep is the last component before the plot forms an elbow and the curve flattens out. Unfortunately, the scree plot often presents some ambiguity. In this case, it does not present a very defined elbow. It suggests keeping between 3 and 4 components.

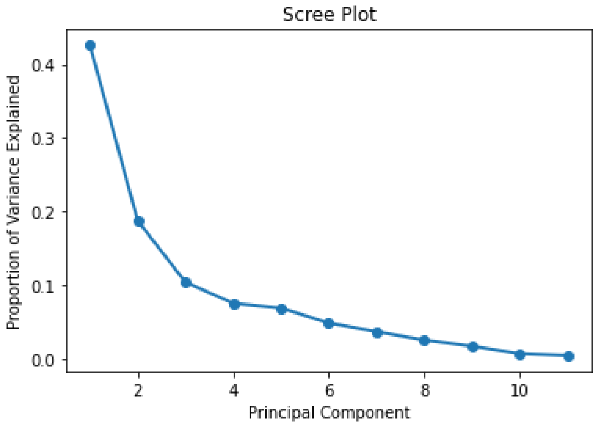


Figure 2.6: Scree plot

The case analyzed does not present a very defined elbow. It suggests keeping between 3 and 4 principal components.

- Use the Kaiser criterion: The Kaiser rule suggests the minimum eigenvalue rule. In this case, the number of principal components to keep equals the number of eigenvalues greater than 1. Eigenvalues of principal components are shown in Table 2.5.

PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**	**PC9**	**PC10**	**PC11**
4,86	2,14	1,18	0,86	0,78	0,55	0,42	0,29	0,19	0,08	0,05

Table 2.5: Variance (eigenvalues) of principal components

The kaiser criterion would suggest keeping 3 components.

It is decided to keep 3 components. The reason is that the scree plot is not very defined. It suggests keeping between 3 and 4 components, and the Kaiser criterion suggests keeping three components.

A total variance of 71.7% is kept. This is under 80%. However, the benefits of having 3 PCs are that data can be observed in a 3D space and that more defined groups would be obtained (As concluded from Table 7.6).

2.5.2.4 Interpretation of the new orthogonal space

In order to analyze the new orthogonal space, loading factors can be obtained. By definition, the principal components are linear combinations of the original variables. Then, loading factors are the weight each original feature contributes to the formation of the principal components.

Before analyzing the loading factors and their meaning, it is important to display how countries are located in this new orthogonal space. From a theoretical point of view, it is important to explain how new technologies penetrate markets and which groups can be classified as adopters of these technologies.

2.5.2.4.1 Using PCA as a visualization tool

PCA is not only a good tool for deleting multicollinearity, but it is also helpful for observing the problem visually. PCA can be used as a visualization tool to observe the data in 2D or 3D. In Figure 2.7, countries are represented in the first two principal components, which account for 61.4% of the total variance of the features.

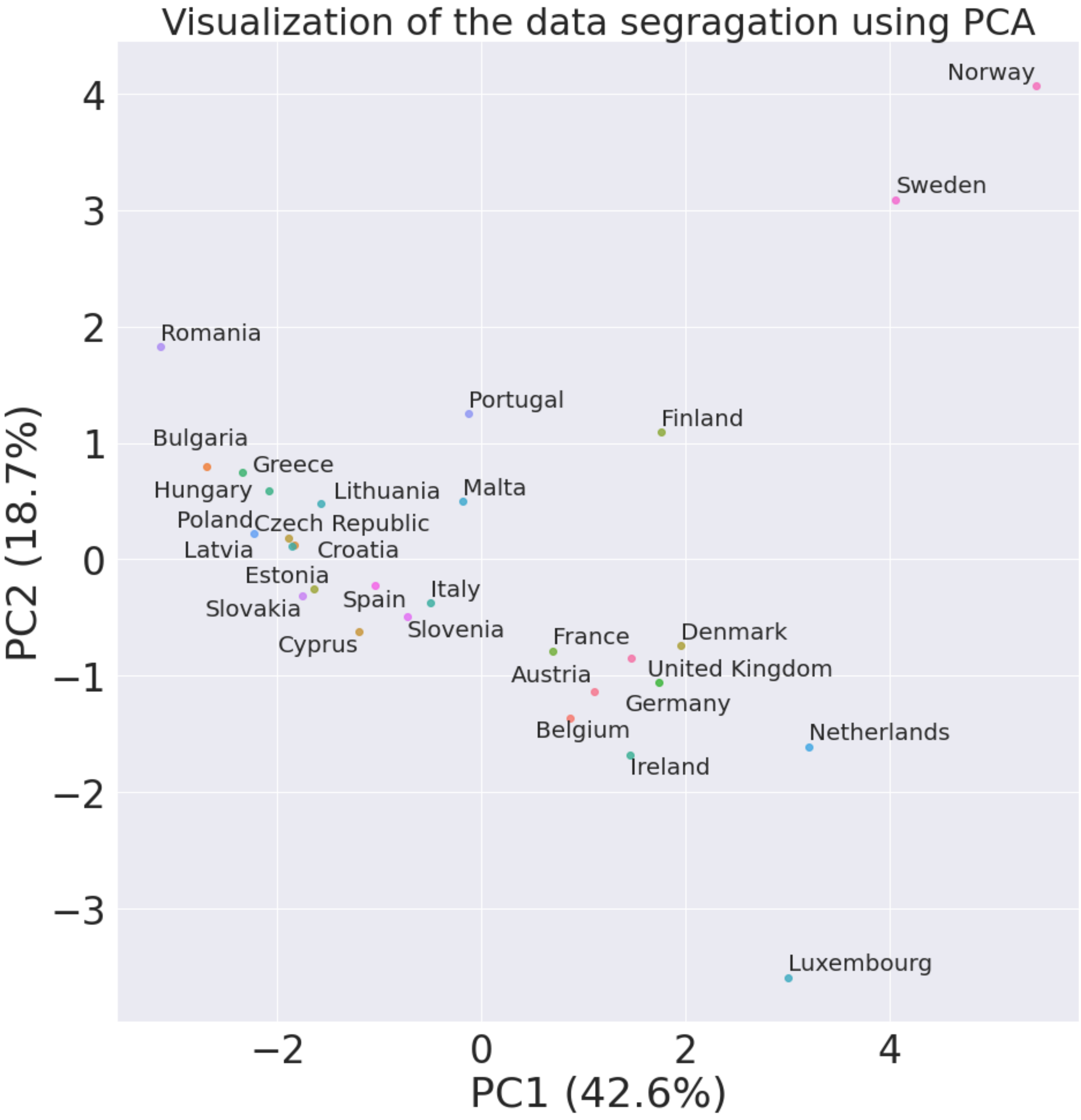


Figure 2.7: Representation of the countries in the first two principal components (PC1 and PC2)

From Figure 2.7 different observations can be done. As it was observed using DBSCAN, Norway and Sweden are outliers because they are not very close to each other, but there are very far away from the rest. Netherlands and Luxembourg present the same situation. Finland is clearly separated from the rest, and Malta and Portugal are relatively far. The United Kingdom and Ireland were outliers in the DBSCAN analysis, but in Figure 3.7 do not look like outliers just observing these two dimensions. (In Notebook, an additional interactive 3D plot can be observed).

Apart from countries relatively far from the rest, others are forming more delimited groups. PC3 would be used as well to separate better clusters. At this point, it is not easy to understand what

PC1 and PC2 represent. This is analyzed in Section 3.5.2.4.3, but before doing that, a theoretical concept related to the penetration of new technologies inside a market and the different existing actors is introduced to understand better what will represent each cluster.

2.5.2.4.2 Innovation adoption curve of Rogers

The clustering purpose is to group countries into 5, 6 groups if possible, following the diffusion of innovations theory presented in 1962 by Rogers [120] and shown in Figure 2.8. The innovation adoption curve of Rogers is a model that classifies adopters of innovations into various categories (innovators, early adopters, early majority, late majority, and laggards) based on the idea that specific individuals are inevitably more open to adaption than others.

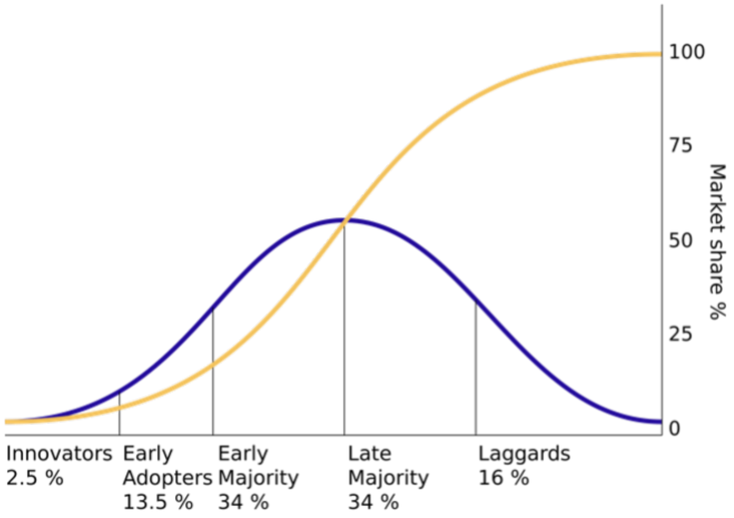


Figure 2.8: The diffusion of innovations described by Rogers [120]

The different features described in 2.4 try to capture how evolved the vehicle market of each country is. A country will be more prone to adopt new technologies depending on the actual market situation, social awareness, and government policies, including investment in infrastructure, subsidies, and tax reductions. A static image of the country has been taken. It considers the technologies already starting to penetrate markets, such as biofuels and electric vehicles. Other alternative fuels like synthetic fuels or hydrogen have not been considered to group countries because they do not have relevance in the current market.

2.5.2.4.3 Loading factors

The loading factors for each of the three principal components are shown in Table 2.6.

It is important to remark that initial variables are standardized, meaning the mean is 0. Then, if GDP per capita is smaller than the mean, it will have a negative value under this scale.

	PC1	**PC2**	**PC3**
GDP_per_capita_2021	0,32	-0,32	-0,07
long_term_interest_rate	-0,23	0,29	0,28
GHG_intensity_reduction_biofuels	0,31	0,42	0,11
ESR_target_2030	-0,39	0,14	-0,07
RES_share_2020	0,20	0,50	0,23
registrations_electric_cars_2020	0,37	0,34	-0,04
charging_points_per_100km	0,30	-0,20	0,10
electric_vehicles_per_charging_point	0,17	0,32	-0,18
average_age_cars	-0,37	0,30	-0,14
social_awareness_ind	0,40	-0,03	0,26
employment_automotive_industry	-0,08	0,14	0,85

Table 2.6: Loading factors of selected features

Once this is clear, the first principal component (PC1), which captures 42.6%, appears to correspond to countries that would be early adopters and the early majority on the right. Higher GDP per capita (0.32) positively influences the alternative vehicle transition to be more evolved. A good public electric charger density infrastructure (0.30) and social awareness (0.40) are also very relevant in the market's current situation. Having high registrations of electric vehicles (0.37) is positive, and the age of cars (-0.37) also, the shorter the average life of cars is, the more right the countries are located. Employment in the automotive industry has a minimum relevance in this axis (-0.08). In summary, higher values in infrastructure, social awareness, CO_2 reduction targets (-0.39), new electric vehicle registrations, and lower values in the average age of cars help the countries to be more on the right of the x-axis. Countries that have worse values in the features are on the left. Sweden and Norway, followed by the Netherlands, are the most to the right countries and Romania the most to the left (See in Figure 2.7).

On the other hand, the second principal component (PC2) has a variance of 20.4%, and it is harder to interpret. It gives a negative value to GDP per capita (-0.37), which means countries like Bulgaria with a negative GDP per capita (in the standardized scale) would be on the upper part of the y-axis. Higher long-term interest rates (0.27) also contribute to be in the upper part. A low ESR share (0.14) and a high employment rate in the automotive industry (0.14) contribute slightly to be in the upper part as well. Having low charging points per 100 km (-0.21) helps to be on the upper part. Having old cars goes in the upper direction (0.31). All these features group countries with lower development to alternative vehicle technologies in the upper part. However, the use of biofuels (0.41), the renewable energy share (0.52), and the registration of new electric cars go in the upper direction(0.34), which are terms that are common for countries with a more decarbonized automotive market. Therefore, afterward, Norway and Sweden are in the upper part. They present very high use of renewable energy sources, high use of biofuels, and high registration of electric cars. They are outliers in these categories. The use of biofuels is high only in Norway, Sweden, and Finland. The rest of the countries have low values (never higher than 1) in this feature, meaning they are not going to the upper part due to this factor. The same happens with renewable energy share, but not that drastically. For the rest of the countries, it can be understood that the upper dimension means to be part of the late majority or laggard, and the down part means to be advanced. The case of Finland is more complex because it has high values, but not enough as Sweden and Norway, so it is kept in-between.

The third principal component, PC3, has a variance of 10.2 % and gives different weighting to each feature. It gives certain importance to the long-term interest rate (0.28) and the renewable energy share (-0.23), which locates countries that tend to be less advanced in the positive part. On the other hand, countries with a dense network of chargers (high charging_points_per_100_km) (0.18) and high availability (low electric_vehicles_per_charging point) (-0.14) will be on a positive axis. High social awareness (0.26) locates countries in the positive axis. Some features push to the upper part more developed countries, and others push them to the downside. However, the most relevant feature in PC3 is the employment in the automotive industry (0.85) which is locating countries with a more prominent automotive industry sector in positive part. The third dimension can be analyzed in a 3d plot.

Once the new reduced orthogonal space determined by the principal components can be understood, the development of the baseline model is started.

2.5.3 Development of the baseline model selected and analysis of the results

Cluster analysis is performed using 3 PCA components. Agglomerative hierarchical cluster analysis is performed, the most widely used technique between hierarchical clustering methods within the literature [63].

Hierarchical clustering has been selected because it does not require defining the number of clusters at the beginning. Here, a bottom-up approach will be followed. This is agglomerative and consists of grouping just the instances that are closest to each other. It is like a reversed tree.

A distance matrix needs to be constructed for clustering, and there is a distance calculation. The distance is defined by the linkage, the distance defined between clusters (method), and the similarity measurement, which is the pairwise distance between observations in n-dimensional space (metric).

In cluster analysis, it is not trivial to determine the most appropriate clustering linkage method and the most suitable similarity measurement. Applying a different method to the same data produces different clusters. The same happens when applying a different metric.

In order to evaluate which is the most appropriate method, a metric for evaluating the clustering results has been selected. Silhouette score has been selected as the most appropriate criterion [121].

2.5.3.1 Silhouette score

Silhouette score is a metric used for clustering when there are no ground-truth cluster labels that can be defined. This is the most typical case while doing clustering. If not, Rand Index or Mutual Information could be used.

Silhouette score is used to measure the separation distance between clusters. It is used to see how densely distributed the data is. The mean intra-cluster distance (Mean distance between the observation and all other data points in the same cluster) with respect to the mean nearest-cluster distance (Mean distance between the observation and all other data points of the next nearest cluster) is calculated. Silhouette coefficient can be expressed as:

$$\frac{(n-i)}{max(i,n)} \tag{2.2}$$

Where n is the distance between each sample and the nearest cluster that the sample is not a part of, while i is the mean distance within each cluster.

It goes between -1 and 1, where 1 means all data is packed in the center of the cluster, and -1 that data may have been assigned to the wrong cluster. It can be used as a graphical tool where each sample's score is observed and how dense and well delimited a cluster is. It can be obtained a global Silhouette score by averaging the Silhouette coefficients. The global Silhouette score describes the entire clustering performance with a single value. Figures of what a Silhouette plot looks like and further analysis can be consulted in Appendix 7.4.1.

Once the Silhouette score has been introduced, linkage and similarity measurement methods can be evaluated, and choose the most appropriate. The highest Silhouette score gives the best solution, which allows checking the model's internal validation.

Apart from observing the Silhouette score, the Silhouette plot and dendrogram can be used to make the final decision.

2.5.3.2 Hyperparameter tuning and model selection

Once the similarity score has been selected and explained, linkage and similarity measurements are presented. There is an extensive list, and the most common ones have been selected.

Linkage methods

- Single link: It measures the distances between closest elements from different clusters.
- Complete link: It is the distance between the furthest elements in clusters. It is the opposite of a single link.
- Average link: Average of all pairwise distances, and the advantage is that the cluster formation is less affected by outliers.

From now on, the following linkage methods presented can only be combined with Euclidean distance (the most common similarity measurement used) because these methods compute centroids in Euclidean space.

- Weighted link: Is a modification of the average link
- Centroids: It measures the distances of the geometric centroids of two clusters
- Median: It is a modification from centroid linkage
- Ward: It measures how the total distance from centroids changes when considering joining two clusters. It uses a variance minimization algorithm.

Similarity measurements

- Euclidean distance: uses Euclidean distance (2-norm) as the distance metric between the points. It is the most typically used.
- Minkowski distance
- Manhattan distance
- Standardized euclidean distance
- Squared euclidean distance

The possible combinations have been evaluated to obtain the highest Silhouette score for 6 clusters. A summary table for each linkage with the similarity measurement, which gives the highest score, is presented. When it was impossible to obtain 6 clusters or the score obtained was extremely low, the result obtained for 5 clusters is shown and indicated with brackets.

linkage (method)	**similarity method (metric)**	**silhouette score**	**Dendrograms (See in Appendix 7.4.2)**
single (5 clusters)	cityblock	0,251	7.9
complete	cityblock	0,398	7.10
average	euclidean	0,375	7.11
weighted	euclidean	0,375	7.12
centroid (5 clusters)	euclidean	0,372	7.13
median	euclidean	0,256	7.14
ward	euclidean	0,401	7.15

Table 2.7: Best similarity score obtained for different linkage methods

It is important to plot the dendrograms to analyze if the formed clusters make sense. The similarity score can be very high because one cluster contains all countries, and the rest contain just one country. This grouping is discarded. The different dendrograms can be seen in Appendix 7.4.2.

Taking into account Table 2.7 and Appendix 7.4.2 the following conclusions can be extracted:

- Single link is discarded because it has a lower Silhouette Score than other methods. It has a value of 0.25 for 5 clusters, and the clusters created do not make any sense.
- Complete link: It presents a relatively high Silhouette Score for 6 clusters, 0.398. The grouping of countries is reasonable. The threshold between 5,6, and 7 clusters is very thin. For 5 and 7 clusters, the Silhouette Score is lower.

- Average link: The highest value obtained using Euclidean distance is 0.375. It separates Sweden and Norway as independent clusters, and the rest of the clusters are bigger.
- Weighted: The highest value is obtained again using Euclidean distance, 0.375. The groups of clusters follow a similar distribution to the complete link, but if countries want to be clustered following Roger's curve, they need to be separated into more groups, 9. In this case, Silhouette Score is 0.29. This approach is discarded due to the low score obtained.
- Centroid: It is calculated with Euclidean distance. The results for the 5 clusters have a score of 0.372, and the clusters are not representative. There are two clusters which are single countries. Again, 8-9 clusters would be required, and the score obtained is very low.
- Median: It is calculated with Euclidean distance. A very low score is obtained, 0.256.
- Ward: It is calculated with Euclidean distance. It presents the highest Silhouette Score, 0.401. Grouping is very reasonable.

Ward method is selected because it presents the highest Silhouette Score, and the grouping looks reasonable. 6 clusters present a score of 0.401, and 7 clusters have a score of 0.398. In both cases, there is only 1 misclassification in the Silhouette Plot. The significance of the results justifies selecting 7 clusters (See Section 2.5.3.3).

The Silhouette Plot for 7 clusters (See Figure 2.9) has only 1 country misclassification. It is Cyprus, representing less than 0.1% of the new vehicle sales in the EU.

When analyzing the 3D plot, Finland might be the country that is being misclassified because it is far away from Malta, Portugal, Spain, Italy, and Slovenia, and it is almost at the same distance from Germany, France, and other clusters. Section 2.5.3.4 shows that one cluster is defined as the early majority and the second as the late majority. Grouping Finland with Spain and Italy group would mean estimating a slower evolvement of its vehicle fleet, which is a conservative decision. Additionally, Finland's vehicle fleet only represents 1% of the European fleet.

The final hierarchical clustering model has the following features:

- **method: ward**
- **similarity measurement: euclidean**
- **number of clusters: 7**
- **Silhouette score: 0.398**

The results obtained are satisfactory. The objectives of dimensionality reduction have been achieved. The purpose was to have a model with a Silhouette score higher than 0.3. All clusters in Figure 2.9 have individual scores over 0.3, except two which have a score around 0.3. There is 1 misclassification. The score using hierarchical clustering has been improved from 0.23 (See previous Section 2.5.2.1) to 0.40, which is quite remarkable.

The final dendrogram obtained is shown, and each cluster is marked with a different color in Figure 2.10.

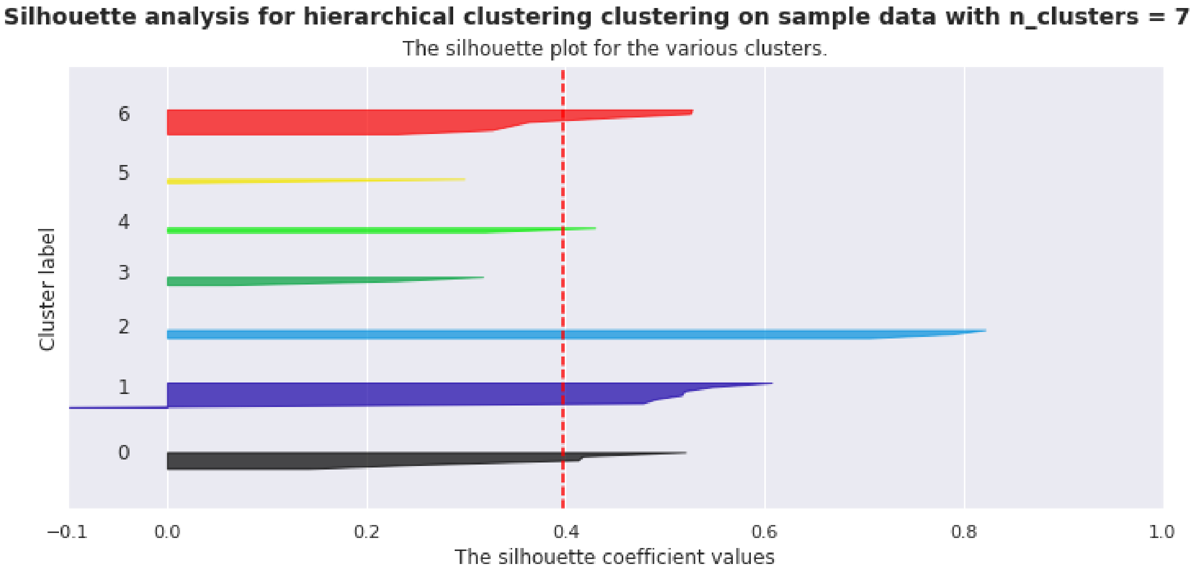


Figure 2.9: Silhouette plot for the final model selected

Stability of data-driven process

Once results are obtained, the stability of the data-driven process is analyzed; what results when deleting certain features, such as new electric vehicle registrations or environmental awareness? A more detailed analysis can be consulted in Appendix 7.4.3.

Deleting these features shows that if a model is created with enough descriptive features, the level of multicollinearity will be high. When deleting the features mentioned, almost identical results to the model with all features are obtained. Then, if there is a lack of data for a specific feature, it should not be a problem not to use it or search for a different one. Extending results to an international level should be possible. Deleting features from a possible international model or selecting others similar should not be a problem, and it is still possible to obtain meaningful results.

2.5.3.3 Significance of results

The hierarchical clustering can allow the grouping of countries and give them labels following an analogy to the Rogers curve, considering how evolved each national vehicle fleet market is. Before doing that, the ANOVA analysis determines if there are statistically significant differences for given average values of identified clusters.

In the one-way analysis of variance tests, under the null hypothesis, it is assumed that all sample means are equal. In contrast, under the alternative hypothesis, it is assumed that there is at least one sample mean which is statistically significantly different than the others. Under the null hypothesis here, it is assumed that all cluster means are equal, observing each principal component separately.

According to the results of the one-way analysis of variance, at a significance level of 5%, it can be concluded that there is a statistically significant difference in means between clusters of

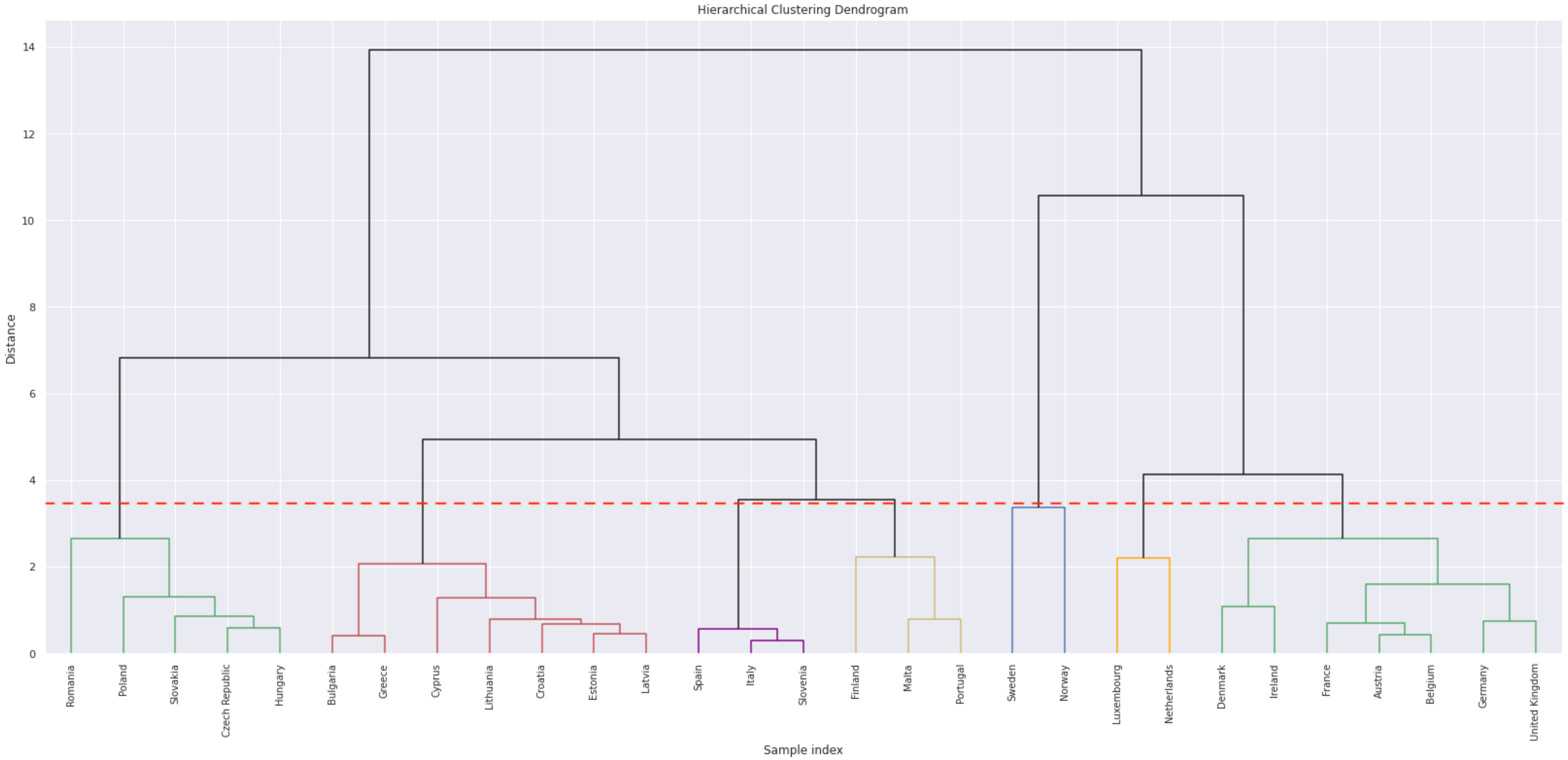


Figure 2.10: Dendrogram of the final model selected: Dendrogram of Hierarchical Clustering based on Ward's criterion. The height of the branches indicates the dissimilarity between clusters. The dendrogram was partitioned.

cluster	1	2	3	4	5	6	7	F_ratio	p_value
PC1	-2,22	-1,87	-0,76	0,48	4,75	3,11	1,33	57,52	$1{,}17\cdot10^{-9}$
PC2	0,50	0,20	-0,36	0,95	3,58	-2,60	-1,09	18,73	$2{,}53\text{E}\cdot10^{-6}$
PC3	1,51	-1,14	0,20	-0,77	0,31	0,33	0,13	11,43	$8{,}03\cdot10^{-5}$

Table 2.8: One-way analysis of variance based on clusters means of principal components

all principal components. At a significance level of 1%, there will still be statistical significance differences for all principal components.

The one-way analysis of variance cannot show which means are different. Consequently, the clusters cannot be ranked according to the features contained in PC levels. Because of that, additionally, 95% confidence intervals for means for each cluster and observed principal components are calculated. The intervals can be observed in Appendix 7.4.4 if 6 clusters are selected and the intervals for the final model which uses 7 clusters can be seen here in Table 2.9.

	Cluster 1		**Cluster 2**		**Cluster 3**		**Cluster 4**		**Cluster 5**		**Cluster 6**		**Cluster 7**	
	lower limit	**higher limit**	**lower limit**	**higher limit**	**lower limit**	**higher limit**	**lower limit**	**higher limit**	**lower limit**	**higher limit**	**lower limit**	**higher limit**	**lower limit**	**higher limit**
PC1	-3,31	-1,12	-2,87	-0,88	-1,31	-0,21	-1,72	2,69	2,81	6,69	2,81	3,41	0,41	2,25
PC2	-1,11	2,11	-0,84	1,24	-0,63	-0,10	0,16	1,74	2,20	4,96	-5,40	0,2	-1,77	-0,4
PC3	0,47	2,55	-1,76	-0,52	0,05	0,35	-0,96	-0,58	-3,90	4,41	-1,01	1,67	-1,50	1,75

Table 2.9: Confidence intervals for 3 principal components and 7 clusters, confidence level 95 %

The significance of the results for the 7 clusters is very satisfactory. It can be concluded that all clusters can be well separated by at least one principal component. Then, interpreting its results and how evolved their markets are is pretty straightforward.

2.5.3.4 Interpretation of results

Once the results of the cluster analysis have been obtained, validated using the Silhouette score, and analyzed if there are statistically significant, results will be interpreted. After explaining the Rogers curve in Section 2.5.2.4.2, an analogy is used for labelling the different countries depending on how evolved the national vehicle market is in each cluster. It is important to know which market share corresponds to each cluster. In Table 2.10, it is shown which countries represent each cluster, the new passenger car sales share of each country in the EU, and each cluster.

Analyzing the means of each cluster (See Table 2.9) and its confidence intervals (See Figure 2.11, 7.20, 7.21) and considering the analysis of the loading factors from the different principal components done in section 2.5.2.4.3 different conclusions can be extracted.

Considering PC1, clusters 5 and 6 have the highest means (See Table 2.8). Then, cluster 7 is the next with the highest mean, but its confidence interval is separated from clusters 5 and 6.

Cluster	Countries	New passenger car sales share by country	New passenger car sales share by cluster
1	**Czechia**	2,1%	**9,9%**
	Hungary	1,2%	
	Poland	4,5%	
	Romania	1,2%	
	Slovakia	0,8%	
2	**Croatia**	0,5%	**2,5%**
	Cyprus	0,1%	
	Greece	1,0%	
	Estonia	0,2%	
	Bulgaria	0,2%	
	Latvia	0,1%	
	Lithuania	0,3%	
3	**Spain**	8,7%	**23,9%**
	Italy	14,8%	
	Slovenia	0,4%	
4	**Malta**	0,0%	**2,5%**
	Portugal	1,5%	
	Finland	1,0%	
5	**Sweden**	3,1%	**4,8%**
	Norway	1,8%	
6	**Netherlands**	3,3%	**3,7%**
	Luxembourg	0,4%	
7	**France**	16,8%	**52,7%**
	Germany	26,6%	
	Ireland	1,1%	
	United Kingdom	-	
	Austria	2,4%	
	Belgium	3,9%	
	Denmark	1,9%	

Table 2.10: Clustering results, market share by country and market share by cluster in 2021. United Kingdom would represent 14,4% of total EU vehicle fleet share

Cluster 5 has a higher mean of PC1 than cluster 6, but confidence intervals cross slightly. Considering PC2 is more complex to interpret but shows that being at the very top meant in the case of cluster 5 that the use of biofuels, the renewable energy share, and the registration of new electric cars is very high. Cluster 5 is classified as the innovator (4.8%). Cluster 6 is classified as the second innovator group or early adopter, but it has as well a tiny market share (3.7%). Then, it is classified as the second innovator.

These two groups of countries formed by only four countries in total are leading the transition

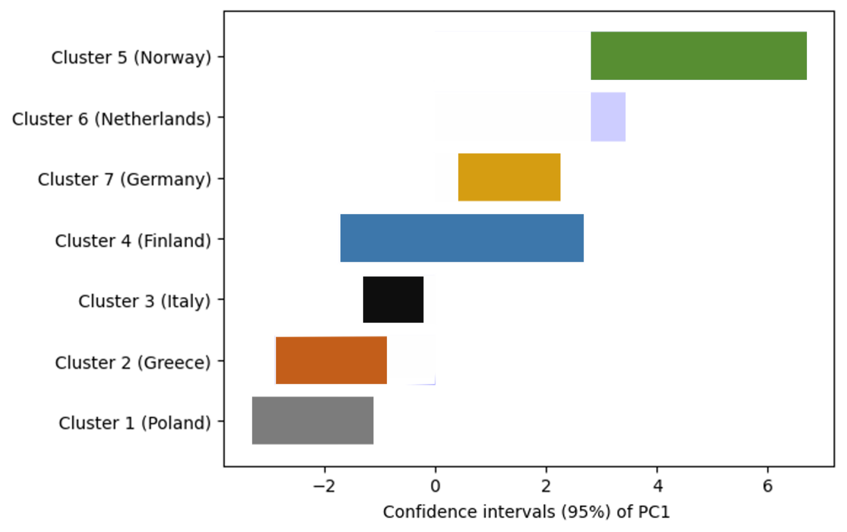


Figure 2.11: Graphical representation of confidence intervals of PC1. A country of each cluster is mentioned to identify each cluster more quicker.

to alternative fuel vehicles or have the conditions to lead but are not leading at the moment. For example, Luxembourg has an intake of electric vehicles of 11%, which is far away from the rest of the countries of these 2 clusters, but it has a very low average vehicle age, making this market very dynamic. Apart from the analysis done using loading factors, if original features are considered, these four countries are the four leading countries in most features. Norway is ranked in the top 3 in 6 out of 11 features, Sweden in 4, Luxembourg in 5, and Netherlands in 3. (If rankings want to be consulted see Appendix 7.5).

The next highest mean in PC1 is from cluster 7. Its PC1 confidence interval is separated from cluster 3 but not from cluster 4. The PC2 confidence interval of cluster 7 is separated from cluster 4. In the case of PC2, it is harder to interpret. Cluster 4 is in the upper part because they have a lower GDP per capita and ESR target but a higher RES share. Considering PC1 and PC2, cluster 7 would be ranked as having a more developed automotive fleet market than cluster 4. Cluster 7 has Europe's biggest countries by population and economic terms. It is the most significant cluster by market size, having France, Germany, and the United Kingdom representing more than 50% of the market share. This group is classified as the early majority (52.7%).

Then, cluster 4 has the next highest mean. Cluster 4 is a small market (3.3%). These countries are relatively small in terms of population and GDP per capita. They have a relatively high use of renewable energy, electric cars sold in 2020 were slightly higher in these countries (in relative terms, not in absolute value) than in cluster 5, and the use of biofuels is more prominent than in cluster 6. However, the targets for reducing emissions and social awareness are considerably lower than in cluster 5. This cluster is below cluster 7 and above cluster 3. Cluster 3 corresponds to a much bigger market where the intake of electric vehicles, the infrastructure, the usage of biofuels, and social awareness are lower. The PC2's confidence interval of both countries is separated. Cluster 4 is cataloged as advanced late majority (2.5%) and cluster 3 as the late majority (23.9%)

Finally, clusters 1 and 2 are labelled. Till now, 87.6% of the market share has been cataloged. Observing PC1, both clusters are the less developed ones. They have a similar PC1 mean and confidence interval. The main difference between these two groups is the power of the automotive industry. PC3 gives considerable relevance to the automotive industry, and cluster 1 and cluster 2 confidence intervals differ in PC3. In cluster 1, employment in the automotive industry is very relevant. Thus, there will be opposition against the intake of electric vehicles because this reduces jobs. Cluster 2 can be cataloged as laggards and Cluster 1 as laggards with the automotive industry or "ICE laggards".

As it has been done before for the highest ranked countries, the original features have been ranked to show which are the countries in the top 3 worst positions for each feature (See rankings in Appendix 7.5). Greece, Romania, Latvia, Lithuania, and Bulgaria are ranked in the top 3, 3 or more times. Romania is part of cluster 1, and Greece, Bulgaria, Latvia, and Lithuania are part of cluster 2.

It has been seen that confidence intervals have significance and can separate the different groups by at least one principal component.

Finally, the labelling selected is summarized here:

- **Innovators**: Cluster 5
- **Second Innovators**: Cluster 6
- **Early Majority**: Cluster 7
- **Advanced late majority**: Cluster 4
- **Late majority**: Cluster 3
- **Laggards**: Cluster 2
- **ICE Laggards**: Cluster 1

In order to observe more visually the different clusters defined in Table 2.10, a color map is presented (See Figure 2.12).

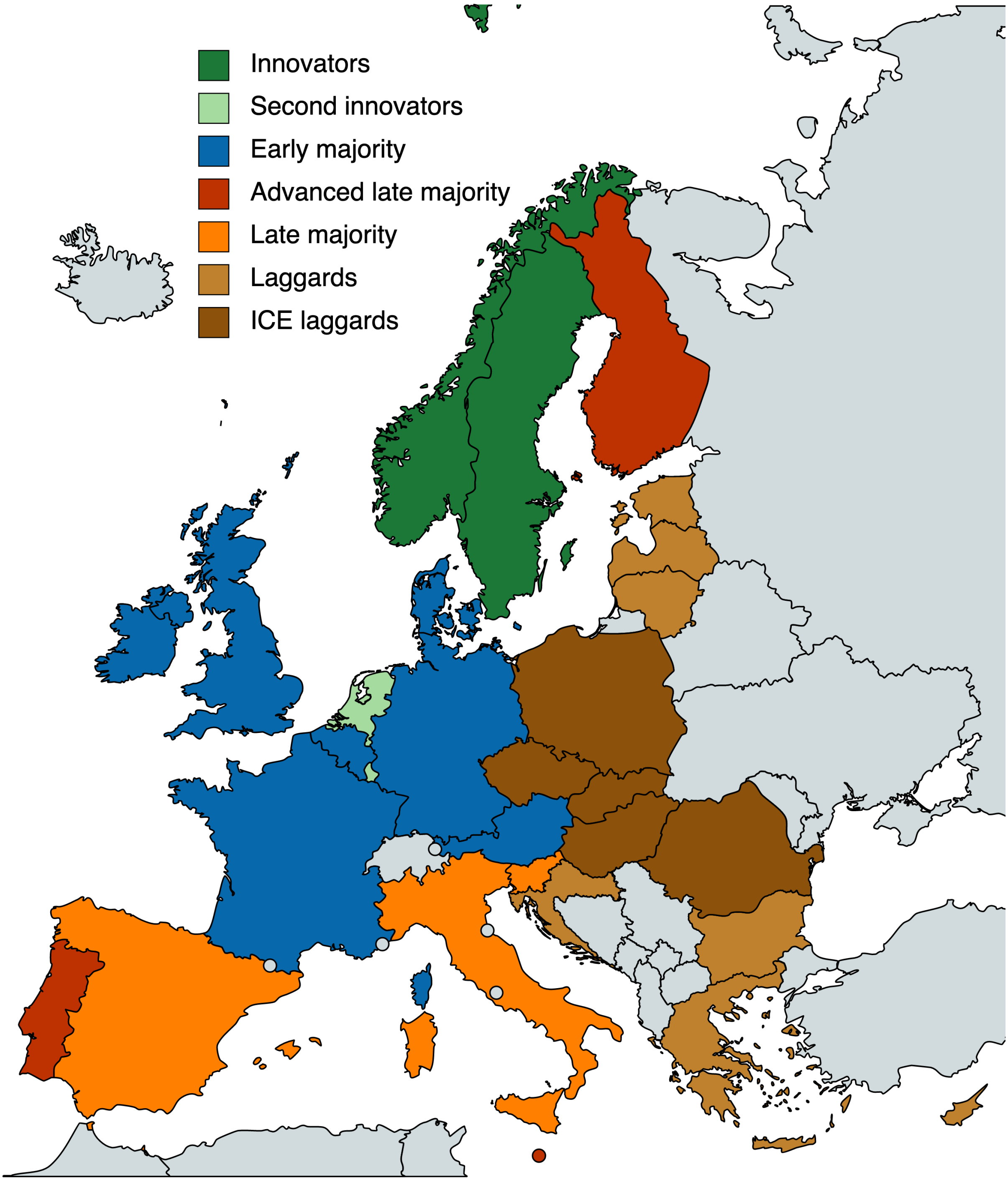


Figure 2.12: Map of Europe with the clusters defined by similarities of the national vehicle market.

2.5.3.5 Comparison with electric vehicle market study

After performing the analysis, it has been found an analysis of BloombergNEF [64] in collaboration with Transport & Environment, where a grouping is done for the different European electric vehicle markets.

The results are slightly different but similar to the ones obtained here (Bloomberg map results can be consulted in Appendix 7.6). In order to understand, it is important to describe specific characteristics of the study.

It is unknown which is methodology followed in order to do this grouping. Then, the scientific rigor of the Bloomberg study cannot be validated. Due to the lack of studies in this field published in scientific magazines, this source is used to compare the results obtained.

Once this has been clarified, the main difference between the study developed here, and the study developed by Bloomberg is that this study is focused on the development of the automotive market of the different countries, considering not only electric vehicles but other alternative powertrains available. This is one of the reasons for the different grouping between both studies.

Additionally, the Bloomberg study uses the new electric vehicle registration share as the main metric in 2020, while this study gives the same relevance to 11 features selected for defining automotive market dynamics.

Both studies use GDP per capita, phase-out targets (in this analysis, precisely ESR target), and charging infrastructure.

In the Bloomberg study, supporting policies have been used as a metric; in this study, it has not been used because this is very detailed data that is difficult to quantify. It will be added in Section 4 where a modelling of the development of vehicle markets at a national and regional level for certain countries selected from each cluster is done.

In this study, a social awareness indicator has been used because it has been found in the literature that it is a relevant parameter when customers buy, for example, a new electric car. The employment share of the automotive industry is also used because it acts against the transition of the vehicle market. Fewer employers are required to produce an electric car than an ICE car.

The last difference relies on the importance the Bloomberg study gives to the size of the country. It includes population, fleet size, and total sales. This makes that bigger countries are grouped instead of grouping countries by their vehicle market characteristics. None of these parameters has been used in this study.

Results comparison

The first difference relies on the innovator group. According to Bloomberg, Norway and Netherlands are innovators, and Sweden, Finland, Denmark, and Iceland are early adopters. They remark on solid support mechanisms, high adoption shares, and charging infrastructure development.

In this analysis, the first innovators are Norway and Sweden. They are different from the Netherlands mainly because of the high use of biofuels, the high renewable share, and the electric vehicles per charging point. These characteristics define them as different markets. None of the studies should be wrong. This study aims to group countries by developing vehicle markets to reach the zero-emission target. In contrast, the Bloomberg study aims to group them by the electric vehicle market intake and the potential for growth of these markets.

The Netherlands is cataloged as the second innovator because it has a very high intake of electric vehicles in 2020, 25%. It is lower than Sweden, 33%, and Norway, 75%. It has the densest infrastructure of chargers in Europe and a relatively low number of electric vehicles per charging point. The Netherlands is grouped with Luxembourg, a tiny country that could be cataloged as the early majority because the electric intake is lower, 11%. However, it has a very high ESR target, a dense charging infrastructure, a lower amount of electric vehicles per charging point than the early majority, and a dynamic market. The age of cars is only 6.7 years, making them able to do a quicker transition of the total fleet than in other countries.

After this group, the early majority for both studies, except Luxembourg, already explained, Finland, afterward explained, and Denmark. In Bloomberg, Denmark is cataloged as an innovator. Denmark has a 16% of new registrations of electric vehicles in front of Germany, which has a 14%. It has a less developed charging infrastructure than Germany, a similar social awareness indicator, and ESR target. After observing the different features, it is much more logical that Denmark is grouped with the early majority. From a qualitative point of view, it can be concluded that the model used for clustering groups correctly classifies Denmark.

In the study done, Malta, Portugal, and Finland are grouped and defined as the advanced late majority. In the hyperspace, Finland is relatively separated from Malta and Portugal, and this group's confidence intervals are spread (See Table 2.9). In Bloomberg Study, Portugal is highlighted for indicating that it is ahead of the other countries of the late majority. It has considerably high new electric vehicle registrations in 2020, 14%, higher use of biofuels, a denser infrastructure, and a higher RES share. All this makes Portugal above them. Malta has an 8% registration of EVs, which is considerably above the late majority average, and it is more than the country with a higher intake, Spain, with a 5%. It has as well a considerably higher social awareness. Both countries are well classified.

Finland is more problematic because it has a GDP per capita similar to the early majority group. It has a higher intensity of biofuels, a higher intake of electric cars with 19%, and a higher RES share than the early majority but considerably lower than Norway and Sweden. The average age of the cars is relatively high and higher than the early majority meaning that the change of the vehicle fleet into electric could be slow. After doing this analysis, it can be concluded that Finland presents different characteristics from Portugal and Malta. It is not clear if its characteristics are above the early majority or below it, but it is clear it is above Portugal and Malta and should not be in the same cluster. Finland could be studied separately from Malta and Portugal to be rigorous and have higher accuracy in the predictions. However, the market share of this cluster is tiny (3.3%), and an additional separation will not be done.

Once the advanced late majority group has been analyzed, the late majority group appears. Spain, Italy and Slovenia define this group. Italy and Spain define this group for Bloomberg. However, the analysis shows how close Italy and Slovenia's automotive markets are. The late majority characteristics are a much lower GDP per capita than for the early majority, lower targets, new registration of electric cars lower than 5%, which is considerably lower than in the early majority (above 10%), reduced infrastructure, and a lower environmental awareness.

Finally, the last group is defined as catching up in Bloomberg, and in this study, there are two groups; laggards and ICE laggards. Laggard's characteristics are the lowest GDP per capita in the EU, lowest targets, registrations of electric cars lower than 3% on average, and in many countries lower than 1%. Additionally, the population sees many other challenges more relevant than climate change, and climate change awareness is very low.

The study presented separates laggard countries with the automotive industry and without the automotive industry. Employment in the automotive industry is higher than 10% in 4 out of 5 countries in ICE laggards. This will make the population and policies reluctant to change to electric vehicles.

2.6 Conclusions and further research

It is the first analysis found in scientific literature to group countries regarding their current automotive fleet situation. In order to analyze if clusters vary in time and verify if progress in the automotive fleet is constant inside clusters, it could be helpful to do the same analysis every two years. This analysis has used mainly data from the year 2020 because it was the one that was available in 2022. An analysis in different moments would allow analyzing how stable these groups are in the future. Stable clusters will allow determining specific supply-side and demand-side policies more efficiently, considering the effects of policies already used in countries of the same cluster.

A second approach could be using variables that are considered in relative terms. For example, the increase of renewable energy share in the last 5 or 10 years for the different countries. This approach has been studied, but much less defined groups, more unstable results, and less defined Silhouette scores were obtained. This approach was discarded.

The clustering model is data-driven analysis. The 11 features selected seem to have been a good approach. A more extensive analysis of the data selected could be done if the resources are larger. In this analysis, employment in the automotive industry was included after receiving feedback from experts in the automotive industry. For selecting more features, additional work with experts in the automotive market could be done, or specific polls, for example, the ones they were done in the case of clustering countries by their construction markets in [102]. This could serve as an example.

Regarding the methodology followed, initially, there was a high multicollinearity among the features. Principal components analysis deleted multicollinearity and reduced the problem into only 3 orthogonal dimensions making a better similarity-based grouping and improving space interpretation. It could happen that multicollinearity and data space is not a problem with other features. Then, it might not require dimensionality reduction techniques as a pre-processing tool. In that case, the analysis would be done directly over the original features. In this case, as it has been said, dimensionality reduction was necessary to obtain a more meaningful space and have more defined clusters. The negative side is that principal component are no longer directly understandable. Then, to understand the results, loading factors have been analyzed and have allowed classifying countries in terms of how evolved their markets are. Finally, clusters have been labelled using an analogy to the Rogers curve.

The clustering model selected has been hierarchical clustering. The tuning of the model has been done by selecting the best linkage and similarity measurement. The results were compared using the Silhouette score, Silhouette plot, and observing the dendrograms. A final model of 7 clusters was selected, and the Silhouette score (0.4) and Silhouette plot obtained were satisfactory. There was only one misclassification, Cyprus. The rest of the countries of the Silhouette plot values were high and uniform. Cyprus represents 0.1% of the EU automotive market's new sales.

After the validation, a one-way analysis of variance based on cluster means was performed and was satisfactory. Additionally, 95% confidence intervals of the means of the principal components were studied. At least, PC1, PC2, or PC3 defined separated 95% confidence intervals for all clusters. If there were a pair of cluster confidence intervals that cross for all principal components, it would not necessarily diminish the significance of the findings. This is because the cluster can be well separated and validated in the hyperspace using the Silhouette score.

The findings allow simplifying the work in further steps from the thesis. A characteristic group for each of the clusters will be selected. Results for the whole EU market sales can now be obtained by selecting one country of each cluster.

3 Country selection and model description

3.1 Characteristic country selection

After grouping national vehicle markets by similarities, selected countries' demand-side factors and external factors are studied. Future new vehicle sales scenarios up to 2035 will be analyzed using Vector21 for one characteristic country of each cluster. Countries of each cluster are selected mainly by their market size (Market share is calculated considering EU-27 countries and Norway). The characteristic country selection can be seen in Table 3.1:

Cluster	Characteristic country selected	Market share
1	Poland	4,5%
2	Greece	1%
3	Italy	14,8%
4	Finland	1,0%
5	Norway	1,8%
6	Netherlands	3,3%
7	France	16,8%
	Germany	26,6%

Table 3.1: Characteristic country selection and new sales market share

Poland has been selected because it is the biggest market in cluster 1. Greece and Italy have been selected for the same reason as clusters 2 and 3. For cluster 4, Finland is selected because it has a simple taxation system. For cluster 5, Norway is selected because of its leading role in EV intake in Europe. For cluster 6, the Netherlands is selected because it has a leading role and is the biggest market in its group. For cluster 7, 2 countries have been selected, France and Germany. This is the most significant cluster in terms of market share representing more than 50% of the total share.

The 7 EU countries selected and Norway represent a market share of 69.8% respect 27-EU countries and Norway. This is already a considerable share of the vehicle market, and it is assumed that countries of the same cluster will have a similar market development. Then, conclusions can be extracted at an EU level and for Norway and UK inside the clustering model.

3.1.1 Regional level analysis

The prediction of the selected European countries is made at a regional level and a lower level. This is done because national countries alone cannot reveal the complete and sometimes complex picture of what is happening at a more detailed level within the European Union. Some regions present very different characteristics within a country, for example, disposable income or car taxation schemes may differ from region to region.

Additionally, a subnational analysis helps analyze changing patterns and learn from the impact that policy decisions can have on the development of the car market.

All the analysis is done using NUTS classification [15], a model developed by European Union to define territorial units in EU-27 countries and UK.

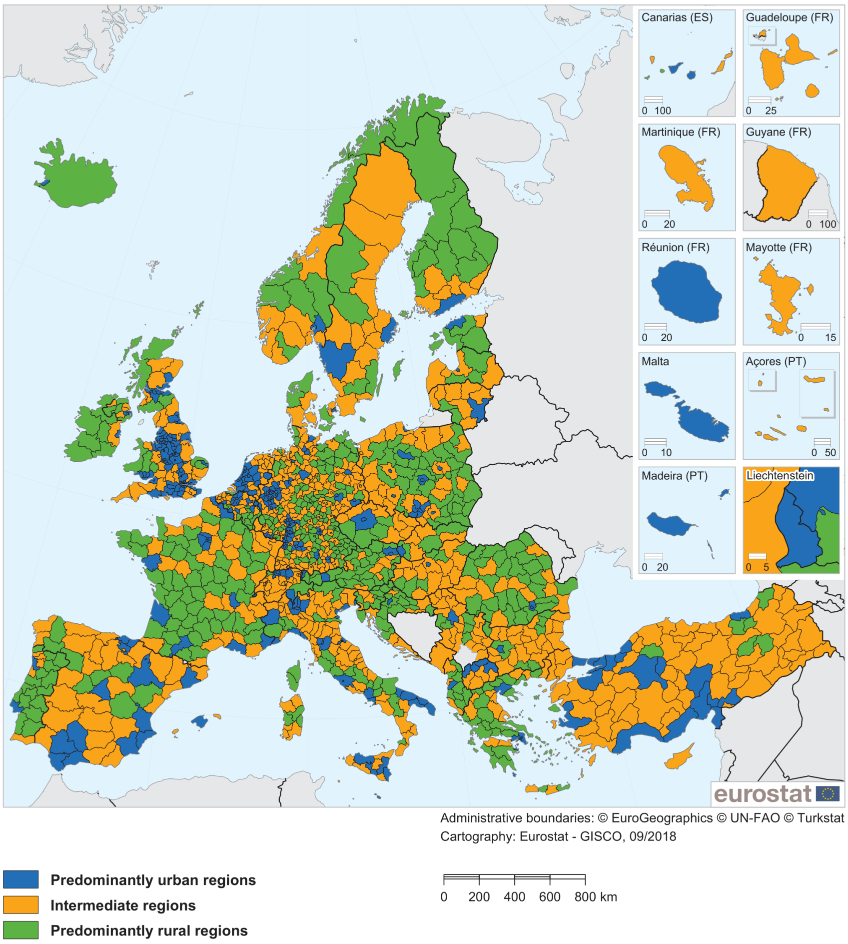
NUTS is a hierarchical system that allows socio-economic analysis at different levels. The classification presented in Eurostat defines the following territorial unit systems:

- NUTS 0: Countries
- NUTS 1: Major socio-economic regions
- NUTS 2: Basic regions for the application of regional policies
- NUTS 3: Small regions for specific diagnoses

The important levels are NUTS 0, NUTS 2, and NUTS 3. NUTS 2 is relevant because taxation schemes and mean socio-economic factors as disposable income varies from region to region. On the other hand, NUTS3 does an additional classification of NUTS2 regions between metropolitan, urban, and rural areas. [99] (See Figure 3.1). The importance of NUTS3 relies on further developing the electrification of transport in rural areas rather than in urban areas. Rural areas present longer average distances and lower average disposable incomes, and bans and regulations are less strict for vehicles because pollution concentration is smaller than in metropolitan and urban areas. All these factors will influence the development of rural areas. There might be significantly different situations in 2030 in the metropolitan and rural passenger car fleets.

Finally, NUTS has been selected because all EU countries use it. It simplifies gathering data and makes possible the task of comparing results among countries using the same territorial unit classification methodology.

Once the selected countries and the level of detail of the data used (when possible) has been defined, the structure of the model used in Vector21 is briefly described.



Administrative boundaries: © EuroGeographics © UN-FAO © Turkstat
Cartography: Eurostat - GISCO, 09/2018

Figure 3.1: NUTS3 classification. Urban-rural typology. Extracted from [98]

3.2 Model description (Redelbach approach [114])

Vector21 is used for modelling the new passenger car sales up to 2035 in the EU. Vector21 is a hybrid of an agent-based and discrete choice market penetration model that assesses the competition among different powertrain alternatives in passenger car markets.

Vector21 models customer choice behavior for conventional and advanced powertrain technologies by combining vehicle simulations, scenario techniques, and discrete choice analyses. Studies on the penetration of alternative penetration are grouped into two major fields:

The first approach focuses on the projected cost of the current and new propulsion systems. It is based on the costumer's total cost of ownership (TCO) minimization [90]. The cost center approach is criticized for assumming an unrealistic *homo economicus*.

The second approach analyzes a dynamic utility-based market model [114]. It uses utility maximization, and the agents consider different criteria (not only total costs) when selecting a vehicle. In contrast to TCO, this approach maps the observed diversity of user characteristics more realistically. Different agents representing real customers are given different weights to each criterion because of socioeconomic characteristics.

The second approach is the one used in this thesis. In Figure 3.2, an overview of the model structure can be seen.

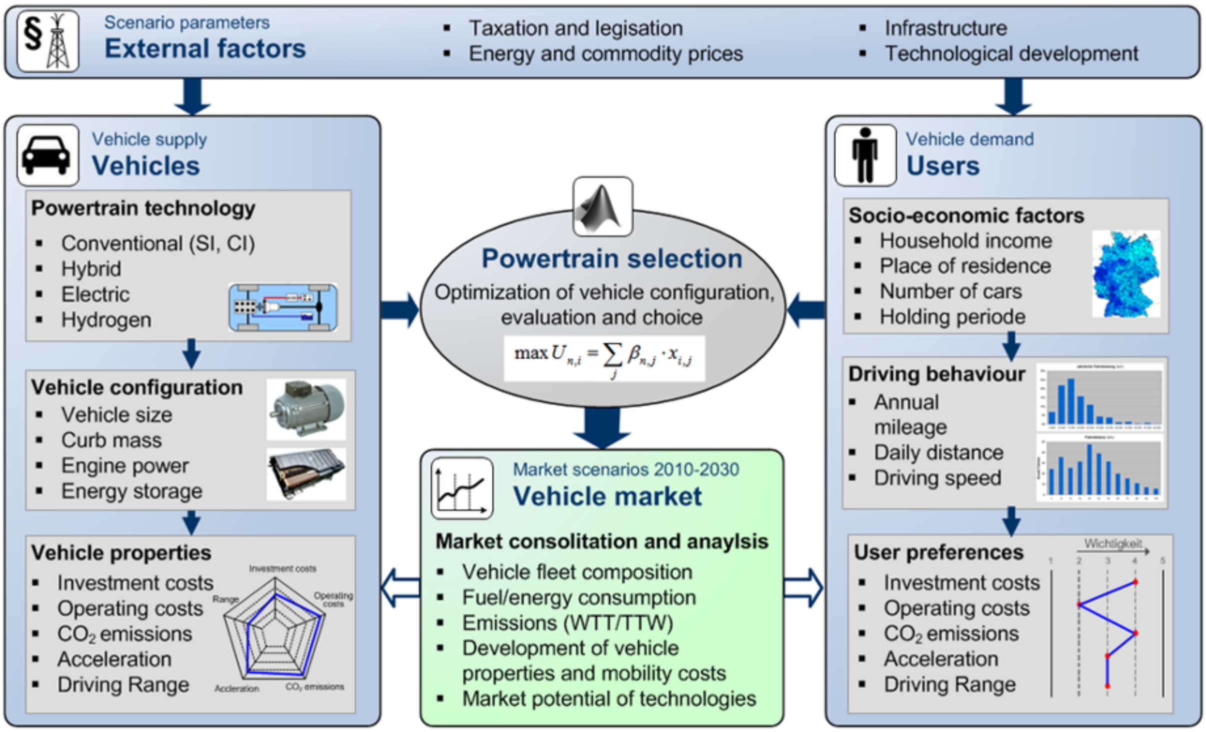


Figure 3.2: Structure of the market model. Extracted from [114]

The structure has different modules:

- **Vehicles**: A wide set of different drivetrain architectures is covered by the model. In this thesis, there are 8 different drivetrain architectures and 3 different segments. In total, there are 24 types of cars available for the agents. The drivetrains architectures are:
  - Battery Electric vehicle (BEV)
  - Compressed Natural gas (CNG)
  - Diesel (D)
  - Diesel hybrid electric vehicle (D-HEV)
  - Fuel cell electric vehicle (FCEV)
  - Gasoline (G)
  - Gasolinehybrid electric vehicle (G-HEV)
  - Gasoline plug-in hybrid vehicle (G-PHEV)

  The segments are:
  - small
  - medium
  - large

The 24 different cars available present a specific power, acceleration, consumption, CO_2 emissions, and price before taxes. It has been used the data set available, and prices of cars have been adjusted to the reality of 2021-2022. The model integrates vehicles and properties and considers the dynamic evolution of prices under different scenarios.

- **User**: The model creates different agents (costumers) depending on socioeconomic factors, driving behavior, and individual preferences by applying stochastic distribution functions. The underlying empirical data is based on actual data extracted from each country modelled. Data provided by DLR has been used, and 3 main features have been searched for all countries:
  - Household income distribution at a regional level (NUTS3)
  - yearly mileage (national level)
  - new registrations of cars (NUTS3)

  The detail level in the data depends on what has been possible to find. Homogenous sources and criteria have been used when selecting data for all countries.

- **External factors**: A large set of scenario parameters that influence the purchase decision is included in the model. The following parameters have been used:
  - CO_2 regulations: Already included in the model. CO_2 regulations are set by the EU. In 2021, the average CO_2 emissions of the new sales for each car manufacturer brand should be below 95 g/100km. If this is not accomplished, car manufacturers must pay high economic penalties. In 2025, and 2030, lower targets are set, and in 2035 the target is 0 g/km.
  - Taxation and incentives: Car taxation and incentives have different schemes in each EU country. Each powertrain option has a different retail price in each EU country. Car taxation is explained in Section 4.1.
  - Energy prices: Energy prices of gasoline, diesel, natural gas, electricity, and hydrogen are different in each country. Gasoline, diesel, natural gas, and electricity prices have been estimated for different countries. Hydrogen prices have been assumed to be constant, and estimated values for Germany have been used for the rest of the countries.
  - Technology learning curves: Learning curves were already available in the V21 model. They were created considering the reference scenario of IEA (STATS) [144].
  - Infrastructure: Infrastructure deployment was already available for Germany in the model. The reference scenario is again used. It was out of the scope to research the future deployment of infrastructure in electricity in each country. It has been used Germany as a reference, and it has been assumed that the deployment in other countries is the same as in Germany, but the year of the deployment can be advanced or delayed with respect to Germany. The infrastructure availability of electricity and hydrogen is assumed to be:
    - Finland: 0 years delay

    - France: 1 year delay
    - Greece: 6 years delay
    - Italy: 5 years delay
    - Netherlands: 1 year advanced
    - Norway: 2 years advanced
    - Poland: 6 years delay

- **Vehicle market**: The results of the individual purchase decision are consolidated in the market module, which scales up the simulation results to represent each national car fleet. Finally, the clustering of countries done in Chapter 2 can be used to obtain representative sales results at an EU level.

The center of the agent-based model developed by Redelbach [114] and used in this thesis is the powertrain selection algorithm which is formulated as a utility maximization problem. Different decision functions can be applied. In this thesis, a linear utility function is applied. U represents the utility a user n assigns to a powertrain option p:

$$maxU_{n,p} = \sum_{i} \beta_{n,i} \cdot x_{p,i} + \varepsilon$$

The vehicle attributes x determines the utility score (depending on the powertrain technology p) and the individual preferences β (depending on the characteristics of user n). In the current model version, the following criteria i are still evaluated: purchase price, operating cost (incl. fuel, electricity, maintenance, and repair), CO2 emissions, acceleration performance, and driving range.

In an iterative process, each agent adjusts the car configuration according to his preferences and then chooses the option with the highest score.

Once the model is clearly defined, inputs and assumptions for obtaining results are explained in Section 4.

4 Demand-side and external factors of the vehicle market

4.1 Car taxation schemes

As a demand-side policy, taxation of vehicles plays a significant role in promoting the use of alternative fuel vehicles. Each country has very different car taxation systems. There are many taxes and considerations that need to be taken into account. Each country has different car taxations. For example, bonus-malus schemes where vehicles that are pollutants pay higher taxes and vehicles which accomplish certain conditions are exempted from the tax or can even get bonuses. Bonuses (or subsidies) help promote and make more affordable alternative vehicles.

The taxes and the incentives can be classified into 3 main groups:

- **Acquisition taxes and subsidies**: These taxes are only paid once at the moment of acquisition of the car. This group includes taxes such as VAT which is a certain % in each country, registration fees which can be constant or depend on different characteristics, and bonus-malus schemes.
- **Ownership taxation (or motor vehicle tax)**: This type of tax is paid annually. In some countries, it is separated between company and passenger cars. CO_2 emissions or engine size are some of the factors countries use to determine the amount to be paid.
- **Company car tax benefits**: The use of a company car for private motoring is treated as a benefit in kind under income tax. It is difficult to quantify precisely the real benefit amount because it depends on the beneficiary's gross income, the national tax system, and other characteristics. This group is neglected due to the complexity of modelling the amount of savings it would generate.

Once the different groups of taxes have been described, analysis for each characteristic country selected is performed. In this thesis, private passenger cars are studied. Only acquisition taxes and subsides and ownership taxes are considered.

4.1.1 Germany

4.1.1.1 Taxes on acquisition

4.1.1.1.1 VAT

VAT is applied at the rate of 19% on the sale of new vehicles.

4.1.1.1.2 Registration fees

The average duties collected at the time of an initial registration amount to €26.30.

4.1.1.1.3 Environmental bonus

A subsidy is given for buying battery electric and plug-in hybrid electric cars to boost the intake of electric vehicles and make the cars more affordable for customers in Germany.

In 2020 and 2021, a summary of the bonus can be observed in Table 4.1

year	2020		2021	
Powertrain	BEV	G-PHEV	BEV	G-PHEV
Net list price below 40.000 €	6.000 €	4.500 €	9.000 €	4.500 €
Net list price above 40.000 €	5.000 €	3.750 €	7.500 €	3.750 €

Table 4.1: Environmental bonus in Germany in 2020 and 2021

In 2022, incentives are the same as in 2021 for BEV and G-PHEV. In 2023, there will be no more incentives for G-PHEV. From 2023 to 2025, incentives are still available for BEV.

If more information about environmental bonus wants to be consulted see Appendix 8.1.

4.1.1.2 Taxes on ownership

4.1.1.2.1 Motor vehicle tax

The annual circulation tax (or ownership tax), which is paid yearly, is based on motor characteristics in Germany. Since July 2009, passenger car taxation has depended on CO_2 emissions and cylinder capacity.

In 2020, the German government changed the annual circulation tax for newly registered cars from 1st January 2021. Before, the CO_2 element had a linear rate, but now it increases progressively. The tax base regarding cylinder capacity remains unchanged.

CO$_2$ component

The CO_2 component has an interval that is tax-free. In 2021 this value was 95 g/km. Between 2009 and 2020, the tax's only modification was related to this threshold. The linear rate remained constant as shown in figure 4.1. The actual intervals with progressive rates are shown in Table 4.2:

The progressive tax has the purpose of making it much less attractive to have vehicles with a high level of emissions.

Capacity component

Since July 2009, the capacity component has remained the same. It is a linear function of the motor's cubic cylinders (cc). Electric and PHEV cars are exempted from this tax. For each additional 100 cc, the tax is:

- 2€ for petrol engines
- 9,5€ for diesel engines

CO_2 (g/km) intervals	Tax rate (€ per additional gram)
0-95	0
96-115	2
116-135	2,2
136-155	2,5
156-175	2,9
176-195	3,4
>195	4

Table 4.2: CO_2 component of the motor vehicle tax for vehicles registered after 1st January 2021

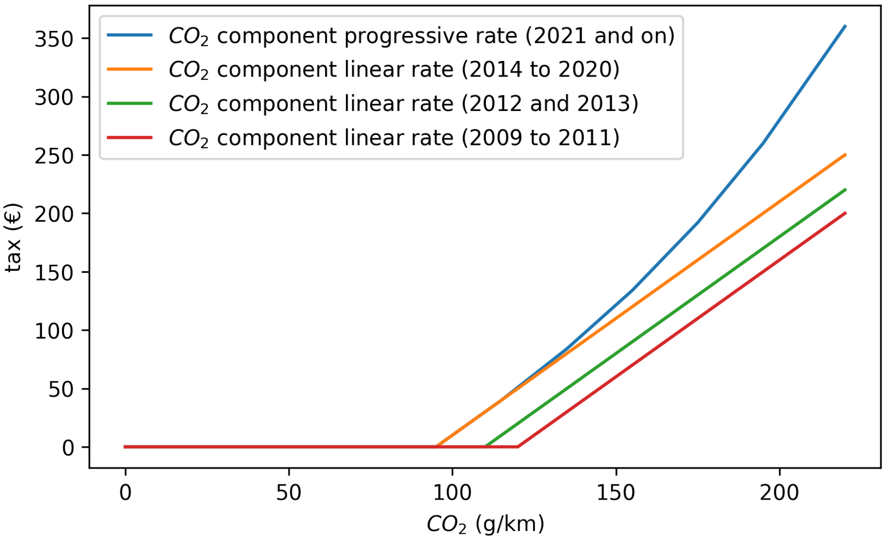


Figure 4.1: CO_2 component of motor vehicle tax in Germany at different registration periods

4.1.2 France

Before introducing taxation in France, fiscal power is introduced. It is a unit defined by the french government and used for tax purposes. Before, it was based on the vehicle's CO2 emissions and engine power characteristics.

For internal combustion engine vehicles approved by the WLTP on 1st November 2019, fiscal power (P_A) is defined only by the engine power (P). The formula is:

$$P_A = 1,80 \cdot (P/100)^2 + 3,87 \cdot (P/100) + 1,34 \tag{4.1}$$

For electric vehicles, the formula is:

$$P_A = 1 + 0,136 \cdot P \tag{4.2}$$

The engine power (P) is given in KW, and the fiscal power (P_A) unit is defined as horsepower, but it is not a horsepower in physical terms.

Once the fiscal power has been explained, the different taxes can be described.

4.1.2.1 Taxes on acquisition

4.1.2.1.1 VAT

VAT is applied at the rate of 20% on the sale of new vehicles.

4.1.2.1.2 Registration fees

A fixed tax, up to €4 per certificate issued, is assigned to the National Agency managing the registration process.

4.1.2.1.3 Regional component of the registration tax

The tax depends on the vehicle's fiscal power and the region's horsepower cost. It can be calculated using Formulas 4.1 or 4.2 in the case of an electric motor. The horsepower cost by region has been collected for each region. Some features can be highlighted:

- The charges per unit of horsepower varied from €27(minimum) to €51.20 (maximum) in 2021.
- Exemptions depend on the type of motorization of the vehicle. The French Ministry of Public Affairs [71] describes two groups receiving exemptions:
  1. **clean vehicles**: This category includes all vehicles running exclusively on electricity, hydrogen, or a combination of these two energies. **These vehicles are completely exempted from the tax independently of the region.**
  2. **so-called "clean" vehicles**: This category includes a wider type of drive trains. It includes plug-in hybrid electric, E85 super ethanol, natural gas, or even liquefied petroleum gas. The exemptions depend on the region. They are 50% or 100% depending on the region except in Overseas France, where there is no exemption. Consult the exemption in Table 8.5

Depending on the region, the taxation is more permissive with ICE vehicles than in others. For example, a SUV with a 150 KW engine (for example, an Audi Q5) would have a fiscal power of 11.2 HP, which means paying 302.4 € in Corse or 571.2€ in Bretagne. The fiscal power increases quadratically. Then, ICE sport cars or SUVs with big engines would be the ones that are paying higher registration taxes.

See fiscal power cost by region and exemptions in Table 8.5 in Appendix 8.2.1

4.1.2.1.4 ecological bonus malus

France firstly introduced the bonus-malus scheme in January 2008. The model has been adjusted in successive years, and it is a model which several countries have adopted.

The ecological malus ("eco-tax") is applied by increasing the cost of the registration certificate for penalizing the purchase of a new vehicle that is considered a pollutant. It was set up alongside the ecological bonus, which encourages the purchase of clean vehicles. Additionally to the bonus malus, there is a scrapping scheme ("prime à la conversion"), a premium granted to car buyers for scrapping old vehicles.

Ecological malus

Only a vehicle's CO_2 emissions determined ecological malus, and in 2022 it is introduced a new component of the malus which is based on the weight of the vehicle.

- **CO_2 based malus**

  CO_2 malus component has been based on WLTP values since 1st March 2020 and is described by an exponential curve that depends on the number of CO2 emissions. Adopted at the end of 2020, the recent Finance law 2021 includes a substantial increase of the maximum payable malus for new high-CO2-emitting vehicles. It increases from €20,000 in 2020 to €30,000 in 2021 and 40.000€ in 2022.

  The values have been extracted from the French Ministry of Public Affairs in [43] and can be consulted in Appendix 8.2.2. A graphical representation of the ecological malus is shown in Figure 4.2.

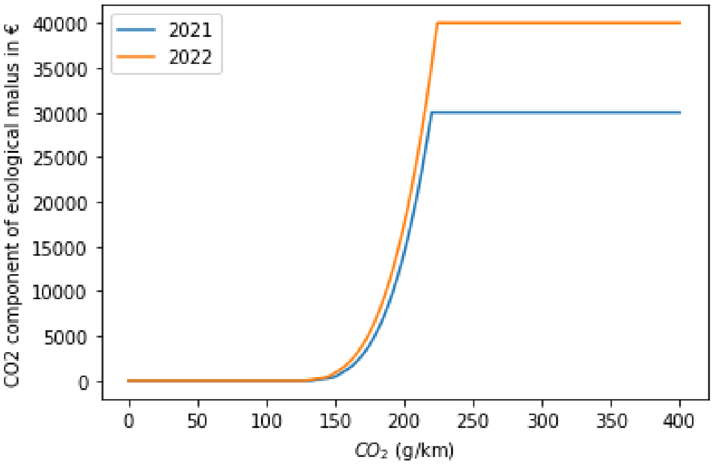


Figure 4.2: CO_2 component of ecological malus tax in France in 2021 and 2022

In Figure 4.2, it can be observed that the rate is very exponential. The difference between 2021 and 2022 lies in the maximum value, which can be 10.000 higher in 2022. The ecological malus is a national registration tax, which is much more relevant than the regional tax because it can reach way higher values.

  Cars that are exempted from paying this tax in 2022 are cars that have emissions equal to or lower than 128 g/km. Additionally, other partial exemptions are:

  - Purchase of a car with at least 5 seats: reduction of 20 g/km per child depending on the CO2 rate. It starts to count from the third child and on.
  - Vehicle running on ethanol: 40% reduction in CO_2 levels.

- **Mass-based malus**

  From 1st January 2022, a new penalty based on the mass on running order (empty weight of the driver + 75 kg of the driver) of new vehicles was implemented to penalize the acquisition of heavier vehicles which are considered to be the most polluting. This new tax will affect categories such as SUVs, 4x4s, certain sedans, and convertibles. This tax is a good measure considering the 15% increase of the average mass of new cars in the EU since 2001 levels [40]. In order to be less pollutant, the change of powertrain is essential, but it is also crucial to have more energetically-efficient cars. This can be done with better aerodynamics, lighter cars, and other improvements.

  This tax affects cars whose weight exceeds 1,800 kg, which will pay €10 per additional kilo beyond this threshold.

  Cars that are exempted from this tax are:

  - Electric vehicles and PHEV with an urban ZEV autonomy over 50km: Even if their total load exceeds the prescribed limit of 1,800 kg.
  - Family vehicle: A reduction of 200 kg per child is recorded to avoid families having to pay a surcharge on the price of their car registration document.

The amount of this new weight penalty will be added to the price of the regional registration tax (See in Section 4.1.2.1.3) and will be added to the ecological penalty detailed earlier. However, the amount of taxes will be capped at €40,000.

$$\left.\begin{array}{l}\textbf{Regional tax} \rightarrow \text{based on fiscal power}\\ \textbf{Ecologic malus component 1} \rightarrow \text{based on CO}_2\text{ emissions}\\ \textbf{Ecologic malus component 2} \rightarrow \text{based on weight}\end{array}\right\} \leq 40.000\text{€}$$

The scrapping scheme is not modelled in Vector21. Information about scrapping scheme can be consulted in Appendix 8.2.3.

Ecobonus

Apart from the scrapping scheme, an ecobonus was ratified on Decree No. 2021-1866 of 29th December 2021. [132] There are subventions for electric vehicles and plug-in hybrid electric vehicles (CO_2 emissions between 20 and 50 g/km). Until July of 2022, the different incentives are shown in Table 4.3.

In July 2022, the incentives are lowered by 1,000 €, and then at the end of 2022. At the end of each year, the scale will be lowered by 1,000 €.

	retail price before taxes	**% of subvention of retail price before taxes**	**maximum amount**
electric	under 45000 €	27%	6.000,00 €
electric	from 45000€ to 60000€	27%	2.000,00 €
PHEV	up to 50000€	27%	1.000,00 €

Table 4.3: Ecobonus for buying a new car in France until 30th June of 2022. The scale will be lowered by €1,000 from 1st July 2022

4.1.2.2 Taxes on ownership

The tax group relies mainly on company cars. The characteristics of this tax can be read in [131]. This tax is not modelled because it requires knowing additional features of the company in order to be able to apply certain deductions.

There is no ownership tax for private passenger cars. From 2012 to 2020, cars are emitting more than 190 g/km needed to pay an annual malus. In 2021, it was suppressed (See more info in Appendix 8.2.4).

4.1.3 Italy

4.1.3.1 Taxes on acquisition

4.1.3.1.1 VAT

VAT is applied at the rate of 22% on the sale of new vehicles.

4.1.3.1.2 Registration fees

The first registration of the vehicle requires paying different fees. The different registration fees are presented in Table 8.8 in Appendix 8.2.5. The total registration fees are 142,98 €. Data has been extracted from ACI [124] and from [70].

4.1.3.1.3 Registration tax

Registration tax or IPT (Imposta Provinciale de Transcrizione) constitutes the greatest economic burden of registration taxes. The tax is divided into two and defined by territorial jurisdiction:

- National: It is the basic charge of the tax
- Provincial: The national amount can be increased by up to 30%

National

The National tax authorities define the national amount [38] and is defined by the power in KW. Passenger cars have a threshold until a fixed tax is paid; after that, a constant price per additional KW is paid. The tax can be seen in Table 4.4:

53 KW or lower	150,81 €
more than 53 KW	3,5119 €/KW

Table 4.4: Registration tax: National amount

Provincial

Once the national tax is defined, there is an increase until 30%, which is the decision of each province. ACI presents a list with the different provinces' percentages in [104]. Regional tax increases and ecological partial exemptions can be consulted in Appendix 8.2.7. Ecological partial exemptions are given to BEV, PHEV, HEV, and CNG cars, depending on the region.

4.1.3.1.4 Bonus Malus scheme

The bonus-malus scheme has substantial differences from the previous year, 2021. In 2019, 2020, and 2021, there was a bonus called ecobonus and a malus called ecotax (in Italian *ecotassa*). These policies expired on 31st December 2021, and there are currently (on 26th April) being approved in the Budget Law for 2022 [18].

Ecobonus

The Ecobonus is a subvention for low-emission cars. It is defined by the level of CO_2 emissions, which mainly depends on the powertrain. It is defined for the period 2022-2024. The budget has been reduced from 700 million per year to 615 million per year. The funds assigned yearly depend on the level of emissions and can be seen in Table 8.9 in Appendix 8.2.8.

The subventions are higher for cleaner vehicles and include a strategy for reducing the amount of old and pollutant cars in the market. The subvention given is higher if the scrapping of an old car is done. Only cars of Euro 5 or lower are accepted. Newer cars are not accepted for the bonus. The subventions given from 2022 to 2024 can be seen in Table 4.5.

The subventions have been considerably reduced from 2021. The maximum ecobonus with scrapping that could be obtained was 10.000 € in the case of electric vehicles. It has been reduced to 5.000 € in the new Budget Law, which is currently being approved. The reduction of the Ecobonus will reduce the electric vehicle intake in Italy, which could have been achieved if previous incentives had been maintained.

Apart from the national ecobonus, there are 2 regions and 1 city which give subventions when purchasing low-emission vehicles (See appendix 8.2.8).

Ecotax

CO_2 (g/km)	**motorisation**	**with scrapping**	**without scrapping**	**price threshold before VAT and IPT**
0-20	electric	**5.000,00 €**	**3.000,00 €**	**35.000,00 €**
21-60	plug-in hybrid electric	**4.000,00 €**	**2.000,00 €**	**45.000,00 €**
61-135	hybrid	**2.000,00 €**	**- €**	**35.000,00 €**
	gasoline			
	diesel			
	methanol			
	GLP			

Table 4.5: Ecobonus of 2022-2024: Incentives for purchasing new cars defined in the Budget Law of 2022 (currently being approved). Data extracted from [72] and [25]

The malus known as ecotax from previous years has not been included in the Budget Law of 2022. However, excluding the ecotax from the Budget Law did not prevent this measure from being re-proposed with a different name. The exact details of the amount to be paid and at which threshold are at 26th April 2022 still unknown.

In different articles [25], [45], it is described how the government may apply even a harder ecotax than the existing one. However, it is unknown. Then, it is decided to use the ecotax of 2021 to model the malus of Italy.

The malus is penalizing the most polluting cars. The level of CO_2emissions defines it, and since 2021 the emissions are measured using WLTP. The different values can be seen in Table 4.6

CO_2 emission ranges	**malus (€)**
191-210	1100
211-240	1600
241-290	2000
291 and more	2500

Table 4.6: Malus tax values of Italy in 2021. CO_2 emissions are measured using WLTP. Data extracted from [44]

4.1.3.2 Taxes on ownership

Ownership tax (in Italian "bollo auto") is paid annually and depends on the power of the vehicle expressed in KW and the euro standards (between Euro 0 and Euro 6). Each regional government determines the amount to pay. A Table of the price that new cars following Euro 6 need to pay can be consulted in Table 8.10 in Appendix 8.2.9.

The tax rate varies by region; if the car has power under 100 KW, it pays a specific rate. If this threshold is overcome, a higher rate is paid for the additional KWs. The case of Piemonte is unique because it creates 6 different tax rates depending on the car's power and can be seen in Table 4.7.

Power < 53 KW	**Power < 100 kW**	**Power < 130 kW**		**Power > 130 KW**	
		first 100 KWs	**Next KWs**	**first 100 KWs**	**Next KWs**
2,58	2,73	2,79	4,18	2,84	4,26

Table 4.7: Annual ownership tax rate in Piemonte separated by power for Euro 5 and 6 vehicles. Rate expressed in €/KW

Super bollo

An additional tax called *super bollo* taxes high-powered cars. Cars that have more than 185 KW pay 20 € per additional KW above 185 KW. [66].

4.1.3.2.1 Exemptions

There are two main exemptions at a national level depending on the vehicle's powertrain.

National exemptions

- **Electric vehicles**: 5 years exemption of the ownership tax after first registration (According to Article 17 of Presidential Decree 39/1953 [31])
- **methane and LPG cars**: 75% reduction of the ownership tax (Law n. 449 of 27: 12: 1997, Art. 17, paragraph 5 [88])

Provincial exemptions

Each province has its exemptions. There are exemptions for hybrid vehicles, for electric vehicles, which increase the period for not paying the tax, or for GLP and methane vehicles which exempt the payment of the tax, amongst others. For a detailed explanation of all the different exemptions by province consult Appendix 8.3.

4.1.3.3 Taxes on motoring

4.1.3.3.1 Periodic review of CNG cylinders

The review of the cylinders is first performed four years after the vehicle registration and then every two years after that.

The cost of auditing and testing CNG cylinders varies depending on the size of the engine of the car. Small cars like Fiat Panda or Volkswagen Golf cost about 120 €, and bigger cars with 4 or 5 cylinders cost between 350€ and 450€.

4.1.4 Netherlands

4.1.4.1 Taxes on acquisition

4.1.4.1.1 VAT

VAT is applied at the rate of 21% on the sale of new vehicles.

4.1.4.1.2 Registration fees

Registration fees for passenger vehicles are paid to Netherlands Vehicle Authority (RDW) when registering a new or used passenger car. The registration charge is 53.40€.

4.1.4.1.3 Registration tax

A registration tax, the Belasting van Personenauto's en Motorrijwielen (BPM), is paid by all passenger cars. The CO_2 emission determines the BPM tariff for a passenger car. Since July 2020, WLTP measurement has been used to calculate gross BPM for new cars. Depending on the powertrain, a different amount is paid. The tax rates and exemptions in 2022 are the following:

- Electric cars: BPM-free. They are completely exempted from the tax.
- Gasoline and Diesel vehicles: There is a base tax of 376€ which all cars pay, and then there is a progressive tax rate defined depending on the g/km, which is determined for different threshold values (See Table 4.8). Additionally, diesel cars pay an additional tax rate of 86,67 €/(g/km) from 75 g/km.

From	Up to	base value in the threshold (€)	Tax per g/km (€)
0	0	0	0
1	84	0	1
85	109	84	62
110	152	1634	137
153	168	7525	224
169		11109	448

Table 4.8: BPM rates for petrol and diesel passenger cars and base values by threshold

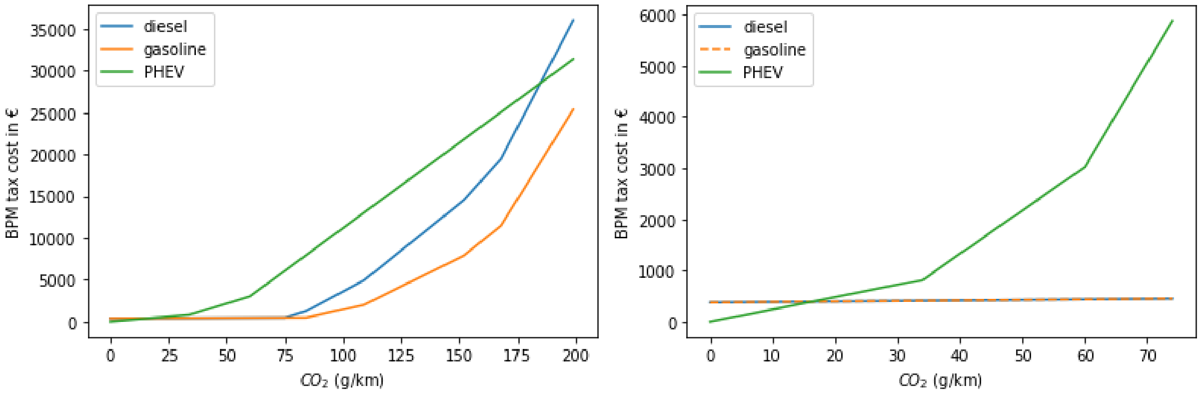
- Alternative fuel engines like LPG, CNG, and E85: They follow the same taxation as gasoline
- Hybrid vehicles: They follow the taxation of gasoline or hybrid depending on the type of fuel it is used.

- Plug-in hybrid electric: Plug-in hybrid electric vehicles are exempted from the base tax of 376€ like electric vehicles. Different bpm rates are applied for plug-in hybrid electric (See Table). It can be observed that PHEVs with emissions lower than 30 g/km have very low taxation, but afterward, the tax rate increase is considerable.

From	Up to	Initial value base in the threshold	Tax per g/km (€)
0	34	0	24
35	60	816	85
61		3026	204

Table 4.9: BPM rates for passenger PHEV and base values by threshold

After analyzing the BPM, a plot of the evolution of the tax cost with respect to the level of emissions depending on the type of car is shown in Figure 4.3. Battery electric cars are not presented in the graph because they have 0 g/km emissions and are exempted from the tax.



(a) Graph

(b) Zoomed in graph

Figure 4.3: BPM tax in € depending on the level of emissions and the powertrain

In Figure 4.3 is very visual of the progressive tax rate of BPM. Diesel pays considerably more than gasoline for the same number of emissions (If they are above 75 g/km). PHEVs are relatively low taxed compared to gasoline and diesel only if the emission are lower than 16 g/km (See Figure 4.3b). After this point, the bpm tax is higher. This can be due to several reasons, as the recent discoveries have shown that declared emissions of PHEV are considerably higher in real driving [107]. The BPM tax shows that the Netherlands government does not want to promote the sale of big large-segment cars.

4.1.4.1.4 Subventions: SEPP

SEPP is a subsidy scheme for electric passenger cars for private individuals in the Netherlands. The scheme runs from 1st July 2020 to 31st December 2024. The budget for each year is determined annually. In 2022, 71 million € for new passenger cars have been given. On 30th

May 2022, less than 12 million are remaining. The requirements for obtaining the subsidy are the following [128]:

- A 100% electric passenger car with a range of at least 120 kilometers (measured by WLTP)
- The retail price before taxes as registered in the vehicle registration register (RDW) is not lower than 12,000 € and not higher than 45,000 €.

If conditions are fulfilled, a fixed subsidy of 3,350€ can be obtained in 2022. The subsidy which can be obtained will be reduced next year and can be seen in Table 4.10

	subsidy	**budget (in millions of €)**
2020	4.000,00 €	10
2021	4.000,00 €	14,4
2022	3.350,00 €	71
2023	2.950,00 €	Unknown
2024	2.550,00 €	Unknown

Table 4.10: SEPP subsidy amounts for different years

4.1.4.2 Taxes on ownership

Ownership tax or motor vehicle tax is in dutch called *Motorrijtuigenbelasting* (MRB). MRB is a tax that depends on:

- **Gross vehicle weight**: The categories are separated every 100 kg
- **Type of fuel used**: Electric, PHEV, gasoline, diesel, LPG, and CNG
- **Region**: Netherland provinces
- **CO_2 emissions**: Exemptions are given depending on the level of emissions

The base of calculation is done at a national level, and the base case is gasoline, where there are fixed tariffs for gasoline depending on the weight. Then it is increased by a provincial general tariff multiplied by a percentage depending on each province (Provincial percentages can be consulted in Appendix 8.4.1.1).

The lack of information in the tax calculation has made that tariffs have been consulted in [23] for each weight category and 2 regions. The calculus for obtaining the different values is explained in Appendix 8.4.2.

Electric and hydrogen

Electric and hydrogen vehicles are completely exempted from the motor vehicle tax in 2022. The exemption will vary in the following years [32] and can be seen in Table 4.11.

The calculation of electric tariff will be calculated using gasoline tariffs which are the base rate.

Plug-in hybrid electric (PHEV)

2022, 2023 and 2024	100 % exemption
2025	75% exemption
From 2026 and on	No exemption

Table 4.11: Electric and hydrogen ownership tax exemption from 2022 and on

Similar to EVs, PHEVs present an exemption. PHEV cars with a CO_2 emission of more than 0 grams per kilometer but not more than 50 grams per kilometer pay half the tax in 2022 [91]. The exemption will be reduced in the following years and can be seen in Table 4.12.

2022, 2023 and 2024	50 % exemption
2025	25% exemption
From 2026 and on	No exemption

Table 4.12: PHEV (with less than 50 g/km) ownership tax exemption evolution

Hybrid or plug-in hybrid vehicles with emissions higher than 50 g/km do not have exemptions.

4.1.5 Norway

4.1.5.1 Taxes on acquisition

4.1.5.1.1 VAT

VAT is applied at the rate of 25% on the sale of new vehicles.

Exemption for battery electric vehicles

Sales and leasing of vehicles that only use electricity for propulsion are exempt from VAT. The exemption for the sale of electric cars was introduced on 1st July 2001. PHEV cars are not included because the exemption does not cover cases where the battery is supplied with power while driving using an external piston-driven internal combustion engine. [134]

A new subsidy scheme has been proposed for 2023. Consumers would pay VAT on all-electric cars that cost more than 50,025.01 € (500,000 NOK) [96]. The VAT charges will be introduced from 1st January 2023 if the budget proposal is not modified and approved. If a car costs 600,000 NOK, VAT would be paid only for the 100,000 NOK above 500,000.

4.1.5.1.2 Registration fees

The registration fees are called scrap deposit tax. Passenger cars pay 240,12 € (2400 NOK) for registering a new car.

4.1.5.1.3 Registration tax

The one-off registration tax is an excise duty that must be paid upon initial motor vehicle registration in Norway.

The motor vehicles are separated into different groups, and the amount they need to pay depending on different factors is defined in [127]. The individual motor vehicle types are divided into tax groups, with, for example, passenger cars placed in tax group a and van vehicles in tax group b.

The one-off registration tax is calculated based on [7]:

- **vehicle's tax group**: Passenger cars are located in the group a.
- **Kerb weight**: The kerb weight is, in most cases, obtained from the motor vehicle's European type approval. The weight component is progressive, and the tax rates are defined for different thresholds in Table 4.13.

Kerb weight in kg	NOK per kg	€ per kg
0-500	0	0,00
501-1200	27,15	2,72
1201-1400	67,68	6,77
1401-1500	211,49	21,16
over 1500	245,97	24,61

Table 4.13: One-off registration tax kerb weight component for different thresholds in NOK per kg and € per kg.

- **CO_2 emissions**: : Vehicle emissions are calculated using WLTP since September 2017. The CO_2 emissions component has a progressive scheme as the weight component. The tax rates are defined in Table 4.14.

CO2 emissions (g/km)	kronen per g/km	€ per g/km
0-87	0	0,00
88-118	1095,4	109,59
119-155	1227,52	122,81
156-225	2382,64	238,38
over 225	3800,83	380,27

Table 4.14: One-off registration tax CO_2 emissions component for different thresholds in NOK per kg and € per kg.

- **NO_X emissions**: Gasoline and diesel must pay an additional tax for NO_X emissions. Gasoline and diesel have a constant tax rate scheme and pay 7.82€/mg (78.4 NOK/mg). This value is used when calculating the one-off tax for motor vehicles with a value determined for NO_X emissions registered in the Vehicle Register. If not, this emission is set at the maximum value that the vehicle in question can have according to the road authorities' regulations. The maximum amount paid for gasoline is 60 mg/km and the maximum for diesel is 80 mg/km.

- **cylinder capacity**: Stroke volume (cylinder volume) stated in cm^3 is used as an alternative tax base instead of CO_2 emissions for passenger vehicles for those cases where the motor vehicle is not obliged to document CO_2 emissions and/or information on CO_2 emissions.

Exemptions and reductions

Vehicles with a specific powertrain have an exemption from the one-off registration tax. The exemption applies to the following vehicles:

- **Electric cars**: 100% exemption
- **Hydrogen cars**: 100% exemption
- **PHEV**: reduction of 15% of its kerb weight when calculating the weight component tax.
- **Ethanol-powered motor vehicles**: 1000,50 € (10,000 NOK) reduction.

4.1.5.2 Taxes on ownership

There was an annual motor vehicle tax in 2017 and earlier. In 2018, the annual motor vehicle tax was replaced by a road traffic insurance tax collected by the insurance companies [9]. The tax has been modified in recent years depending on the engine type. Electric and hydrogen cars paid less in the past, but since March 2022, the amount paid is the same for electric and hydrogen cars than for diesel and gasoline (See road traffic insurance tax from previous years in Appendix 8.5.1). Since March 2022, the road traffic insurance tax is for all types of passenger cars except diesel without particle filters 297,62€. Diesel cars without particle filters pay 349.48€.

The region of Svalbard is exempted from the road traffic insurance tax.

4.1.6 Poland

4.1.6.1 Taxes on acquisition

4.1.6.1.1 VAT

VAT is applied at the rate of 23% on the sale of new vehicles [84]. The base of VAT for calculating the tax is the sum of the price of the vehicle plus the amount of the excise tax (See excise tax in Section 4.1.6.1.3).

4.1.6.1.2 Registration fees

There are different fixed fees which can be consulted in Table 8.15 in Appendix 8.6.1. The amount of the registration fees is 52.87€. The fees are low compared to the excise duty and VAT.

4.1.6.1.3 Registration tax

Any new car sold in Poland or registered in Poland for the first time must pay the excise duty tax [139]. The tax rate is calculated by multiplying the tax rate by the vehicle value. In 2022, the following tax rates (See Table 4.15) apply depending on the size of the engine (cubic centimeters) and the type of motorization.

size of engine	type of motor	tax rate
0-2000 cm^3	general	3,1%
2001cm^3 or more	general	18,6%
0-2000 cm^3	hybrid	1,55%
2001cm^3 to 3500 cm^3	hybrid	9,3%
0-2000 cm^3	PHEV	0,00%
2001cm^3 to 3500 cm^3	PHEV	9,3%
0-2000 cm^3	electric	0,0%
2001cm^3 or more	electric	0,0%
0-2000 cm^3	hydrogen	0,0%
2001cm^3 or more	hydrogen	0,0%

Table 4.15: Excise duty rate depending on the size of the engine and the type of powertrain. Data extracted from [139]

The excise duty exemption exists for electric, hydrogen, and PHEVs with an engine smaller than 2000 cm^3. Hybrids present a lower excise duty than gasoline, diesel, or another type of engines.

4.1.6.1.4 Subventions: "my electric" program

The program "my electric" started in 2020, and it finishes at the end of 2025 [92]. It consists in a reduction of 15% of the price of the car which the Polish Government subsidizes when purchasing zero-emission cars (electric or hydrogen) with a price lower than 49,464.68€ (225,000 PLN). The maximum subsidy that can be obtained is 4.122,06 € (18,750 PLN) for natural persons and 5,935.76 € (27,000 PLN) for natural persons if they possess a large family card (at least 3 sons).

The program's budget is more than 100 million € (500 million PLN). However, in the first year, there were only 261 applications, and only 8% of the budget of the program for 2020 was used [19] showing that this measure is not adequate for having a higher electric vehicle intake in Poland.

4.1.6.2 Taxes on ownership

There are no existing taxes on ownership in Poland. There have been proposals in different cities like Warsaw, one of the most air-polluted cities in the world, but there is still no annual motor vehicle taxation or bans on driving inside cities.

4.1.7 Greece

4.1.7.1 Taxes on acquisition

4.1.7.1.1 VAT

VAT is applied at the rate of 24% on the sale of new vehicles.

4.1.7.1.2 Registration fees

The registration fees in Greece consist of issuing the registration document and the license plate. This costs 75€. [116] Additionally, there is a transfer fee, which varies depending on the car's cubic centimeters. It is on average 120€. Then, a passenger car pays an average of 195€ in Greece.

4.1.7.1.3 Registration tax

The registration or excise tax is paid once when a car is registered. The registration tax for new cars depends on Euro exhaust emission standards, CO_2 emissions, and the car's retail price defines a progressive rate [103]. The registration tax of a new car can be calculated by finding the tax rate using the formula 4.3

$$registration\, tax\, rate = coeff.\, portion \cdot CO_2\, coeff. \tag{4.3}$$

The tax rates depending on the retail price before taxes (in Formula 4.3 coeff. portion) can be seen in Table 4.16.

Retail price before taxes	**tax rate**
up to 14000	4%
From 14001 to 17000	8%
From 17001 to 20000	16%
From 20001 to 25000	24%
From 25001	32%

Table 4.16: Registration levy rate depending on the retail price.

CO_2 emissions coefficient (in Formula 4.3 CO_2 coeff.) depends on the level of emissions calculated with WLTP system since 2021. The coefficients for euro 6-d standards in 2022 can be seen in Table 4.17.

CO2 emissions (WLTP)	CO2 emissions coefficient (2022)
0-100	95%
101-120	100%
121-140	110%
141-160	120%
161-180	130%
181-200	140%
201-250	160%
>250	200%

Table 4.17: CO_2 emissions coefficient in 2022 for euro 6d passenger cars

With the CO_2 emissions coefficient and tax rate for different retail price groups, the registration tax rate can be calculated, and registration tax can be obtained. An example of how the registration tax is calculated can be consulted in Appendix 8.7.1.1

Exemptions and reductions

Vehicles with certain types of powertrains can receive an exemption or reduction of the registration tax. The exemption applies to the following vehicles:

- **Electric cars**: 100% exemption
- **Hybrid cars**: 50% exemption

4.1.7.1.4 Subventions

I move electrically 2 is a subvention program for electric vehicles, and it is currently being approved in May 2022. It will give 75 million for 2022 and 2023. The requests can be done until 31st December 2023.

The subvention rate increased from 20% in 2021 to 30% in 2022 and a maximum retail price has not been established. The maximum amount of the subvention is 8,000 €. [33]

Subventions for PHEV or hybrid have not been mentioned in the government's declaration. The declaration has not yet been ratified.

4.1.7.2 Taxes on ownership

4.1.7.2.1 Annual circulation tax

The annual circulation tax depended on engine capacity until 2010. Since 2010 it has depended on CO_2 emissions. Before, the NEDC cycle was used, but since January 2021, the annual circulation tax depends on CO_2 emissions measured with WLTP [11].

The payment of annual circulation tax is determined in accordance with Table 4.18.

CO2 emissions (g/km)	annual road tax per grams of CO2
0-122	0
123-139	0,64
140-166	0,7
167-208	0,85
209-224	1,87
225-240	2,2
241-260	2,5
261-280	2,7
>281	2,85

Table 4.18: Annual circulation tax coefficient in 2022 depending on the level of CO_2 emissions measured with WLTP

The tax coefficient is paid for all the emissions and not by thresholds. For example, the owner of a vehicle emitting 149g/km of CO2 (WLTP) is liable to pay an annual circulation tax of: 149·0.70€ = 104.30€.

Since November 2010, there have been no specific exemptions for electric and hybrid vehicles. Electric vehicles do not pay annual tax because the emissions are 0 g/km, and hybrid, gasoline, or diesel vehicles with emissions lower than 122 g/km neither pay an annual tax.

4.1.7.2.2 Luxury living tax

There is an additional yearly tax in Greece when owning a vehicle. The tax is called luxury living tax and depends on the presumed personal income.

Personal income presumption system

A certain income for passenger cars is presumed to depend on the engine capacity (cc). There is a reduction of 30% for cars older than 5 years, and cars older than 10 years do not pay the tax. The presumed income is defined in Article 16 of Law 2238/1994 [80] and Table 4.19 is shown for observing how the presumed income is calculated.

When calculating, the engine capacity is rounded to the nearest hundred.

Tax

Once presumed income can be calculated, luxury living tax can be determined. This tax is defined in Article 44 of Law 4111/2013 [81]. It has existed since 2012 and is paid only by cars

From (cc)	Up to (cc)	Initial value base in the threshold	Tax per 100cc
0	1200	4000	0
1201	2000	4000	600
2001	3000	8800	900
3001		17800	1200

Table 4.19: Accumulated income presumption system for passenger cars. Income for each range and increase every 100 ccs is shown

engine capacity	% of presumed income
cc <= 1929	0%
cc >1929 & cc <= 2500	5%
cc >2500	10%

Table 4.20: Luxury living tax rates. Data extracted from [81]

with an engine capacity above 1929 cc. Cars above this threshold pay a relative percentage with respect to the presumed income. The different cases are presented in Table 4.20

Exemptions

Electric cars do not pay luxury living tax; hybrid cars pay only by the internal combustion engine part.

4.1.8 Finland

4.1.8.1 Taxes on acquisition

4.1.8.1.1 VAT

VAT is applied at the rate of 24% on the sale of new vehicles.

4.1.8.1.2 Registration fees

The registration fees in Finland consist of the vehicle registration plate. It costs 5,50€ [36].

4.1.8.1.3 Registration tax

Finland's registration tax is called car tax. Car tax is applied to new cars which have not been registered yet. The car tax is determined based on the general consumer price of the car (retail price plus VAT). The basis is multiplied by a rate that depends on the CO_2 emissions (measured with the WLTP method) [26]. In Figure 4.4, the rate can be observed. Before October 2021, rates were the same, except for zero-emissions cars. BEV and FCEV are in 2022 exempted from paying the tax, and prior to this modification, they were paying the car tax with a 2.7% rate.

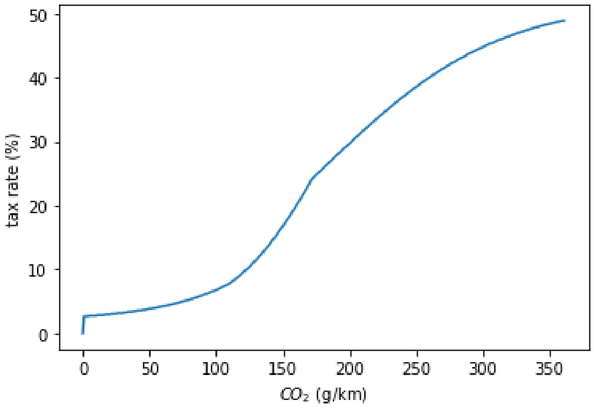


Figure 4.4: Tax rates of Finnish registration tax depending on the level of CO_2 emissions (WLTP). Tax rates have applied since October 2021. Data extracted from [140].

4.1.8.1.4 Subventions

Purchase incentives have been given only for BEV since 2018. An initial plan ended in November 2021, and in 2022 it was extended until March 2023 [21]. The conditions for receiving the subsidy are:

- It is a fully electric passenger car
- The total price of the car considering taxes is not higher than 50000 €
- The car is new and unregistered

Given the following conditions, the subsidy of 2000€ can be obtained.

4.1.8.2 Taxes on ownership

Ownership tax is called vehicle tax. Vehicle tax is paid annually. Vehicle tax consists of a basic tax and a tax on driving power. The tax on driving power is imposed on vehicles that are powered by some other force or fuel than motor petrol [22].

4.1.8.2.1 Basic tax

The base tax is calculated based on the vehicle's carbon dioxide emission levels as reported by the manufacturer. The taxation is based on WLTP CO_2 values and the WLTP tax table since 1st January 2020. The yearly amount of the basic tax in € can be seen in Figure 4.5.

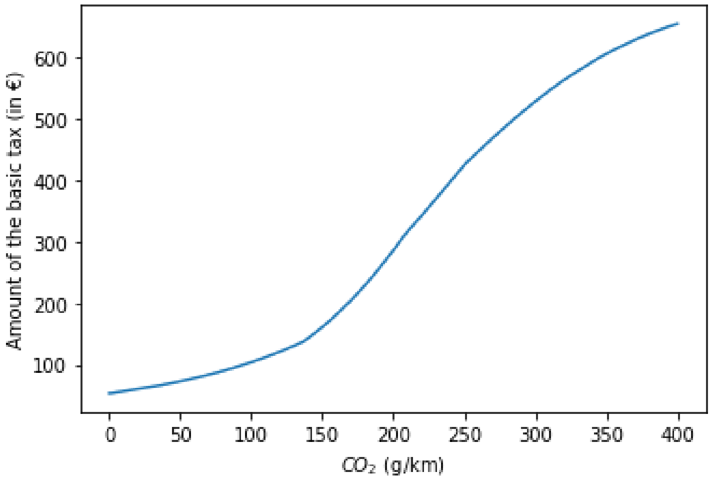


Figure 4.5: Yearly amount of basic tax depending on the level of CO_2 emissions (WLTP). The values have been valid since January 2020. Data extracted from [22]

4.1.8.2.2 Tax on driving power

In addition to base tax, tax on driving power is to be paid for passenger cars which use a fuel other than petrol.

The tax on driving power for passenger cars is normally set at 5.5 cents per day for each partial or 100 kilograms of total vehicle mass (e.g., diesel-powered vehicles). However, the tax level for particular power sources may be lower than this and can be observed in Table 4.21.

Powertrain	**cents per day per each 100 kg**	**€ per year per each 100 kg**
Diesel	5,5	20,08 €
Electricity	1,5	5,48 €
Electricity and petrol	0,5	1,83 €
Electricity and diesel	4,9	17,89 €
Methane	3,1	11,32 €

Table 4.21: Tax on driving power for different power sources. Data extracted from [22]

4.2 Fuel and energy prices

Other external factors besides the taxation influencing the customer when selecting a specific powertrain are the fuel and energy prices. The fuel and energy prices can be estimated under different scenarios, which determine the evolution of the fuel prices. This price depends to a large extent on the assumptions made in the scenarios.

In addition to taxation, energy prices significantly impact the development of the vehicle markets.

4.2.1 Scenario of IEA selected [144]

Fuel and energy price in the future is not specific. IEA's World Energy Outlook (WEO) explores various scenarios, each built on a different set of underlying assumptions about how the energy system might evolve. The price of electricity and fossil fuels will differ in the future depending on the scenario.

For the scope of this thesis, only one scenario is consulted. The scenarios proposed by IEA can be consulted in Appendix 8.8.1. The scenario selected is the Stated Policies Scenario (STEPS). It reflects current policy settings based on specific policies in place, as well as those that governments worldwide have announced and defined. It provides a more conservative benchmark for the future because it does not take it for granted that governments will reach all announced goals.

This scenario considers the current policies in use and the future policies if they have been officially announced and declared. STEPS scenario is used in the thesis to see which would be the future European passenger car sales considering the current policy settings.

4.2.2 Gasoline and diesel prices

4.2.2.1 Components of gasoline and diesel price

There are two main components to determine the price of fuels:

- **price of fuel before taxes**: It is based on the crude oil price determined in the international market. For example, the Brent oil barrel. Additionally, the refining process and the transport and marketing are additional costs of the price of fuel before taxes. In European Countries, fuel price before taxes usually represents less than 50% of the retail price of the fuel [51]. Then, the price of fuel before taxes can be divided into:
    1. **price of crude oil**: It is defined in commodity markets. It has international values. IEA estimates the future price of Brent oil barrel under the different scenarios presented in 4.2.1.
    2. **price of refining**: This can differ for each country.
    3. **price of transportation and marketing**: This can differ for each country.
- **taxes**: In European Union, taxes represent the major part of fuel price. They are divided into:
    1. **fuel tax or excise tax**: Excise duties have a fixed price and remain unchanged independently of the fuel price. They are modified after some years or in some countries

yearly. Ukraine's war has drastically increased the price of fuels, and some countries have reduced the excise tax price for months. However, this has not happened before in the last 20 years.

2. **CO_2 tax**: This tax remains unchanged as well, independently of the price of crude oil. Most countries do not apply CO_2 tax for particulars; it is only applied to companies. Germany and Norway are countries that apply CO_2 tax as an additional component of the fuel tax.
3. **VAT**: The VAT is a tax that depends on the fuel price and is applied to the previously fixed taxes (fuel tax and CO_2 tax) mentioned here.

It is difficult to obtain the fuel price breakdown for each country and the specific cost of refining, transportation, and marketing. Then, an approach that relates the price of crude oil with the price of fuel before taxes is analyzed.

4.2.2.2 Correlation between crude oil and gasoline and diesel price before taxes

A close correlation can be established between the oil price and the petrol station price for petrol and diesel in the US between 1999 and 2022 [75]. Therefore, the development of future fuel prices can be depicted in the model using a linear regression depending on the oil price.

Average yearly values of the price before taxes of gasoline and diesel have been obtained since 2005 from the Weekly Oil Bulletins presented by the European Commission [143]. The average yearly price of crude oil, specifically from a Brent barrel, is obtained from [35]. Norway's price before taxes is difficult to find. Then, it has been obtained the retail price of fuels in [108] and VAT, excise taxes and CO_2 tax has been deduced [49].

Then, it is evaluated if a correlation between the EU-27 price of gasoline and each specific country that is part of the study presents a correlation with the Brent barrel. In Figure 4.6, the correlation between the price of gasoline in the EU and the price of the Brent barrel oil can be seen.

In Figure 4.7, the correlation between the price of diesel in the EU and the price of the Brent barrel oil is presented as well.

It can be observed a R^2 of 0.92 for gasoline and a R^2 of 0.97 for diesel. It is relatively high and shows a clear correlation between the price of gasoline and the price of Brent-barrel oil. The same comparison between the gasoline and diesel prices in all countries of the EU with respect to Brent barrel oil is made. Scatter plots of gasoline price with respect to Brent oil barrel price with linear trend line can be consulted in Appendix 8.8.2. Both gasoline and diesel prices present a high correlation with Brent barrel price (See Table 8.16 and Table 8.17 in Appendix 8.8.2). The lowest correlation is for the gasoline price in Germany, which has an R^2 of 0.78 and is already relatively high.

Once correlations are calculated, it is observed that the estimation done is not good enough because data seems to be time dependant (See Figure 8.17 in Appendix 8.8.2.1). From 2005 to 2011, it overestimates the real value. From 2012 and on, the regression underestimates values of the fuel price before taxes.

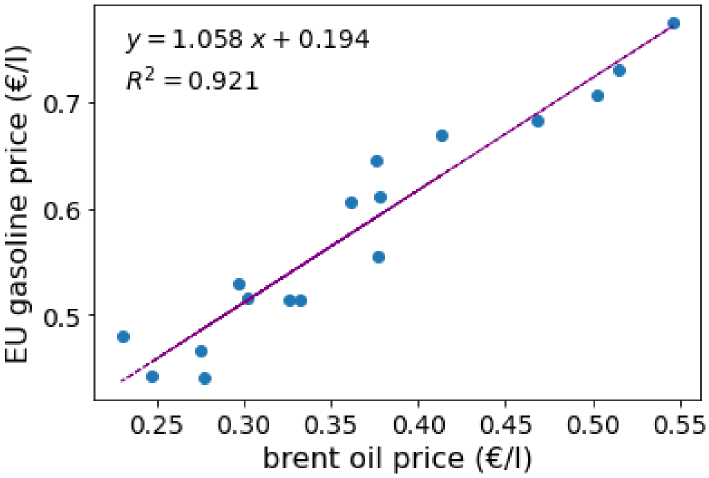


Figure 4.6: Correlation between the average price of gasoline before taxes in European Union and Brent barrel oil price between 2005 and 2021

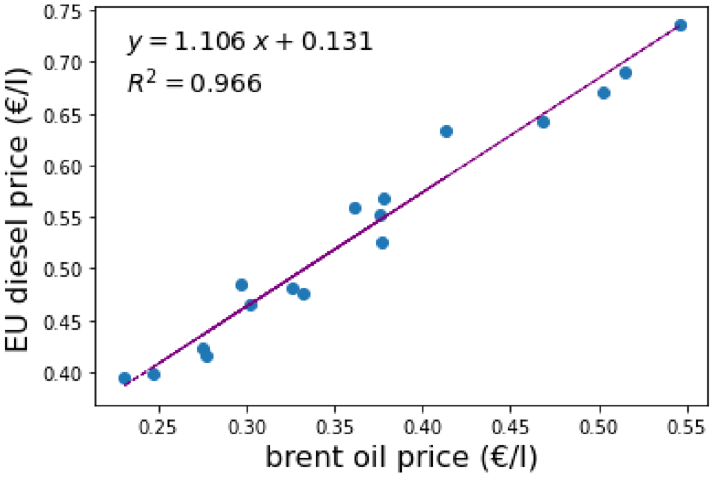


Figure 4.7: Correlation between the average price of diesel before taxes in European Union and Brent barrel oil price between 2005 and 2021

4.2.2.2.1 Chow test and ARIMAX model

Chow test

It is decided to perform a Chow test to observe if a structural change in the data can be confirmed. Two separate data groups were selected, one from 2005 to 2011 and the other from 2012 to 2021. The model has 17 observations (N=17) and only one exogenous variable. (k=2). Then, the Chow test hypothesis is:

- Null hypothesis H_0: Structural permanence
- Alternative hypothesis H_a: No structural permanence or structural change

The Chow test evaluates the sum of squared errors (SSE) of the whole regression data and the sum of squared errors for each of the subgroups defined. The test statistic for performing a Chow test with two subgroups is:

$$F_0 = \frac{\frac{SCE_T-(SCE_1-SCE_2)}{k}}{\frac{SSE_1+SSE_2}{N-2k}} \tag{4.4}$$

The statistical test follows a Snedecor's F-distribution with k in the numerator and N-2k in the denominator. Then, $\mathrm{F}_{critical}$ is $F_{2,13,0.99}$= 6.701.

If the statistical test F_0 is higher than $F_{2,13,0.99}$, there is no structural permanence.

Country	F_0
Finland	27,8505463
France	50,17625
Germany	24,6362167
Greece	26,6222029
Italy	28,3571564
Netherlands	10,4884462
Poland	25,675379

Table 4.22: Statistical test of Chow test for gasoline and crude oil splitting data into two-time subgroups. First, from 2005 to 2011, and second, from 2012 to 2021

It can be concluded that there is no structural permanence with a confidence level of 99% in the following periods between crude oil and gasoline. Other periods are evaluated to see what happens: 2005 and 2010, 2005 and 2012, 2005 and 2013, or 2005 and 2014. In all cases, the F_0 obtained is higher than the $\mathrm{F}_{critical}$, but F_0 is lower than the selected period, showing that the optimum year has been selected.

For diesel, it cannot be seen graphically an increase in the price of crude oil over the years. However, to use the same procedure for both, a chow test is performed for the same two time periods.

Country	F_0
Finland	16,9573069
France	10,1274236
Germany	15,7743794
Greece	25,513424
Italy	4,73886429
Netherlands	8,67535659
Poland	11,2359506

Table 4.23: Statistical test of Chow test for diesel and crude oil splitting data into two-time subgroups. First, from 2005 to 2011, and second, from 2012 to 2021

It can be observed in Table 4.23 that F_0 values are much lower for diesel than for gasoline. With a confidence level of 99%, Italy will not reject the null hypothesis. However, with a confidence level of 95% $F_{2,13,0.95}$=3.804, it is rejected. Diesel presents in relative terms better results if data is split between 2005 and 2010 and 2011 and 2021.

However, to use the same data period for both, the same subgroups as for gasoline are considered.

Finally, the price of gasoline and diesel before taxes will be estimated using data from 2012 to 2021.

ARIMAX

Before applying a Chow test, it was evaluated if ARIMAX could capture the relationship between the price of gasoline in a certain year with the price of previous years. Auto ARIMA function was used in R language, which selects the ARIMAX model with the lowest AIC (For more information about the model, see [67] and [68]).

The results were not satisfactory at first and were not capable of explaining why the price ratio between fuel and crude oil was increasing through the years. It was obtained a model (0,1,0) for most countries which is a random walk. It had a very broad confidence interval when defining future values, and this model could be improved.

Then, a Chow test is done, and only data since 2012 is used. Then, an ARIMA model (0,0,0) is obtained for all countries. (0,0,0) means that errors are uncorrelated across time. Using the Chow test and ARIMA slightly reduces the global error through the years, but it is not able to capture the price spike of gasoline in 2021 (See Figure 4.24).

	ARIMA	**linear regression**
2012	-2%	0%
2013	0%	2%
2014	1%	3%
2015	-5%	-2%
2016	2%	5%
2017	0%	2%
2018	2%	4%
2019	3%	6%
2020	-1%	1%
2021	-16%	-14%

Table 4.24: Relative error of ARIMA estimation of the price of gasoline in Germany and a linear regression model with respect to the real price of gasoline.

In the case of diesel, the relative error of ARIMA is better than with linear regression, and the error in 2021 is for all countries smaller than 15% and for all countries except two smaller than 10%.

Once the ARIMA model has been implemented and fitted, the model can forecast the future values of gasoline and diesel considering the brent oil price. In Figure 4.8, it is represented the ARIMAX forecast of the gasoline price of France under the STATS scenario. In Figure 4.9, it is shown the ARIMAX forecast and the IEA forecast of the price of crude oil.

The price of natural gas only affects CNG sales which represents a minor part of total European sales. The CNG price calculation can be consulted in Appendix 8.8.3.

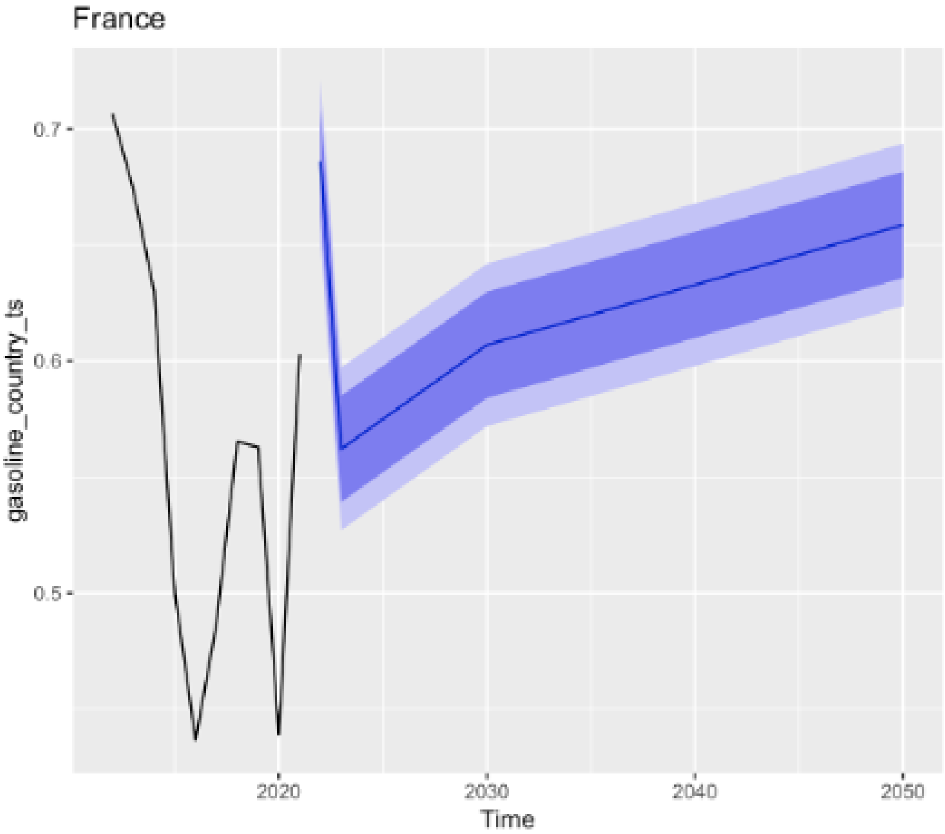


Figure 4.8: ARIMAX forecast of the price of Gasoline in France under STATS scenario. Dark blue represents 80% confidence interval and light blue represents 95% confidence interval.

4.2.3 Electricity prices

Electricity prices depend on the pricing systems used. There are four main pricing schemes in the global charging market, the last three of which are relevant for public charging [112]:

- **At home**: Drivers would pay according to their electricity contracts, so prices vary substantially.
- **Energy consumption pricing**: The users pay according to the amount of energy transferred into the battery. It is measured in €/kWh.

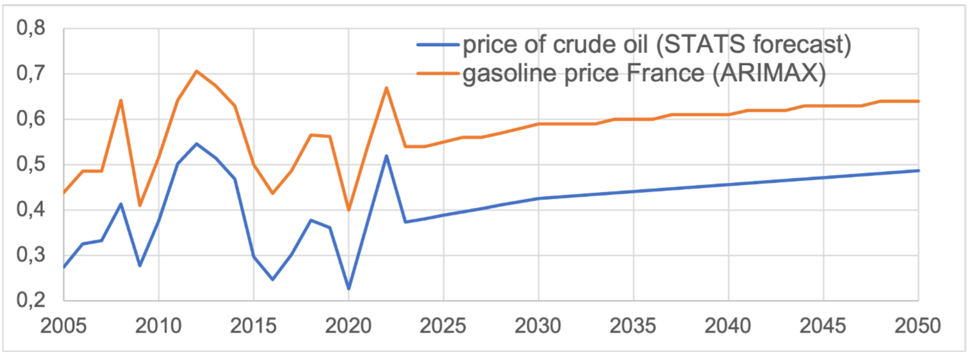


Figure 4.9: ARIMAX forecast of the price of Gasoline in France under STATS scenario (orange) and IEA forecast of crude oil price under STATS scenario

- **Time consumption pricing**: It is commonly associated with parking, where the user is charged according to the time spent with the plug-in, regardless of the amount of electricity transferred, measured in €/min.
- **Subscription based**: Users pay a fixed monthly rate and may be entitled to charge for free, or at a reduced rate, at a network of charging points.

As the number of charging points increases and charging service providers increase, tariffs become more complex. At the moment, the complexity makes it difficult to compare among countries public charging electricity prices.

Then, it has been selected to estimate electricity prices when charging at home. Electricity prices are presented in € per kWh. There are different prices per kWh when charging at home depending on the consumption band each household is (Consumption bands are, for example, between 1,000 KWh and 2500 Kwh; or between 2500 kWh and 5000 kWh).

In [69], it specified the price paid by a household based on a weighted average of the consumption band spectrum or based on the consumption band which is most frequently used. Taking this into account is estimated the average household electricity price before taxes for each country.

Values from 2011 to 2021 are taken into account, and the relative electricity price difference with respect EU average can be seen in Table 4.25.

The household average electricity price before-tax difference to the EU is assumed to stay constantly up until 2035. Once before tax price of electricity is defined, there are additional components that influence the price of electricity at a national level:

- **Network costs**: This has already been included in the before-tax electricity price difference.
- **Excise tax**: A fixed tax per kWh.
- **VAT**: VAT and excise tax have already been described for other fuels.

The taxes are assumed constant up to 2035. The estimation of household electricity prices before taxes are extracted from the central scenario of [122]. This is the most similar scenario found compared to the reference scenario of IEA.

Finland	-17%
France	-20%
Germany	10%
Greece	-13%
Italy	21%
Netherlands	8%
Norway	-45%
Poland	-13%

Table 4.25: Average household price of electricity difference respect EU average

With all these parameters, average electricity prices for EVs are calculated nationally.

4.2.4 Hydrogen prices

Hydrogen prices have been calculated for a previous project in DLR at a detailed level for Germany. Germany is expected to deploy infrastructure quicker than most other EU countries and to have a relatively low price. It has been assumed that all countries have the same prices as Germany.

This assumption is beneficial for the intake of FCEV at an EU level. However, it is observed that in any country, FCEV sales are present up to 2035 under the scenario studied. The main reasons are the still high purchase price and the low infrastructure deployment compared to other technologies such as BEV.

4.3 Income distribution

Another important demand-side factor when purchasing a vehicle is the customer's income. It is required to determine the purchasing power of a customer to determine if he or she will buy a vehicle and which category of vehicle most likely will be.

Before analyzing each country's income distribution, it is essential to know different ways to quantify income. It exists the national income and disposable income. The key difference between them is that national income is the total value of the total output of a country, including all goods and services produced in one year. In contrast, disposable income is the net income available to a household or an individual for spending, investing, and saving purposes after income taxes are paid. Disposable income only accounts for household incomes. An exact definition is given by OECD library [65] *"Disposable income can be seen as the maximum amount that a household can afford to spend on the consumption of goods or services without reducing its financial or non-financial assets or increasing its liabilities."*

After observing the differences between national income and disposable income, disposable income will be selected per household because it defines the real purchasing power of a customer when buying a car. The customer's decision is affected by the household situation.

The disposable income (in €) per household is calculated at a NUTS2 level. Total disposable income (Extracted from [73]) is divided by the number of households (Extracted from [97]). Household disposable income distribution can be obtained at a NUTS2 level.

The national average disposable income per habitant is obtained for the first decile, each quantile, and the last decile.(Extracted from [123]). The median income of each quintile is obtained as a % of the average disposable income of the total population (See in Appendix 8.9.1). The first decile, quintiles, and tenth decile are obtained only at a national level. The quintiles with respect to the total average income are quite similar between different countries. It is supposed that this income distribution at a national level remains the same at the NUTS2 and NUTS3 levels. The only difference considered at the NUTS2 and NUTS3 levels is the average income, which will change all regions' disposable income in absolute terms.

At this point, there is sufficient data for performing the analysis at a NUTS2 level.

Some additional hypotheses are required to obtain disposable income distribution at a NUTS3 level. The methodology proposed for obtaining income distribution at a NUTS3 level can be consulted in Appendix 8.9.4.

4.4 Vehicle size distribution

Vehicle size distribution is different by country and by type of region (urban, rural, or metropolitan).

The data source is Marklines which classifies the cars using segment classification by Euro NCAP. Marklines data does not provide information about the difference by type of region. Then, vehicle size distribution is used at a country level. Euro NCAP classifies cars with letters, representing cars of a specific size, weight, power, and retail price.

The vehicle size distribution is relatively constant and stable in time among countries. It has been used the average values from 2019 to 2021 as the vehicle size distribution of the vehicle fleet. Values have been obtained for each country and assumed constant from 2020 to 2035.

This letter classification has been converted into a more straightforward classification, the one used in V21.

In the V21 model, there are 8 types of powertrain and 3 vehicle sizes: small, medium, and large. This gives a total of 24 reference vehicles which the customers will be able to buy depending on the external conditions and their personal preferences.

It has been checked that the most representative vehicles of each Euro NCAP category have similar weight, size, power, and price to the equivalent V21 vehicle sizes. Then, it has been obtained the Table 4.26.

Using Table 4.26, vehicle size distribution is adapted to V21 format and directly integrated in the model.

Euro NCAP categories	denomination	v21 categories
A	basic	small
B	small	small
C	lower medium	medium
D	upper medium	medium
E	large	large
F	luxury	large
MPV		large
SUV-A		small
SUV-B		small
SUV-C		medium
SUV-D		medium
SUV-E		large
Pickup Truck		large

Table 4.26: Vehicle size car classification by Euro NCAP and in V21

4.5 Yearly mileage distribution

The annual mileage of cars differs in each country, and the demand varies by car segment. The following parameters are considered to be relevant in the mileage distribution of each country:

- Country: Each country presents different geographical and socioeconomic conditions. Then, it is the first relevant factor for defining yearly mileage distribution.
- Car segment: Small, medium, and large present different yearly mileage distributions. Usually, small vehicles drive fewer kilometers than medium vehicles, and medium vehicles tend to drive less than large ones.
- NUTS3: NUTS3 defines small territorial units as metropolitan, urban or rural. In rural areas, typically, higher yearly mileage is present.
- Powertrain: Typically, large diesel cars are used for customers with a high yearly mileage demand.

4.5.1 Limitations of data

4.5.1.1 Measurement of yearly mileage

Average yearly mileage can be calculated by odometer readings of the road authorities, surveys, or other methods. The procedure selected will vary the final results when comparing countries. Thus, it has been selected data that was calculated using a similar procedure. The procedure selected has been odometer reading. Even using odometer reading, the process of calculation can make results differ. However, there is no database where all countries are available. Then, different data sources are used with the condition that they use odometer readings. The different country sources and their characteristics can be consulted in Appendix 8.10.1.

Once sources of average yearly mileage data are obtained, it is shown in Table 4.27, a summary of the sources and the relative error when it is possible to compare values for the same country and the same year.

Country	Source used	Source of comparison	Relative error between sources
Finland	National Statistics Office	Enerdata	Less than 2% (2018)
France	National Statistics Office	Enerdata	Less than 1%
Netherlands	National Statistics Office	Enerdata	Less than 1%
Norway	National Statistics Office		No comparison available
Italy	Enerdata		No comparison available
Greece	Enerdata		No comparison available
Germany	Enerdata	OECD	Less than 3%
Poland	Enerdata	OECD	Less than 10%

Table 4.27: Sources used for average yearly mileage in 2019. All sources measure yearly mileage by odometer reading. The relative error between a source of comparison is shown when it is possible

From Table 4.27, it can be concluded that National Statistics Office and Enerdata give very similar results. The error is in 3 cases smaller than 2%. On the other hand, OECD differs more from Enerdata, at least a 3% and in the case of Poland, a 10%. Then, data has been selected from National Statistic Offices when possible, and when it has not been possible, data has been selected from Enerdata. Final data selected values can be consulted in Appendix 8.10.2.

4.5.1.2 Availability of yearly mileage data

Most countries do not have data more recent than 2019. Then, 2019 was selected as the reference year for all countries for obtaining yearly mileage.

The objective is to find yearly mileage distribution for each country. However, it was only possible to find yearly mileage distribution in 2019 for the Netherlands in an official source [78].

Additionally, other parameters considered relevant for the yearly mileage distribution as the car segment, NUTS3, and powertrain were not available in all countries. Some sources provided yearly mileage by powertrain, others by car segment, and others at a NUTS2 or NUTS3 level. However, data introduced in the model must be homogeneous for all countries.

Then, it has been used only the average yearly mileage as data.

4.5.2 Fitting yearly mileage distribution of Netherlands to a Probabilistic Distribution Function

It is required the probabilistic distribution when creating synthetic customers with Vector21. In Italy, the probability of having a customer with low yearly mileage (for example, 6.000 km per year) is much higher than in Finland.

At this point, the passenger car yearly mileage distribution is only available for the Netherlands.

The yearly mileage distribution in the Netherlands is given by the number of passenger cars that drive a certain amount of km. The intervals are created for every 5000 km. The values can be seen in Table 4.28.

yearly mileage	**passenger car distribution**
<=5,000 km	24,14%
5 001 - 10 000 km	24,41%
10 001 - 15 000 km	19,63%
15 001 - 20,000 km	12,78%
20 001 - 25 000 km	7,64%
25 001 - 30,000 km	4,45%
30 001 - 35,000 km	2,66%
35 001 - 40 000 km	1,60%
>=40 001 km	2,67%

Table 4.28: Yearly mileage distribution in 2019 in the Netherlands. Extracted from [78]

It is required yearly mileage distribution every 1,000 km until 60,000 km for using it in the V21 model.

The data in the Netherlands only goes up to 40,000 km, and the rest is grouped in only one group. It is estimated that the rate of decrease of the next categories until 60,000 km can fit a probabilistic distribution. It is obtained a rate of decrease of the passenger car intervals, around 1,66 before 40,000. At the tail, the distribution ends up disappearing. It is supposed to have a similar rate but avoid having a very long tail. Then, a rate of 1,5 is assumed. Then, the distribution goes up to 100% and can be seen in Table 4.29.

The distribution goes up to 60,000 km and has an average of 12,818 km (data of CBS), while the average of Enerdata is 12,851 km. The values of the distribution give a good result.

Once this estimation is done, a plot of the distribution supposing uniform distribution inside each quantile group (This is a hypothesis) is calculated. The histogram obtained can be seen in Figure 4.10.

4.5.2.1 Properties of the distribution

Certain properties of the distribution of data are calculated and presented in Table 4.30.

yearly mileage	**passenger car distribution**
<=5,000 km	24,14%
5001 - 10 000 km	24,41%
10001 - 15 000 km	19,63%
15001 - 20,000 km	12,78%
20001 - 25 000 km	7,64%
25001 - 30,000 km	4,45%
30001 - 35,000 km	2,66%
35001 - 40 000 km	1,60%
40001 - 45000 km	1,07%
45001 - 50000 km	0,71%
50001 - 55000 km	0,48%
55001 - 60000 km	0,42%

Table 4.29: Yearly mileage distribution in 2019 in the Netherlands assuming a progressive decrease. The distribution above 40,000 km is estimated.

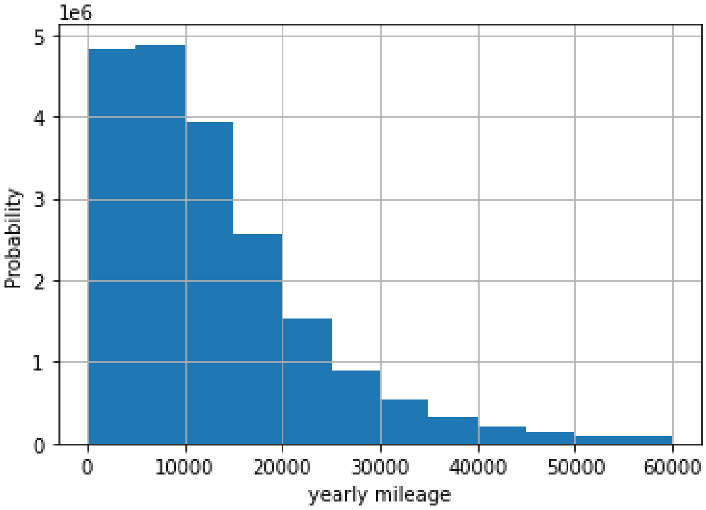


Figure 4.10: Histogram of yearly mileage in 2019 in the Netherlands

It obtained a mean of 12.819,9 km, which is the expected value considering the average yearly mileage given by [27] and a standard deviation bigger than 10.000 km showing the yearly mileage is spread. The skewness measures the asymmetry of the distribution about its mean. Skewness is 1.44. It is considerably positively skewed or right-tailed. The kurtosis is 2.49. It has a higher propensity to produce more outliers than if it was normally distributed.

skewness	1,44
kurtosis	2,49
mean	12819,9 km
standard deviation	10407,4 km
median	10368 km

Table 4.30: Properties of yearly mileage distribution data in 2019 in the Netherlands.

4.5.2.2 Distribution fitting

For fitting the data to a probabilistic distribution, different distributions are evaluated. Each distribution has parameters that modify its shape, and additionally, the distribution can be reallocated and rescaled to fit the data. In Appendix 8.10.3.2, a beta distribution is studied. Parameters are defined, properties are obtained, and reallocated and rescaled. The shape and the properties are more similar to the yearly mileage but still different.

Instead of evaluating distribution functions one by one, an auto-fitter that fits distributions automatically in an optimal way to the data is used. Before that, distribution functions that want to be evaluated must be selected. Criteria of selection of different distribution functions to be tested can be consulted in Appendix 8.10.3.3.

4.5.2.2.1 Auto-fitter

SciPy provides a method .fit() for every distribution object individually. A script for an automatic fitter procedure is used to set up a multi-model evaluation process.

First, a function for the sum squared error is called 80 times to compare the distance of a given candidate distribution from the actual data. Another statistic is calculated as well, Akaike information criterion (aic).

The distribution of yearly mileage data is created discretely. Each quartile representing 5,000 km has a uniform distribution with a certain probability (See Table 4.29). Only 10,000 points are used for representing the domain because fitting 80 functions take too much time if the resolution (the number of points) is increased.

The 5 functions with the smallest sum of squared errors concerning the yearly mileage of the Netherlands are shown in Figure 4.11.

In Figure 4.11 it is observed different functions which capture relatively good the path of the data. The biggest difference between all functions is between 0 and 10,000. Genhalflogistic and halflogistic cannot capture that between 0 and 5,000 km, there is more or less the same amount of drivers than between 5,000 and 10,000 km in the Netherlands. Then, these two functions can be discarded. Mielke is discarded because it can not be seen and gives problems when trying to fit it into the data. Burr and kappa 3 are capable of describing better the yearly mileage. After analyzing Figure 4.11, it is shown the value of statistics in Table 4.31.

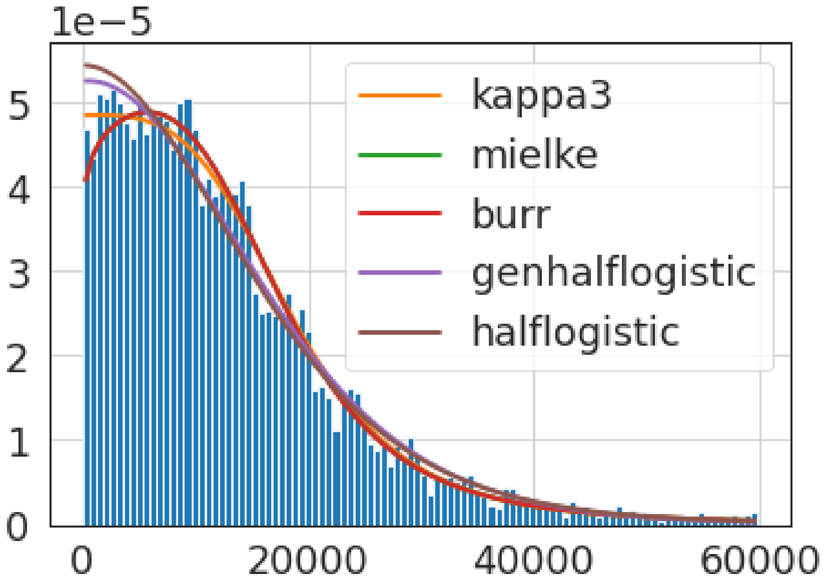


Figure 4.11: 5 distribution functions with the lowest sum squared errors

distribution	sum squared errors	aic
kappa3	4,48466E-10	2405,3
mielke	5,04612E-10	2407,0
burr	5,06126E-10	2407,0
genhalflogistic	6,79364E-10	2401,2
halflogistic	8,0899E-10	2392,2

Table 4.31: Sum squared errors and aic of 5 distributions with the lowest sum of squared errors

It can be obtained that kappa3 presents the lowest sum squared error and a better aic than burr; The lowest the aic, the better. However, the results are relatively similar.

Burr function is selected as the function to best fit yearly mileage distribution (See more details of the decision in Appendix 8.10.3.4).

4.5.2.2.2 Tuning selected distribution function

Once burr has been selected and parameters are obtained, the cumulative distribution function of the distribution function selected (burr) and the actual cumulative distribution of the yearly mileage data are studied.

Observing Figure 4.12, burr function is following very precisely the different quantiles every 5,000 km. It is analyzed the relative error for burr, kappa3, and gen halflogistic to compare how the function is fitted to data. (See Appendix 8.10.3.5)

The cumulative distribution function average error of burr is lower than 1%. This error is

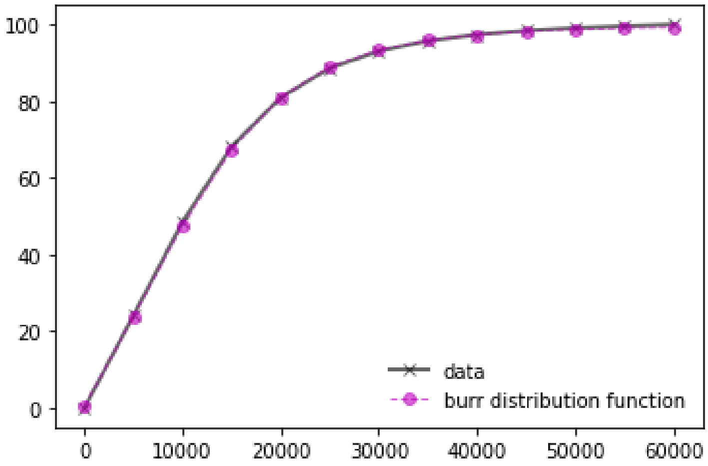


Figure 4.12: Cumulative distribution function of burr function and yearly mileage data

considered a good result when fitting the function. Netherlands yearly mileage distribution every 1,000 km can be obtained from this function.

4.5.2.2.3 Adjusting distribution function of Netherlands to other countries

Once the yearly mileage distribution function of the Netherlands has been fitted, it is considered that the distribution function remains the same for the yearly mileage of all countries. The burr function is used as the distribution function of all countries. Location and scale parameters can be used to shift or rescale the distribution to adapt it to the average yearly mileage of each country.

After trying both parameters, it is observed that using location is problematic for all countries which have a lower average yearly mileage than the Netherlands. The reason is that when shifting the function to the left, there is a limit at 0. Then, the cumulative distribution would start before 0, and this should not happen. On the other hand, if the function is rescaled, the cumulative distribution function obtained for all countries respects the 0 limits, and functions with a higher average are more spread in the mileage domain. The ones which have a lower average have a more narrow peak.

Each country's yearly mileage distribution is obtained by adjusting the burr distribution function only by modifying the scale parameter. Each distribution function is adjusted to have the average yearly mileage known in each country. The yearly mileage probabilistic distribution function of the different countries can be seen in Figure 4.13.

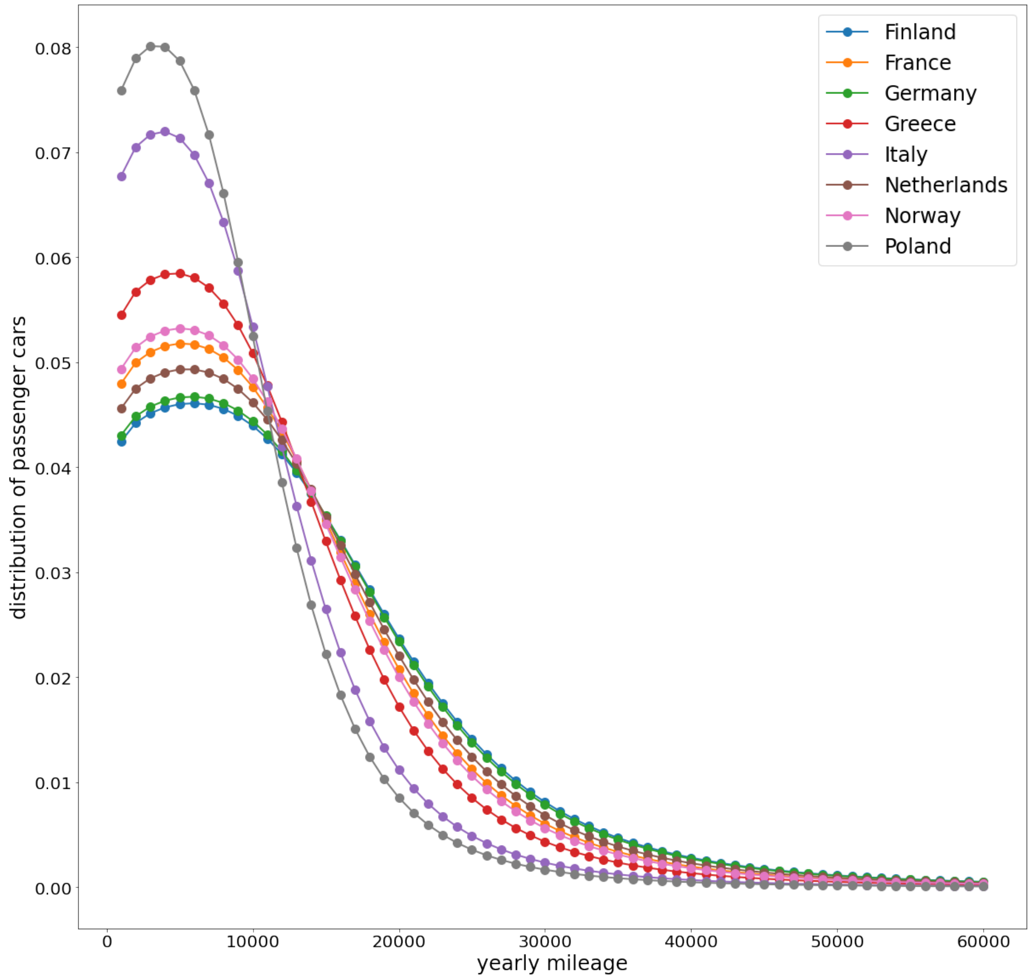


Figure 4.13: Yearly mileage probabilistic distribution function (PDF) of countries studied. PDF is obtained using burr and adjusting the distribution to the average yearly mileage of each country using scale parameter.

5 Results

Model structure is explained in Chapter 3 and inputs gathered are explained in Chapter 4. Results under the STATS scenario are presented here (See Section 4.2.1 for more details about the scenario)

5.1 Purchase price including taxes and incentives

One of the main contributions of this thesis has been modelling the taxation system of 8 different EU countries. Taxation has a major influence on alternative powertrain intake.

Taxation influences operating costs through ownership tax and purchase price through registration tax and VAT.

The purchase price, including taxes and incentives, is analyzed. It is divided into three components:

- Gross price before VAT: This is the price of the vehicles before registration taxes and VAT. The gross price before VAT of a specific vehicle through all EU has been assumed constant.
- Registration tax: It includes taxes and subsidies paid when purchasing a vehicle.
- VAT: Value-added tax is a fixed rate applied to the gross price before VAT plus the registration tax.

The purchase price, including taxes and incentives for the medium segment, is shown in Figure 5.1.

Different conclusion can be observed of Figure 5.1:

1. VAT is paid in all countries for gasoline vehicles and BEV except for Norway. In the medium segment, VAT exemption in Norway for BEV represents a reduction of more than 8,000€.
2. In 2022, BEV medium present subsidies in almost all countries going from around 2,000€ in Italy to 9,000€ in Germany. After 2025, subsidies are gone in all countries.
3. BEV medium does not pay registration tax in any country.
4. Registration tax for gasoline vehicles is very high in Finland, Netherlands, Norway, and Greece. On the other hand, there is Germany, where gasoline vehicles do not pay a registration tax.

Norway's case is characteristic. The combination of VAT exemption for BEV and high registration tax for gasoline vehicles make BEV cheaper than gasoline already in 2022. Netherlands and Finland have high registration taxes and subsidies, which make the prices of gasoline and BEV close to each other. Finally, the Greece case is surprising. In April 2022, taxation has been modelled, which is favorable for electric vehicles. Considering current taxation in Greece, infrastructure will be the main barrier to EV intake.

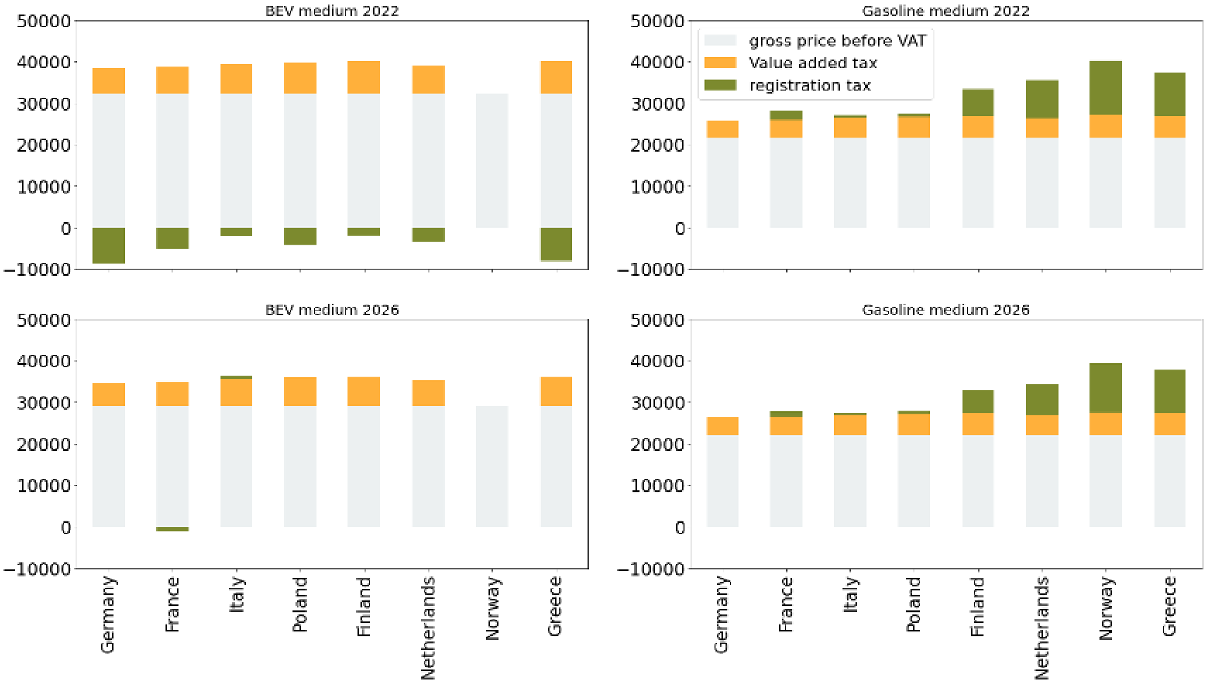


Figure 5.1: Purchase price including taxes and incentives for gasoline vehicles and battery electric vehicles of the medium segment in 2022 and 2026

The large segment has certain differences from the medium segment. The main differences can be summarized:

1. Gross price before VAT is similar in 2022 for BEV and gasoline. In 2026, the current scenario expects battery prices to decrease. The decrease will make the price of electric vehicles in the large segment can be equal to or even cheaper than gasoline vehicles in 2026.
2. There are no subsidies in large segment. Most countries do not give subsidies for vehicles with a gross price before VAT higher than 50,000€.
3. Registration tax is higher because it is normally based on parameters like CO2 emissions, power, and/or weight which are higher in this segment

Intake of electric vehicles in the large segment is expected to be quicker for the reasons explained. Figure of purchase price after taxes and incentives of the large segment can be consulted in Appendix 9.1.

The small segment taxation system and effects are very similar to the medium segment. The gross price before VAT is considerably higher for BEV than for gasoline vehicles. Then, taxation schemes substantially affect the final purchase price and the customer decision. Figure of purchase price after taxes and incentives of the small segment can be consulted in Appendix 9.1.

5.2 New passenger car sales up to 2035 under the STATS scenario

5.2.1 EU sales by type of powertrain

8 countries that represent 69.8% of the EU sales have been modelled in the V21. Each country represents a cluster. Then, sales per country are weighted, assuming the relative sales by type of powertrain and segment are constant inside each cluster. With this assumption, 100% of EU sales up to 2035 are modelled. EU new passenger car sales up to 2035 under the STATS scenario are presented in Figure 5.2.

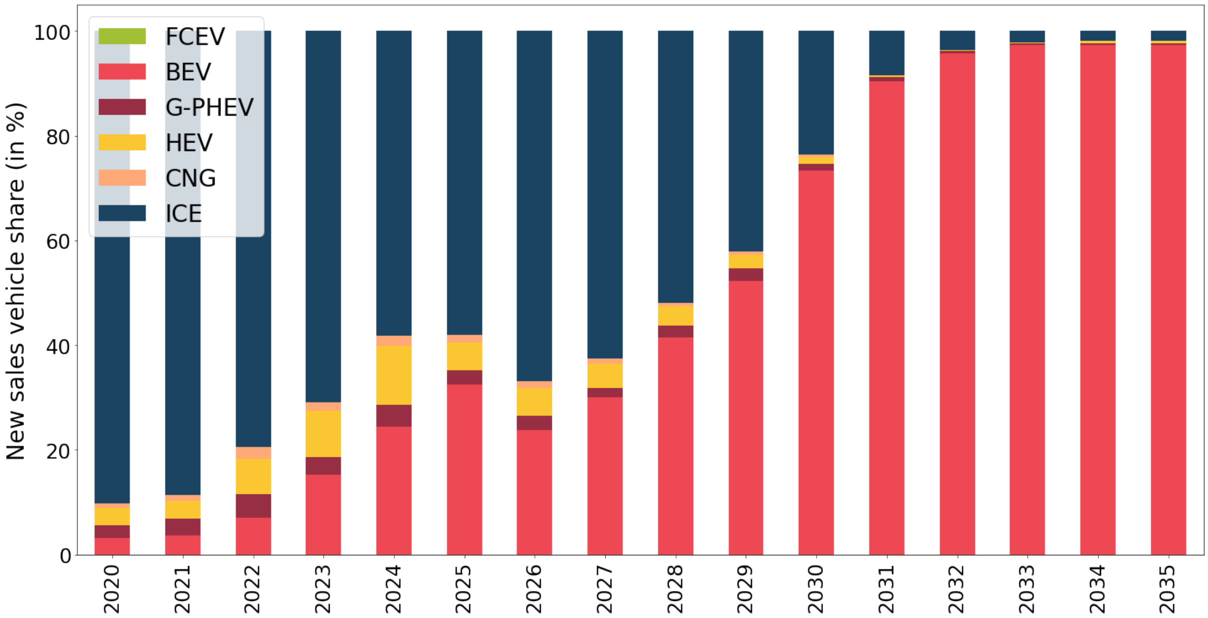


Figure 5.2: EU new passenger car sales up to 2035 under STATS scenario

Different conclusions can be extracted:

1. There are no FCEV sales
2. Subsidies are boosting the intake of electric vehicles up to 2025. EV sales decrease when incentives are phased out. After 2025, the relative share of EV is not recovered until 2028.
3. G-PHEV and HEV are temporary solutions for reaching CO_2 targets, but BEV is increasing its share considerably over the years.
4. Almost 100% of EV sales is already reached in 2032-2033.

Under the STATS scenario and the hypothesis assumed for infrastructure deployment (See Section 3), it is possible to transition to zero-emission new vehicle sales in 2035.

However, each country will follow a different path to reaching zero-emission vehicle sales in 2035. The electric vehicle is crucial in this transition, and a detailed analysis of EV sales at a national level is performed.

5.2.2 New EV sales by country

EV sales are the main contributor to the transition to zero-emission vehicle sales in 2035 under the STATS scenario. However, each country has different socioeconomic conditions, infrastructure, taxation schemes, and energy prices.

The current situation shows different EV intakes in each EU country. Each country has a different situation and will follow different strategies to reach the goal.

Figure 5.3 shows the EV intake from 2020 to 2035 at the selected countries.

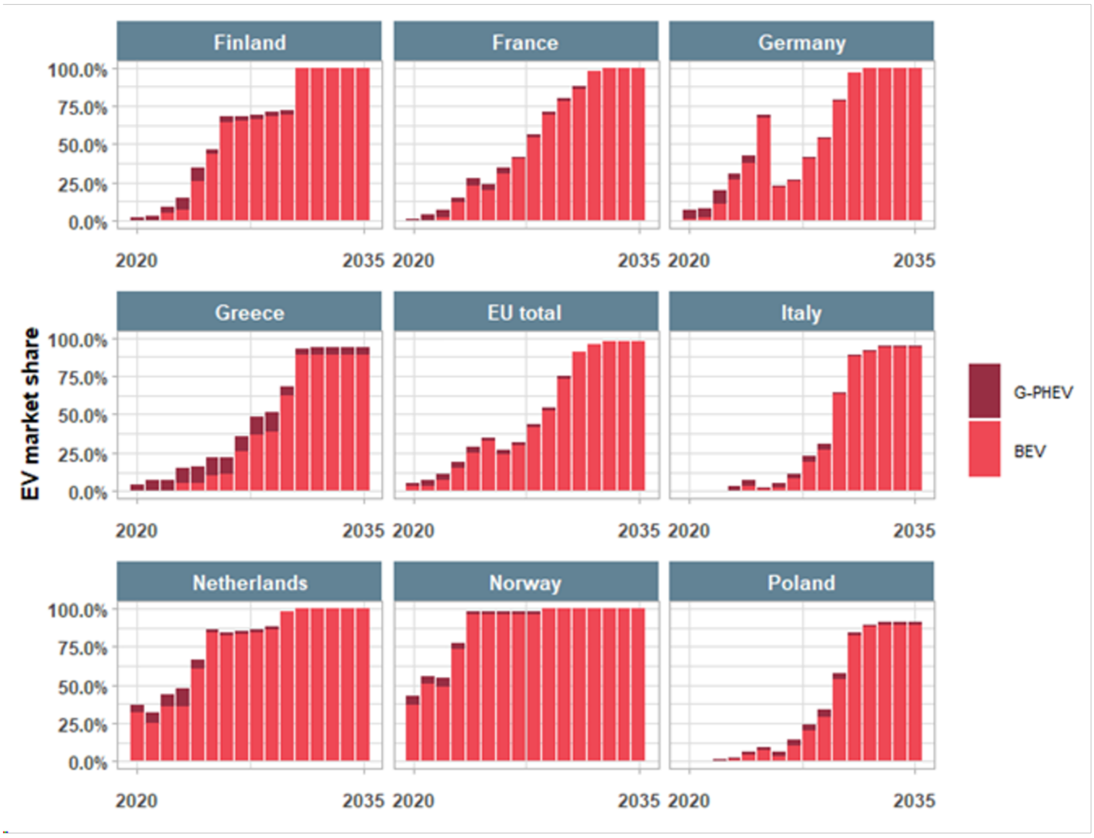


Figure 5.3: Electric vehicle new passenger car sales for different countries up to 2035 under STATS scenario

It can be observed that EV intake is different in each country.

- Norway and Netherlands are leading EV intake. Norway presents 97.9% EV sales in 2025, and the Netherlands presents EV sales higher than 80% in 2025.
- Finland is the following country with the highest EV intake. It presents almost a 50% EV intake in 2025. Then, the growth is sustained until 70% in 2030. France and Germany reach the same levels or even higher in 2030.
- Germany and France are part of the same cluster but present different EV intakes. The difference is mainly due to the taxation scheme. Germany promotes EV intake through a 9,000€ subsidy which lasts until 2025. In 2025, EV sales reach a maximum of 68.8%.

Once incentives are gone, EV sales do not reach this maximum again until 2030. In the case of France, the EV intake is slower but constant. The taxation scheme is mainly based on a high registration tax for ICE cars, not subsidies. From 2026 to 2030, EV sales are higher in France than in Germany.

- Italy presents a very low EV intake. Italy's results are slightly better than Poland's. In both cases, EV intake is lower than 10% until 2027. In 2029, EV sales are slightly higher than 30%. In 2030, around 60%. Italy reaches almost 95% EV sales in 2035 while Poland reaches 91%. The main reasons for the low EV intake are the slow infrastructure deployment and the taxation schemes. Additionally, household income is low in these two countries.
- Greece case is characteristic and unexpected. Taxation schemes were changed in April of 2022. The taxation is favorable for EV vehicles, which explains the higher intake compared to Italy and Poland. Infrastructure is assumed to be as slow as in Poland. In 2025, more than 20% of sales are electric, in 2028, almost 50%, and 2031, more than 93%.

It can be concluded that the European passenger car market is entirely different at a national level. Norway and Netherlands are above 80% EV sales in 2025, while Italy presents a 2% of EV sales in 2025.

Taxation has a strong influence on EV intake. Germany is one of the market leaders in the transition to electric vehicles. However, if EV incentives are phased out in 2025 as planned in Germany, EV sales are expected to reduce significantly. Then, Germany might lose its leadership position.

Infrastructure availability is a key parameter that hugely influences a slow BEV intake in Greece, Italy, and Poland. It can be observed the high G-PHEV share in Greece. Greece has favorable taxation, but the lack of infrastructure makes customers buy more G-PHEV than BEV until 2026.

6 Conclusion and further research

6.1 Conclusion

70% of the EU market (8 countries) has been modelled using the V21 utility maximization model. All EU market has been modelled using automotive market clusters under the STATS scenario. The first EU automotive market study that combines clustering with a utility model.

7 defined clusters have been obtained using hierarchical clustering in combination with principal component analysis. A relatively high Silhouette Score (0.4) has validated the model internally.

Using clustering is cost-wise due to the time required for gathering data. Gathering data for 8 countries has been difficult. Gathering data for 27 would be very slow, and there might be a lack of data for certain countries.

The purchase price after taxes and incentives has been the most important factor for the customer in selecting a vehicle. Modelling the taxation of each country has been a significant contribution to V21.

The EU 2035 target is possible with current policies and the necessary infrastructure deployment. Countries will follow different paths to reach 100% EV sales in 2035. Under the STATS scenario, it is recommended that certain countries change their taxation scheme. Germany, if it wants to keep its leadership position; Poland and Italy, if they want a smoother EV intake, should try to follow similar paths to Greece.

6.2 Further research

Further development of the STATS scenario is necessary. Infrastructure deployment should be included in detail. It has a significant influence on the purchase decision. Hydrogen prices should be studied individually for each country.

An analysis of new passenger car sales is necessary, and a stock model should be developed. These improvements can be made under V21 software. The stock model becomes relevant to observe how countries follow the path to having net-zero emission automotive fleets in 2050.

Other scenarios with different policy assumptions and with other technology cost development should be carried. For example, scenarios with different policy assumptions are Announced Pledged Scenarios of IEA (See Appendix 8.8.1). Scenarios with different technology cost development are scenarios where, for example, hydrogen cars are assumed to have quicker developments than expected and have a more competitive price.

Bibliography

[1] *12558: Households, by size of income. Highest value in decile, number and per cent (M) 2005 - 2020*. Mar. 1, 2022. URL: `https://www.ssb.no/en/statbank/table/12558/tableViewLayout1/` (visited on 08/08/2022).

[2] *2021 progress report: Making the transition to zero emission mobility*.

[3] *2050 long-term strategy*. Nov. 1, 2018. URL: `https://ec.europa.eu/clima/eu-action/climate-strategies-targets/2050-long-term-strategy_en` (visited on 04/04/2022).

[4] *Additional taxes on registration cards*. 2022. URL: `https://www.carte-grise.org/taxe_additionnelle.php` (visited on 04/22/2022).

[5] "The problem of multicollinearity". In: *Understanding regression analysis*. Ed. by Allen, Michael Patrick. New York: Plenum Press, 1997, pp. 176–180. ISBN: 978-0-306-45648-0. DOI: `10.1007/978-0-585-25657-3{\textunderscore}37`.

[6] *Amounts of the fine dust surcharge for motor vehicle tax from 2022*. URL: `https://www.belastingdienst.nl/wps/wcm/connect/nl/auto-en-vervoer/content/bedragen-fijnstoftoeslag-motorrijtuigenbelasting` (visited on 05/16/2022).

[7] *Annual circular for one-off registration tax 2022*. 2022. URL: `https://www.skatteetaten.no/rettskilder/type/rundskriv-retningslinjer-og-andre-rettskilder/avgiftsrundskriv/engangsavgift-2022/` (visited on 05/23/2022).

[8] *Annual detailed enterprise statistics for industry*. Feb. 28, 2022. URL: `https://ec.europa.eu/eurostat/databrowser/view/SBS_NA_IND_R2__custom_2643533/default/table?lang=en` (visited on 05/09/2022).

[9] *Annual motor vehicle tax and weight-based motor vehicle tax*. 2022. URL: `https://www.skatteetaten.no/en/person/duties/cars-and-other-vehicles/annual-motor-vehicle-tax-and-weight-based-motor-vehicle-tax/` (visited on 05/24/2022).

[10] *Annual report on transport in 2019*. Dec. 16, 2020. URL: `https://www.statistiques.developpement-durable.gouv.fr/bilan-annuel-des-transports-en-2019-0` (visited on 07/06/2022).

[11] *Annual road tax for private cars*. 2022. URL: `https://www.gov.gr/en/sdg/vehicles/buying-and-selling-motor-vehicle/road-tax/annual-road-tax-for-private-cars` (visited on 05/29/2022).

[12] Arellano, Manuel and Bond, Stephen. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations". In: *The Review of Economic Studies* 58.2 (1991), p. 277. ISSN: 00346527. DOI: `10.2307/2297968`.

[13] *Assegnazione di contributi comunali per l'acquisto di auto elettriche - ottavo avviso, riapertura dei termini*. 2022. URL: `https://www.comune.laquila.it/index.php?id_oggetto=19&id_cat=0&id_doc=860&id_sez_ori=217&template_ori=12&&gtp=1` (visited on 05/16/2022).

[14] *Average annual population to calculate regional GDP data (thousand persons) by NUTS 3 regions*. Apr. 18, 2022. URL: `https://ec.europa.eu/eurostat/databrowser/view/NAMA_10R_3POPGDP$DEFAULTVIEW/default/table` (visited on 05/09/2022).

[15] *Background: NUTS 2021 Classification*. 2021. URL: `https://ec.europa.eu/eurostat/web/nuts/background` (visited on 05/11/2022).

[16] Beyer, Kevin. et al. *When Is "Nearest Neighbor" Meaningful?* Vol. 1540. 1997. URL: `https://www.researchgate.net/publication/2845566_When_Is_Nearest_Neighbor_Meaningful`.

[17] *Biofuels (RED and RED II)*. URL: `https://energy.ec.europa.eu/topics/renewable-energy/biofuels/biofuels_en` (visited on 04/05/2022).

[18] *Budget Law 2022*. 2022. URL: `https://www.mef.gov.it/focus/Legge-di-Bilancio-2022/#cont4_scheda` (visited on 04/26/2022).

[19] Bugla, Katarzyna. "Electric cars 2021 - surcharges for electric cars". In: *superauto* (Nov. 9, 2020). URL: `https://www.superauto.pl/artykuly/elektryczne-samochody-2020-doplaty-do-samochodow-elektrycznych` (visited on 05/25/2022).

[20] "Bulgarians are Driving the Oldest Cars in Europe". In: *Sofia News Agency* (Oct. 13, 2017). URL: `https://www.novinite.com/articles/184252/Bulgarians+are+Driving+the+Oldest+Cars+in+Europe%20-%20https://www.iene.eu/bulgarians-are-driving-the-oldest-cars-in-europe-p3744.html#:~:text=The%20average%20age%20of%20cars,to%20data%20from%20the%20association.` (visited on 03/22/2022).

[21] *Buying a new all-electric car? You can apply for a procurement subsidy of EUR 2,000 from Traficom*. June 21, 2022. URL: `https://www.traficom.fi/fi/ajankohtaista/ostamassa-uutta-tayssahkoautoa-2-000-euron-hankintatukea-voi-hakea-traficomista` (visited on 01/10/2022).

[22] *Buying a new all-electric car? You can apply for a procurement subsidy of EUR 2,000 from Traficom*. June 21, 2022. URL: `https://www.traficom.fi/en/transport/road/structure-and-quantity-vehicle-tax` (visited on 2022).

[23] *Calculate motor vehicle tax*. 2022. URL: `https://www.belastingdienst.nl/wps/wcm/connect/nl/auto-en-vervoer/content/hulpmiddel-motorrijtuigenbelasting-berekenen` (visited on 05/16/2022).

[24] "Car incentives 2022, how the new Ecobonus works until 2024". In: *newsauto* (Apr. 12, 2022). URL: `https://www.newsauto.it/notizie/incentivi-auto-moto-furgoni-ecobonus-come-funzionano-2022-348499/` (visited on 04/25/2022).

[25] *Car incentives 2022: the news of the Budget Law*. Apr. 18, 2022. URL: `https://www.automobile.it/magazine/acquisto-auto/incentivi-auto-12762#ecotassa` (visited on 04/25/2022).

[26] *Car taxation – when you buy or bring into Finland a car not earlier registered in Finland*. June 21, 2022. URL: `https://www.vero.fi/en/individuals/vehicles/car_tax/` (visited on 2022).

[27] *Change in distance travelled by car*. Odyssee-Mure, 2020. URL: `https://www.odyssee-mure.eu/publications/efficiency-by-sector/transport/distance-travelled-by-car.html` (visited on 07/06/2022).

[28] Charu C. Aggarwal, Alexander Hinneburg, and Daniel A. Keim. *On the Surprising Behavior of Distance Metric in High-Dimensional Space*. 2002. URL: `https://www.researchgate.net/publication/30013021_On_the_Surprising_Behavior_of_Distance_Metric_in_High-Dimensional_Space`.

[29] Chipeva, N. "The average age of cars in Sofia is 19 years, their number is nearly a million – Bulgaria". In: *Darik.News* (). URL: `https://darik.news/en/the-average-age-of-cars-in-sofia-is-19-years-their-number-is-nearly-a-million-bulgaria-2.html` (visited on 03/22/2022).

[30] *CO_2 emission performance standards for cars and vans*. 2022. URL: `https://ec.europa.eu/clima/eu-action/transport-emissions/road-transport-reducing-co2-emissions-vehicles/co2-emission-performance-standards-cars-and-vans_en` (visited on 03/29/2022).

[31] *Consolidated text of motor vehicle tax laws. Article 20 of Presidential Decree of 5 February 1953, n. 39*. 1953. URL: `https://www.normattiva.it/uri-res/N2Ls?urn:nir:presidente.repubblica:decreto:1953-02-05;39`.

[32] *Corporate Electric transport and your company*. 2022. URL: `https://business.gov.nl/running-your-business/environmental-impact/making-your-business-sustainable/corporate-electric-transport-and-your-company/#:~:text=Fully%20electric%20cars%20are%20exempt,and%20motorcycle%20tax%20(Bpm).` (visited on 05/17/2022).

[33] *Costas Skrekas: 30% increase in the subsidy in the second cycle of "I move Electricly*. Feb. 2, 2022. URL: `https://government.gov.gr/kostas-skrekas-30-afxisi-tis-epidotisis-ston-deftero-kiklo-tou-kinoume-ilektrika/` (visited on 05/30/2022).

[34] *Costs for registering with the Pubblico Registro Automobilistico (PRA)*. URL: `https://www.aci.it/i-servizi/guide-utili/guida-pratiche-auto/costi-per-la-registrazione-al-pra.html` (visited on 04/25/2022).

[35] *Crude Oil Prices: Brent - Europe*. Aug. 1, 2022. URL: `https://fred.stlouisfed.org/series/MCOILBRENTEU` (visited on 08/12/2022).

[36] *Customer prices for registration*. Jan. 3, 2022. URL: `https://www.traficom.fi/en/transport/road/customer-prices-registration` (visited on 08/03/2022).

[37] Das, Udaibir S., Papaioannou, Michael G., and Trebesch, Christoph. "Sovereign Default Risk and Private Sector Access to Capital in Emerging Markets". In: *IMF Working*

Papers 10.10 (2010), p. 1. ISSN: 1018-5941. DOI: `10.5089/9781451961942.001`.

[38] *Decreto 27 novembre 1998, n. 435*. URL: `https://www.aci.it/i-servizi/guide-utili/guida-pratiche-auto/decreto-27-novembre-1998-n-435.html` (visited on 04/25/2022).

[39] Demeulenaere, Xavier. "The use of automotive fleets to support the diffusion of Alternative Fuel Vehicles: A Rapid Evidence Assessment of barriers and decision mechanisms". In: *Research in Transportation Economics* 76 (2019), p. 100738. ISSN: 0739-8859. DOI: `10.1016/j.retrec.2019.100738`.

[40] Díaz, S. et al. "European vehicle market statistics 2021/2022". In: (2021).

[41] *Digest of UK Energy Statistics. Annual data for UK, 2020*. July 29, 2021. URL: `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1060151/DUKES_2021_Chapters_1_to_7.pdf` (visited on 03/22/2022).

[42] Ebadian, M., McMillan, J., and Saddler, J. *Commercializing Conventional and Advanced Transport Biofuels from Biomass and Other Renewable Feedstocks*. Dec. 2021. URL: `https://task39.sites.olt.ubc.ca/files/2021/12/Final-Draft-IEA-Bioenergy-Task-39-Newsletter-Issue-58.pdf` (visited on 03/22/2022).

[43] *Ecological malus: all about the ecotax*. 2022. URL: `https://www.carte-grise.org/ecotaxe.htm` (visited on 04/21/2022).

[44] *Ecotax 2021: all the news and penalized cars*. Apr. 26, 2022. URL: `https://www.automobile.it/magazine/in-primo-piano/ecotassa-12521`.

[45] "Ecotax 2022 for cars starting from 160 g / km: concrete hypothesis". In: *sicurato* (Mar. 24, 2022). URL: `https://www.sicurauto.it/news/attualita-e-curiosita/ecotassa-2022-per-auto-a-partire-da-160-g-km-ipotesi-concreta/` (visited on 04/26/2022).

[46] *Effort sharing 2021-2030: targets and flexibilities*. July 14, 2021. URL: `https://ec.europa.eu/clima/eu-action/effort-sharing-member-states-emission-targets/effort-sharing-2021-2030-targets-and-flexibilities_en` (visited on 04/05/2022).

[47] *Estonia long term interest rate*. 2021. URL: `https://www.ceicdata.com/en/indicator/estonia/long-term-interest-rate#:~:text=Estonia%C2%BA%5C%20Long%5C%20Term%5C%20Interest%5C%20Rate%5C%3A%5C%20Month%5C%20Avg%5C%3A%5C%20Estonia%5C%3A%5C%20ECB,Jun%5C%202020%5C%20to%5C%20Jul%5C%202022.` (visited on 03/20/2022).

[48] European Parliament and European Council. *Regulation (EU) 2018/842 of the European Parliament and of the Council of 30 May 2018 on binding annual greenhouse gas emission reductions by Member States from 2021 to 2030 contributing to climate action to meet commitments under the Paris Agreement and amending Regulation*. May 30, 2018. URL: `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32018R0842` (visited on 04/04/2022).

[49] *Fees History 2020*. Jan. 1, 2020. URL: `https://www.skatteetaten.no/rettskilder/type/rundskriv-retningslinjer-og-andre-rettskilder/avgiftshistorie/` (visited on 08/10/2022).

[50] Figenbaum, Erik, Kolbenstvedt, Marika, and Elvebakk, Beate. *Electric vehicles - environmental, economic and practical aspects: As seen by current and potential users*. Vol. 1329/2014. TØI rapport. Oslo: Institute of Transport Economics, 2014. ISBN: 9788248015567.

[51] *Fuel price breakdown*. 2015. URL: `https://www.fuelseurope.eu/knowledge/refining-in-europe/economics-of-refining/fuel-price-breakdown/` (visited on 08/12/2022).

[52] *Future of Europe*. Sept. 2021. URL: `https://europa.eu/eurobarometer/surveys/detail/2554` (visited on 03/22/2022).

[53] *Gas prices for household consumers - bi-annual data (from 2007 onwards)*. Apr. 8, 2022. URL: `https://ec.europa.eu/eurostat/databrowser/view/NRG_PC_202/default/table?lang=en` (visited on 08/20/2022).

[54] *Gas prices for household consumers - bi-annual data (from 2007 onwards)*. URL: `https://ec.europa.eu/eurostat/databrowser/view/NRG_PC_202__custom_3256194/default/table` (visited on 08/22/2022).

[55] *GDP, current prices (US$ PPP): Purchasing power parity; billions of international dollars*. 2022. URL: `https://www.imf.org/external/datamapper/PPPGDP@WEO/ADVEC` (visited on 03/20/2022).

[56] *Global price of Natural gas, EU*. Aug. 16, 2022. URL: `https://fred.stlouisfed.org/series/PNGASEUUSDM#0` (visited on 08/18/2022).

[57] *Gray card: fiscal power rates in 2022*. Feb. 1, 2022. URL: `https://www.service-public.fr/particuliers/actualites/A14825`.

[58] *Greenhouse gas emission intensity of fuels and biofuels for road transport in Europe*. Nov. 18, 2021. URL: `https://www.eea.europa.eu/ims/greenhouse-gas-emission-intensity-of` (visited on 03/22/2022).

[59] Grein, A. F., Prakash Sethi, S., and Tatum, L. G. “A Dynamic Analysis of Country Clusters, the Role of Corruption, and Implications for Global Firms”. In: (2008). URL: `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.554.2275&rep=rep1&type=pdf` (visited on 03/24/2022).

[60] Grube, Thomas et al. “Passenger car cost development through 2050”. In: *Transportation Research Part D: Transport and Environment* 101 (2021), p. 103110. ISSN: 13619209. DOI: `10.1016/j.trd.2021.103110`.

[61] Grübler, Arnulf, Nakićenović, Nebojša, and Victor, David G. “Dynamics of energy technologies and global change”. In: *Energy Policy* 27.5 (1999), pp. 247–280. ISSN: 0301-4215. DOI: `10.1016/s0301-4215(98)00067-6`.

[62] Hagman, Jens et al. “Total cost of ownership and its potential implications for battery electric vehicle diffusion”. In: *Research in Transportation Business & Management* 18 (2016), pp. 11–17. ISSN: 2210-5395. DOI: `10.1016/j.rtbm.2016.01.003`. URL: `https://www.sciencedirect.com/science/article/pii/S2210539516000043`.

[63] Hand, D. J., Mannila, Heikki, and Smyth, Padhraic. *Principles of data mining*. Adaptive computation and machine learning. Cambridge Mass.: MIT Press, 2001. ISBN: 026208290X.

[64] *Hitting the EV inflection point*. May 2021. URL: `https://www.transportenvironment.org/wp-content/uploads/2021/08/2021_05_05_Electric_vehicle_price_parity_and_adoption_in_Europe_Final.pdf` (visited on 04/28/2022).

[65] *Household disposable income: Income and savings*. URL: `https://www.oecd-ilibrary.org/docserver/factbook-2015-18-en.pdf?expires=1652256057&id=id&accname=guest&checksum=46218B5F9755543F3CA4332F9F348CF8` (visited on 05/11/2022).

[66] *How, when and how much to pay: "Superbollo" since 2012*. URL: `https://www.agenziaentrate.gov.it/portale/it/web/guest/schede/pagamenti/addizionale-erariale-sulle-tasse-automobilistiche-c.d.-superbollo-/come-quando-e-quanto-versare` (visited on 05/20/2022).

[67] Hyndman, R. and Athanasopoulos, G. *Forecasting: Principles and Practice*. 2019. URL: `https://otexts.com/fpp2/` (visited on 08/24/2022).

[68] Hyndman, R. and et al. *package 'forecast': Forecasting Functions for Time Series and Linear Models*. July 25, 2022. URL: `https://cran.r-project.org/web/packages/forecast/forecast.pdf` (visited on 08/24/2022).

[69] IEA. *Energy Prices and Taxes for OECD Countries*. 2020. URL: `https://www.oecd-ilibrary.org/energy/energy-prices-and-taxes-for-oecd-countries-2020_dbf6150b-en;jsessionid=TE4vF3qAMzZlC2F3LhQXWuCK92Sw5ScLb2V7Qorl.ip-10-240-5-166` (visited on 08/23/2022).

[70] *Immatricolazione auto: come funziona e quali sono le tempistiche (Car matriculation)*. Jan. 27, 2021. URL: `https://www.automobile.it/magazine/burocrazia/immatricolazione-auto-5315` (visited on 04/25/2022).

[71] *Immatriculation d'un véhicule propre : les exonérations de carte grise*. Apr. 9, 2021. URL: `https://guichetcartegrise.com/blog/immatriculation-dun-vehicule-propre-les-exonerations-de-carte-grise`.

[72] *Incentives: the decree signed* . Apr. 6, 2022. URL: `https://www.lautomobile.aci.it/articoli/2022/04/06/incentivi-firmato-il-decreto.html` (visited on 04/26/2022).

[73] *Income of households by NUTS 2 regions*. Apr. 18, 2022. URL: `https://ec.europa.eu/eurostat/databrowser/view/NAMA_10R_2HHINC__custom_3165589/default/table` (visited on 08/08/2022).

[74] *Interest Rate | Europe*. 2022. (Visited on 03/20/2022).

[75] *Is price gouging behind the rise in gas prices?* July 11, 2022. URL: `https://rsmus.com/insights/economics/is-price-gouging-behind-the-rise-in-gas-prices.html` (visited on 08/12/2022).

[76] *Italy: New Law Reduces Excise Taxes and VAT on Fuels to Ameliorate Financial Crisis Caused by War in Ukraine*. May 2, 2022. URL: `https://www.loc.gov/item/`

global-legal-monitor/2022-05-31/italy-new-law-reduces-excise-taxes-and-vat-on-fuels-to-ameliorate-financial-crisis-caused-by-war-in-ukraine/#:~:text=The%20rate%20of%20excise%20duty,transport%20is%20set%20at%205%25 (visited on 08/25/2022).

[77] Kemp, René and Soete, Luc. "The greening of technological progress". In: *Futures* 24.5 (1992), pp. 437–457. ISSN: 0016-3287. DOI: 10.1016/0016-3287(92)90015-8. URL: https://www.sciencedirect.com/science/article/pii/0016328792900158.

[78] *Kilometers and numbers of vans and passenger cars 2019*. June 16, 2022. URL: https://www.cbs.nl/nl-nl/maatwerk/2022/24/kilometers-en-aantallen-bestel-en-personenauto-s-2019.

[79] *Klimabarometer 2021 (Norway)*. URL: https://kantar.no/kantar-tns-innsikt/klimabarometeret-2021/ (visited on 03/22/2022).

[80] *Law 2238/1994: Article 16*. URL: https://www.taxheaven.gr/law/2238/1994 (visited on 05/30/2022).

[81] *Law 4111/2013: Article 44*. URL: https://www.taxheaven.gr/circulars/17539/arora-enhmerwmeno-foros-polyteloys-diabiwshs-n-4111-2013-analysh-dieykriniseis-paradeigmata (visited on 05/30/2022).

[82] Lever, Jake, Krzywinski, Martin, and Altman, Naomi. "Principal component analysis". In: *Nature Methods* 14.7 (2017), pp. 641–642. ISSN: 1548-7091. DOI: 10.1038/nmeth.4346. URL: https://www.nature.com/articles/nmeth.4346.

[83] Ligterink, N. E. "The fleet composition on the Dutch roads relevant for vehicle emissions". In: (2017). URL: https://repository.tno.nl/islandora/object/uuid%3Abdbdf01c-c3bc-4f2a-b66b-a44baaaa63ca (visited on 08/04/2022).

[84] *List of VAT rates*. Dec. 13, 2021. URL: https://www.podatki.gov.pl/en/value-added-tax/general-vat-rules-and-rates/list-of-vat-rates/ (visited on 05/25/2022).

[85] Maugeri, Andrea et al. "Applying a hierarchical clustering on principal components approach to identify different patterns of the SARS-CoV-2 epidemic across Italian regions". In: *Scientific Reports* 11.1 (2021), p. 7082. ISSN: 2045-2322. DOI: 10.1038/s41598-021-86703-3. URL: https://www.nature.com/articles/s41598-021-86703-3#Sec4.

[86] *Mean and median income by degree of urbanisation*. Apr. 29, 2022. URL: https://ec.europa.eu/eurostat/databrowser/view/ILC_DI17__custom_2659745/default/table (visited on 05/09/2022).

[87] Meissner, Hans Rudolf. *Elektromobilität: Szenarien von Beschäftigungsauswirkungen*. 2011. URL: http://www.fastev-berlin.org/igm_17-11-2010_hrm.pdf..

[88] *Misure per la stabilizzazione della finanza pubblica: LEGGE 27 dicembre 1997, n. 449*. 1997. URL: `https://www.federmetano.it/wp-content/uploads/2018/01/Legge-n.-449-del-27121997-Art.-17-comma-5.pdf`.

[89] *Mobilità elettrica*. 2022. URL: `https://www.provincia.bz.it/turismo-mobilita/mobilita/mobilita-sostenibile/mobilita-elettrica.asp` (visited on 05/16/2022).

[90] Mock, P. *Development of a scenario model to simulate future market shares and CO2*. Köln, 2010.

[91] *Motor Vehicle Tax Act 1994*. URL: `https://wetten.overheid.nl/BWBR0006324/2022-01-01/#HoofdstukIII` (visited on 05/16/2022).

[92] *My electric program- Questions and Answers*. 2022. URL: `https://www.gov.pl/web/elektromobilnosc/program-moj-elektryk--pytania-i-odpowiedzi` (visited on 05/25/2022).

[93] *Natural gas price statistics*. URL: `https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Natural_gas_price_statistics` (visited on 04/01/2022).

[94] *New registrations of electric vehicles in Europe*. Nov. 18, 2021. URL: `https://www.eea.europa.eu/ims/new-registrations-of-electric-vehicles` (visited on 03/22/2022).

[95] *Norway's National Plan: related to the Decision of the EEA Joint Committee No. 269/2019 of 25 October 2019*. Dec. 2019. URL: `https://www.regjeringen.no/contentassets/4e0b25a4c30140cfb14a40f54e7622c8/national-plan-2030_version19_desember.pdf` (visited on 03/22/2022).

[96] Norwell, Frazer. "Norway to remove VAT exemption for electric cars". In: *The Local* (May 12, 2022). URL: `https://www.thelocal.no/20220512/norway-to-remove-vat-exemption-for-electric-cars/` (visited on 05/24/2022).

[97] *Number of households by degree of urbanisation and NUTS 2 regions*. July 4, 2022. URL: `http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfst_r_lfsd2hh&lang=en` (visited on 08/08/2022).

[98] *NUTS 3 rural development*. Eurostat, 2021. URL: `https://ec.europa.eu/eurostat/web/rural-development/background` (visited on 05/11/2022).

[99] *NUTS3: Regions and cities — Overview*. 2021. URL: `https://ec.europa.eu/eurostat/web/regions-and-cities/overview` (visited on 05/11/2022).

[100] Offer, G. J. et al. "Techno-economic and behavioural analysis of battery electric, hydrogen fuel cell and hybrid vehicles in a future sustainable road transport system in the UK". In: *Energy Policy* 39.4 (2011), pp. 1939–1950. ISSN: 0301-4215. DOI: `10.1016/j.enpol.2011.01.006`.

[101] *Ownership tax guide*. URL: `https://www.aci.it/i-servizi/guide-utili/guida-al-bollo-auto.html` (visited on 04/26/2022).

[102] Ozyurt, Beste, Dikmen, Irem, and Birgonul, M. Talat. "Clustering of host countries to facilitate learning between similar international construction markets". In: *Engineering,*

Construction and Architectural Management 27.1 (2020), pp. 66–82. ISSN: 0969-9988. DOI: `10.1108/ECAM-11-2018-0495`. URL: `https://www.emerald.com/insight/content/doi/10.1108/ECAM-11-2018-0495/full/pdf` (visited on 03/24/2022).

[103] *Passenger vehicles: REGISTRATION LEVY RATES*. 2022. URL: `https://www.gov.gr/en/sdg/vehicles/buying-and-selling-motor-vehicle/vehicle-registration-levy/passenger-vehicles` (visited on 05/29/2022).

[104] *Percentages of increase of the IPT*. URL: `2022-04-25`.

[105] Philip, Johan. "The probability distribution of the distance between two random points in a box". In: (2007). URL: `https://www.researchgate.net/profile/shubham-sharma-90/post/pdf-of-distance-between-two-random-points/attachment/59d6420279197b807799dd79/as%3a437282965594114%401481267791552/download/habc.pdf`.

[106] Picarelli, Mattia Osvaldo, Vanlaer, Willem, and Marneffe, Wim. "Does Public Debt Produce a Crowding out Effect for Public Investment in the EU?" In: *SSRN Electronic Journal* (2019). DOI: `10.2139/ssrn.3376471`.

[107] Poliscanova, Julia. *Plug-in hybrids in new emissions scandal as tests show higher pollution than claimed*. Nov. 23, 2020. URL: `https://www.transportenvironment.org/discover/plug-hybrids-new-emissions-scandal-tests-show-higher-pollution-claimed/` (visited on 05/17/2022).

[108] *Prices on engine fuel (NOK per litres)*. 2022. URL: `https://www.ssb.no/en/statbank/table/09654` (visited on 08/09/2022).

[109] *Progress towards national greenhouse gas emissions targets in Europe*. Oct. 26, 2021. URL: `https://www.eea.europa.eu/data-and-maps/indicators/progress-towards-national-greenhouse-gas/assessment` (visited on 03/22/2022).

[110] *Provinces; rates for motor vehicle tax surcharges*. Jan. 19, 2022. URL: `https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80889NED/table` (visited on 05/15/2022).

[111] *Quantile-Quantile Plots*. URL: `https://www.ucd.ie/ecomodel/Resources/QQplots_WebVersion.html#introduction` (visited on 07/06/2022).

[112] Rademaekers, K. et al. *Study on energy prices, costs and their impact on industry and households*. Oct. 2020. URL: `https://d1owejb4br3l12.cloudfront.net/sites/default/files/about-us/case-study/energy-prices-costs-eu-study.pdf` (visited on 09/07/2022).

[113] *Recovery plan for Europe*. July 2020. URL: `https://ec.europa.eu/info/strategy/recovery-plan-europe_en` (visited on 04/04/2022).

[114] Redelbach, M. et al. *Modelling customer choice and market development for future automotive powertrain technologies*. Nov. 17, 2013. (Visited on 09/01/2022).

[115] Région Autonome Vallée d'Aoste, ed. *Contributi mobilità sostenibile*. URL: `https://www.regione.vda.it/trasporti/Contributimobilitasostenibile/default_i.aspx` (visited on 05/16/2022).

[116] *Registration certificate and plates for passenger cars and motorcycles*. July 1, 2022. URL: `https://www.gov.gr/en/sdg/vehicles/buying-and-selling-motor-vehicle/general/registration-certificate-and-plates-for-passenger-cars-and-motorcycles` (visited on 08/03/2022).

[117] *Renewable energy directive*. 2009-09. URL: `https://energy.ec.europa.eu/topics/renewable-energy/renewable-energy-directive-targets-and-rules/renewable-energy-directive_en` (visited on 04/05/2022).

[118] *Road statistics on infrastructure and fleet data until 2021*. 2022. URL: `https://alternative-fuels-observatory.ec.europa.eu/transport-mode/road` (visited on 04/06/2022).

[119] *Road traffic volumes*. 2022. URL: `https://www.ssb.no/en/statbank/table/12577/` (visited on 07/06/2022).

[120] Rogers, Everett M. *Diffusion of innovations*. 5th ed. New York: Free Press, 2003. ISBN: 0743222091.

[121] Rousseeuw, Peter J. "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". In: *Journal of Computational and Applied Mathematics* 20 (1987), pp. 53–65. ISSN: 0377-0427. DOI: `10.1016/0377-0427(87)90125-7`. URL: `https://www.sciencedirect.com/science/article/pii/0377042787901257` (visited on 03/26/2022).

[122] Schmitt, A. *EU Energy Outlook 2050: How will the European electricity market develop over the next 30 years?* Apr. 11, 2022. URL: `https://blog.energybrainpool.com/en/eu-energy-outlook-2050-how-will-the-european-electricity-market-develop-over-the-next-30-years/` (visited on 08/20/2022).

[123] *Share of households and economic resources by income, consumption and wealth quantiles - experimental statistics*. June 30, 2021. (Visited on 05/11/2022).

[124] *Sheet 19: Rates and payments*. Mar. 30, 2021. URL: `https://www.aci.it/laci/la-federazione/amministrazione-trasparente/archiviofile/aci/Disposizioni_Generali/Atti_generali_Doc_Unico/SCHEDA_19_tariffe_e_pagamenti_Banner_540_2021%20%281%29.pdf` (visited on 04/25/2022).

[125] Shen, Heng Tao. "Dimensionality Reduction". In: *Encyclopedia of database systems*. Ed. by LIU, LING and ÖZSU, M. TAMER. Springer reference. New York: Springer, 2009, pp. 843–846. ISBN: 978-0-387-35544-3. DOI: `10.1007/978-0-387-39940-9{\textunderscore}551`.

[126] *Still fine growth in road traffic mileage*. Mar. 22, 2019. URL: `https://www.stat.fi/tietotrendit/artikkelit/2019/tieliikenteen-ajokilometreissa-edelleen-hienoista-kasvua/?listing=simple` (visited on 07/06/2022).

[127] *Storting resolution on excise duties for 2022: Taxes on motor vehicles*. 2022. URL: `https://lovdata.no/dokument/STV/forskrift/2021-12-14-3577/KAPITTEL_3-1#%5C%C2%5C%A72_2` (visited on 05/23/2022).

[128] *Subsidy Scheme Electric Passenger Cars for Private Individuals (SEPP)*. Apr. 7, 2022. URL: `https://www.rvo.nl/subsidies-financiering/sepp` (visited on 05/30/2022).

[129] Summers, Lawrence H. "Have we Entered an Age of Secular Stagnation? IMF Fourteenth Annual Research Conference in Honor of Stanley Fischer, Washington, DC". In: *IMF Economic Review* 63.1 (2015), pp. 277–280. ISSN: 2041-4161. DOI: `10.1057/imfer.2015.6`.

[130] *Survey on Household Income and Wealth - 2016*. Mar. 12, 2018. URL: `https://www.bancaditalia.it/pubblicazioni/indagine-famiglie/bil-fam2016/index.html` (visited on 05/09/2022).

[131] *Tax on company vehicles (TVS)*. Jan. 1, 2021. URL: `https://entreprendre.service-public.fr/vosdroits/F22203` (visited on 04/22/2022).

[132] *The amount of the ecological bonus and the conversion bonus are extended*. Apr. 28, 2022. URL: `https://www.service-public.fr/particuliers/actualites/A14391` (visited on 05/30/2022).

[133] *The most important issues facing the country (UK)*. Dec. 2021. URL: `https://yougov.co.uk/topics/politics/trackers/the-most-important-issues-facing-the-country` (visited on 03/22/2022).

[134] *The VAT Handbook 2022: 6-7 Vehicles*. 2022. URL: `https://www.skatteetaten.no/rettskilder/type/handboker/merverdiavgiftshandboken/gjeldende/M-6/M-6-7a/` (visited on 05/24/2022).

[135] *Total length of roads at 31 December by type of road (2019)*. URL: `https://w3.unece.org/PXWeb2015/pxweb/en/STAT/STAT__40-TRTRANS__11-TRINFRA/ZZZ_en_TRInfraRoad_r.px/` (visited on 03/22/2022).

[136] *Traffic performance*. 2020. URL: `https://www.aut.fi/tilastot/liikennesuoritteet` (visited on 07/06/2022).

[137] *Transport Statistics 2021 - Reference Year 2020*. URL: `https://nso.gov.mt/en/nso/Media/Salient-Points-of-Publications/Documents/2022/Transport%20Statistics%202021/transport-2021.pdf%3E` (visited on 03/28/2022).

[138] *Use of renewable energy for transport in Europe*. Mar. 10, 2022. URL: `https://www.eea.europa.eu/ims/use-of-renewable-energy-for` (visited on 04/05/2022).

[139] *Vehicle registration tax*. 2022. URL: `https://www.podatki.gov.pl/en/vehicles-in-eu/vehicle-registration-tax/` (visited on 05/25/2022).

[140] *Vehicle tax rates*. June 21, 2022. URL: `https://www.vero.fi/henkiloasiakkaat/auto/autoverotus/autoveron_maara/ajoneuvojen-veroprosentit/#henkiloauto` (visited on 2022).

[141] *Vehicles in use Europe 2022*. Jan. 2022. URL: `https://www.acea.auto/files/ACEA-report-vehicles-in-use-europe-2022.pdf` (visited on 03/22/2022).

[142] Wappelhorst, S. et al. *Analyzing policies to grow the electric vehicle market in European cities*. 2020. URL: `https://theicct.org/wp-content/uploads/2021/06/ev_city_policies_white_paper_fv_20200224.pdf`.

[143] *Weekly Oil Bulletin*. Energy section, 2022. URL: `https://energy.ec.europa.eu/data-and-analysis/weekly-oil-bulletin_en` (visited on 08/12/2022).

[144] *World Energy Outlook 2021*. Oct. 1, 2021. URL: `https://www.iea.org/reports/world-energy-outlook-2021` (visited on 08/09/2022).

[145] Yan, Shiyu. "The economic and environmental impacts of tax incentives for battery electric vehicles in Europe". In: *Energy Policy* 123 (2018), pp. 53–63. ISSN: 0301-4215. DOI: `10.1016/j.enpol.2018.08.032`. URL: `https://www.sciencedirect.com/science/article/pii/S0301421518305597`.

7 Appendix of Chapter 2

7.1 Sources of feature selection

The sources of data for the 11 features are presented below:

- **GDP per capita**: Data extracted from International monetary Fund. [55].
- **long-term interest rate**: Data extracted from Trading Economics for long term interest rate at the beginning of 2022 (between January and March) [74]. Estonia and Norway values were not available. Values for 1-year period between 2020 and 2021 are available in CEIC [47]. An average of the long interest rate during this period have been used.
- **greenhouse gas (GHG) intensity reduction of road transport - % of decrease between 2010 and 2019**: Data extracted from European Environment Agency for all countries except Norway. [58] Norway data has been extracted from Newsletter IEA Bioenergy [42].
- **Effort Sharing Regulation target for 2030**: Data extracted from European Environment Agency [109] for all countries except Norway and UK. Norway ESR target for 2030 is obtained from Norway's National Plan (October 2019) [95]. UK ESR target for 2030 is obtained from Regulation (EU) 2018/842 of the European Parliament [48].
- **Renewable Energy Source share in 2020**: Data extracted from European Environment Agency [138] and for the UK it has been extracted from Annual data statistics for UK (2020) done by the Department for Business, Energy & Industrial Strategy [41].
- **New registrations of electric cars**:Data extracted from European Environment Agency [94].
- **Average of vehicle age**: Data extracted from European Automobile Manufacturer's Association (ACEA) [141]. Cyprus data was incomplete and an approximation has been done (Calculus can be seen in Appendix 7.2.1). Malta's average age of cars can be obtained from Malta's National Statistics Office [137]. Bulgaria's average age of cars is not available by the government. In 2017 it was over 20 years according to Sofia News Agency [20] and according to [29] in 2020 average age of cars is 19 years. It is selected 19 years to be the average age of cars for whole Bulgaria, although it may be even older.
- **Automotive industry size**: Data extracted from Eurostat [8]. Ireland, Luxembourg and Malta values were estimated based on the size of the automotive industry respect to GDP compared to other EU countries where the employment was known.
- **Public electric chargers infrastructure**: Data of number of public chargers extracted from European Alternative Fuels Observatory (EAFO). Data extracted from year 2020. Data has slightly been modified for year 2020 in April 2022. Previous source is no longer available. Data can now be extracted from [118] for year 2020. Additionally, data for year 2021 has been included. In 2021, there has been a drastic increase in public chargers modifying the situation of 2020. It could be interesting to add how infrastructure is increasing in relative terms to capture certain evolution of the market within the cross-sectional model being used.

It has been adopted an approach were AC public chargers were considered and data has been collected for 2020 between normal chargers, which are catalogued as having less than 22 KW and faster chargers with higher power than 22 KW. EAFO is including AC chargers and separating between low, medium and fast charge. DC chargers data is available, but represents a much smaller quantity. In order not to have many variables it was selected the total number of chargers as a whole. Since april 2022 EAFO provides 3 defined different categories. Further work could be interesting to weight the availability of public chargers considering an average time of charging for each category. This would allow to capture better the availability of public chargers each country has.

In order to calculate, the public charger availability it was necessary the total number of electric cars. This data has been extracted from EAFO. Source is no longer available and data for 2021 is since april 2022 in EAFO available [118]. Malta data was not available and National Statistics office data has been used [137]. Due to a difference in the way of counting cars in Malta, an approximation has been done (See Section of Appendix 7.2.2).

Finally, geographical electric public charger density has been extracted from [2]. Norway and UK data has not been provided. Km of road could be extracted from UNECE [135], total kms calculated and charger density obtained (See Section of Appendix 7.2.3).

- **climate change awareness indicator**: Data extracted from Eurobarometer 517 [52]. Poll demanded by the European Comission and done by Kantar. Norway and UK are not included in this survey. Norway survey was done as well by Kantar [79] and UK poll was done by YouGov [133]. Norway and Great Britain polls, are in the same time period between 2020 and 2021 and ask a very similar question and the answer options are similar.

7.2 Data calculus of certain features and approximations

7.2.1 Vehicle age of Cyprus fleet

The vehicle age for different countries is given in [141]. However, in the case of Cyprus certain data is given, but the average age is not given. Considering the cars which are less than 5 years, less than 10 and older than 10 years, an approximation can be done.

After analyzing Table 7.1 it can be observed that Cyprus has a total of cars under 5 years of 14% and a total of cars older than 10 years of 65.7%. This locates te fleet of Cyprus to be older than Portugal which is 13.2 years. it can be observed that the fleet is younger than Hongary because Hongary has the same amount of cars of less than 5 years, but only 12.1% of cars between 5 and 10 years while Cyprus has 20.3% of cars in this group. Then, the average age of cars of Cyprus is between 13.2 and 14.2 years. It has been selected an approximative value of 13.5 years.

7.2.2 Electric vehicles of Malta

National Statistics Office of Malta change the criteria of electric cars between Q1 and Q2. Then, the calculus and suppositions are explained below:

	2020	2019	2018	2017	**2016 (<=5 years)**	2015	2014	2013	2012	**2011(5-10 years)**	**>10 years**	Total	**Average age**
Italy	1481569	2044754	1993189	2020182	1858850	1579376	1335431	1250519	1302675	1603798	23247531	39717874	**11,8**
Percentage per year	3,7	5,1	5,0	5,1	4,7	4,0	3,4	3,1	3,3	4,0			
Percentage per period					**23,7**					**17,8**	**58,5**		
Finland	92074	114242	123330	127960	132810	123047	117409	113890	120152	136586	1546948	2748448	**12,5**
Percentage per year	3,4	4,2	4,5	4,7	4,8	4,5	4,3	4,1	4,4	5,0			
Percentage per period					**21,5**					**22,2**	**56,3**		
Spain	82192	1142134	1202336	1126974	1059774	963725	796528	660435	625558	725489	16044285	25169158	**13,1**
Percentage per year	0,3	4,5	4,8	4,5	4,2	3,8	3,2	2,6	2,5	2,9			
Percentage per period					**18,3**					**15,0**	**63,7**		
Portugal	133067	223175	233392	240205	239774	231305	196479	158108	140511	188304	3315680	5300000	**13,2**
Percentage per year	2,5	4,2	4,4	4,5	4,5	4,4	3,7	3,0	2,7	3,6			
Percentage per period					**20,2**					**17,3**	**62,6**		
Hungary	98802	128308	117278	108066	97925	95031	91583	88209	96326	101115	2896280	3918923	**14,2**
Percentage per year	2,5	3,3	3,0	2,8	2,5	2,4	2,3	2,3	2,5	2,6			
Percentage per period					**14,0**					**12,1**	**73,9**		
Latvia	11921	16005	15081	16254	15420	15705	16634	16418	19593	22562	507369	672962	**14,3**
Percentage per year	1,8	2,4	2,2	2,4	2,3	2,3	2,5	2,4	2,9	3,4			
Percentage per period					**11,1**					**13,5**	**75,4**		
Cyprus	9696	13116	15161	19767	23283	–	–	–	–	117506	379629	578158	–
Percentage per year	1,7	2,3	2,6	3,4	4,0								
Percentage per period					**14,0**					**20,3**	**65,7**		

Table 7.1: Passenger car by age of countries similar to Cyprus

	electric (BEV+PHEV)	Hybrid
Electric cars 2020 quarter 1 (Q1)	2445	2294
	electric (BEV)	Hybrid (+PHEV)
Electric cars 2020 Q2	2068	2795
Electric cars 2020 Q4	2533	3395

Table 7.2: Electric and hybrid cars in Malta in 2020

- In Q1, electric vehicles (EV) are formed by battery electric vehicles (BEV) and plug-in hybrid electric (PHEV).
- In Q2, change the designation of EV. EV is only BEV and PHEV is part of hybrid category.
- It is considered that the difference in electric between 2020 Q1 and Q2 is aprox PHEV. Then, PHEV in 2020 Q2 are 377. 377 represents 13.5% of total hybrid vehicles.
- Supposing the relation of PHEV respect to hybrid stays constant BEV+PHEV can be calculated for 2020 Q4.
- In 2020, Malta had approximately 2991 electric cars

7.2.3 Road kms of UK and Norway

According to [2], roads km include motorways, state, provincial and communal roads. Then, from UNECE [135], it has been extracted the following data and calculated the total number of kms (See Table 7.3).

The total kms of road are calculated and the public chargers per 100 kms can be obtained for Norway and UK. Different countries were evaluated to observe if the result following this process and it was very precise. The only problem is UK kms of road are calculated for 2019.

United Kingdom (2012)	
Motorways	3733
State roads	8507
Provincial roads	38235
Communal roads	344531
Total kms	**395006**
Charging points per 100 km	**8,5**

Norway (2019)	
Motorways	1008
State roads	10753
Provincial roads	44696
Communal roads	39715
Total kms	**96172**
Charging points per 100 km	**20,1**

Estonia (2019): used for verification	
Motorways	161
State roads	16609
Provincial roads	24060
Communal roads	18398
Total kms	**59228**
Charging points per 100 km verification	**0,72**

Table 7.3: Charging points per 100 km calculus for Norway and UK

Extrapolations could not be done because there are countries where the kms of roads increase and in others decrease. However, it is a quite good approximation. The variation of the kms in 8 years is not big for countries like Germany or France. Nor should be for the UK.

7.3 Dimensionality reduction benefits

Deleting multicollinearity and using an optimum amount of dimensions is relevant for distance-based algorithms. It is important to reduce the effect of the curse of dimensionality, phenomena appearing when having a high-dimensional space. This effect has been studied and in high dimensional space data becomes more sparse, more separated and algorithmic techniques fail from an efficiency and/or effectiveness perspective. Research shows that the concept of proximity, distance or nearest neighbor may not even be qualitatively meaningful. In Paper [28], it is shown the dimensionality curse from the point of view of the distance metrics.

In [16], it is shown that under a wide set of conditions, when dimensionality increases, the distance to the nearest data point approaches the distance to the farthest data point. Empirical results are presents and it is shown that this effect can occur for as few as 10-15 dimensions. The model defined has 11 features, so it can already be affected by this phenomenon.

Data becomes more sparse and separated. This can be easily visualized by considering which amount of data concentrates within 10% of distance to the edges. This is purely a geometrical feature of a N-dimensional space and if there is a sufficient number of points following a uniform distribution, the data which concentrates within this borders is presented in Table 7.4

Number of dimensions	**Data concentrated within 10% of distance to the edges**
1	80%
2	64%
3	51%
4	41%
5	33%
6	26%
7	21%
8	17%
9	13%
10	11%
11	8.6%

Table 7.4: Percentage of data concentrated within 10% of distance to the edges

Data in space is more sparse and concentrates in the edges. Average distance between two points of data increases. Additionally, higher dimensional space is less meaningful because if it is considered the average distance between two random points following a uniform distribution in a N-dimensional space, it is observed that the average distance between the two points increases with higher number of dimensions.

The average distance between two random points has been calculated analytically for certain number of dimensions by different authors [105]. Here, a numerical approach is used. It is considered data has been normalized, the mean is 0 and standard deviation is 1. Then, the space coordinate axis is defined between -1 and 1. The average distance between two random points for a N-dimensional spaces with coordinates [-1,1] is shown in Table 7.5 .

After observing data is more sparse, average distance increases with higher number of dimensions and distance between two random points increases as well, it can be understood why distances between points lose their meaning.

At this point, it is calculated the distance between a country and all other countries before applying dimensionality reduction and after it. Distances are normalized respect the average distance between two random points for a N-dimensional space. A boxplot is presented in Figure 7.1 for the original space and in Figure 7.2 a boxplot for a 4D space after applying PCA. In Appendix 7.3.1 a boxplot for a 3D space after applying PCA (Figure 7.3) and a boxplot for a 2D space after applying PCA (Figure 7.4) can be consulted.

Number of dimensions	Average distance between 2 points
1	0,668
2	1,046
3	1,322
4	1,556
5	1,761
6	1,934
7	2,099
8	2,255
9	2,400
10	2,546
11	2,666

Table 7.5: Average distance between two points in N-dimensional space with coordinate axis [-1,1]

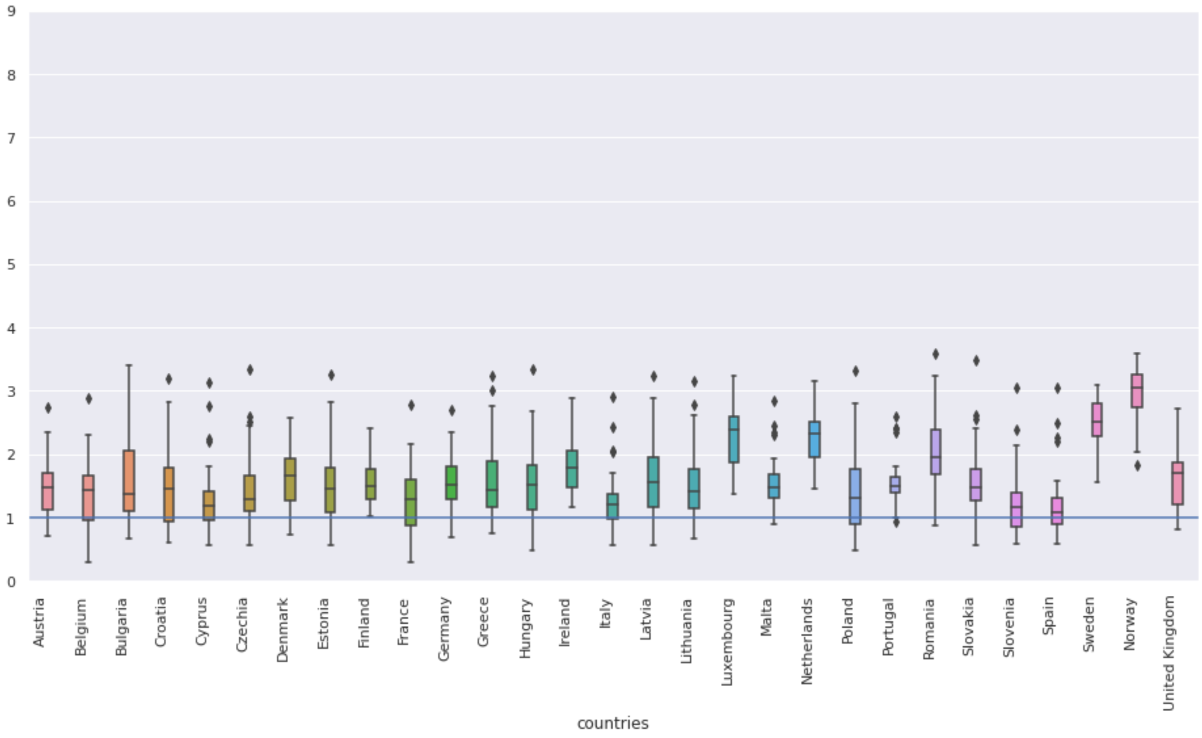


Figure 7.1: Boxplot of the distance in the original space (11D) between countries normalized respect to the average distance between two random points.

In both figures a significant difference can be observed, the interquartile distance (difference between third and first quartile) and the whiskers are larger in the 4D (Figure 7.2) case than in 10D case (Figure 7.1). For the 3D case (Figure 7.3) and the 2D case (Figure 7.4) this is even larger. Country distance presents a higher variance when reducing the number of dimensions.

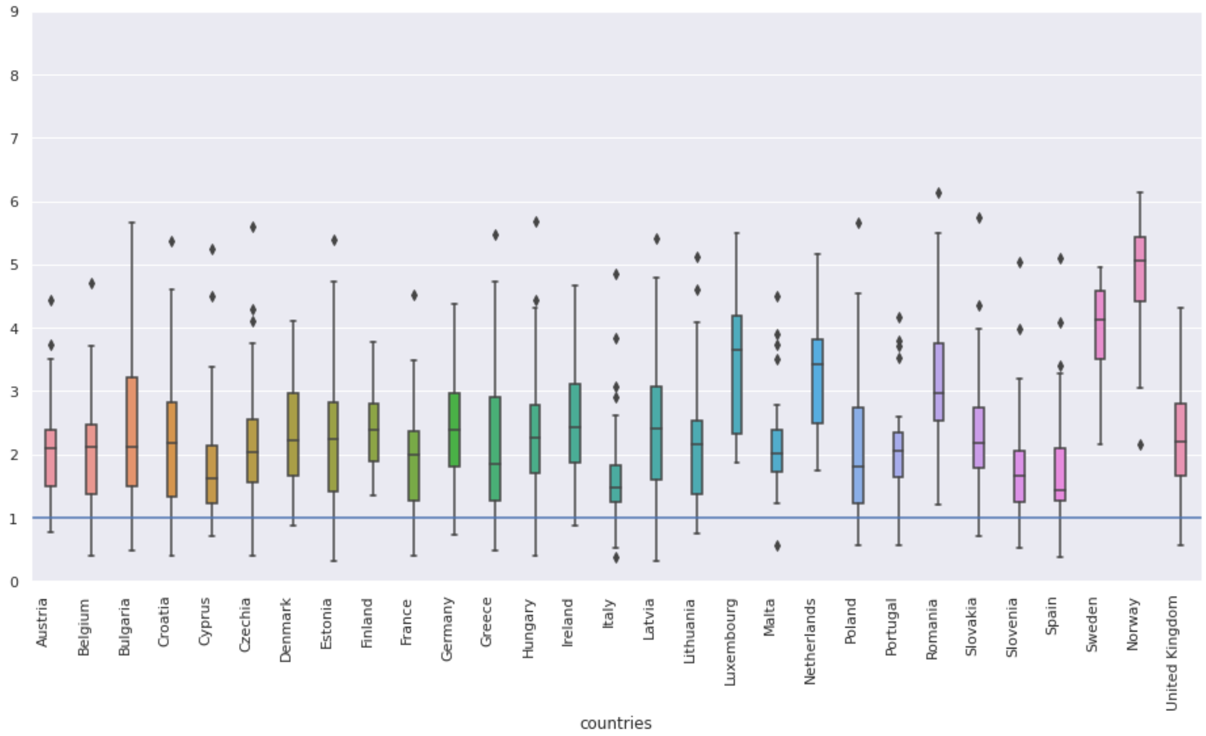


Figure 7.2: Boxplot of the distance after applying PCA (4D) between countries normalized respect to the average distance between two random points.

For quantifying results, Table 7.6 is presented where average distance between countries and between two random points is presented for a different number of dimensions. Normalised distances are presented.

	10D (original features)	**4D (PCA)**	**3D (PCA)**	**2D (PCA)**
average distance between countries	1,66	2,46	2,69	2,95
standard deviation	0,54	0,99	1,23	1,65
average distance between two random points	2,66	1,55	1,33	1,04
normalised average distance between countries	0,66	1,59	2,05	2,81
normalised standard deviation	0,21	0,64	0,94	1,57

Table 7.6: Average distance and standard deviation between countries for the original space and for a different number of dimensions after applying PCA

It can be observed in Table 7.6 that in 10D the normalised average distance between countries is 0.66. This is under 1 which means that it is inside a threshold where the grouping can be done randomly. Additionally, standard deviation is really small making it very difficult for distance-based algorithms like Kmeans or hierarchical clustering to determine different clusters. By reducing the number of dimensions, normalised average distance between two countries becomes bigger than 1 and standard deviation increases considerably. The information is more meaningful in this new space and a considerable improvement in the performance of similarity based algorithms is achieved.

7.3.1 Boxplots of distance between countries in a N-dimensional space

Here, boxplots of the distance between countries normalized with the random average distance between two points after applying PCA and keeping 3 principal components is presented in Figure 7.3 and keeping 2 principal components is presented in Figure 7.4

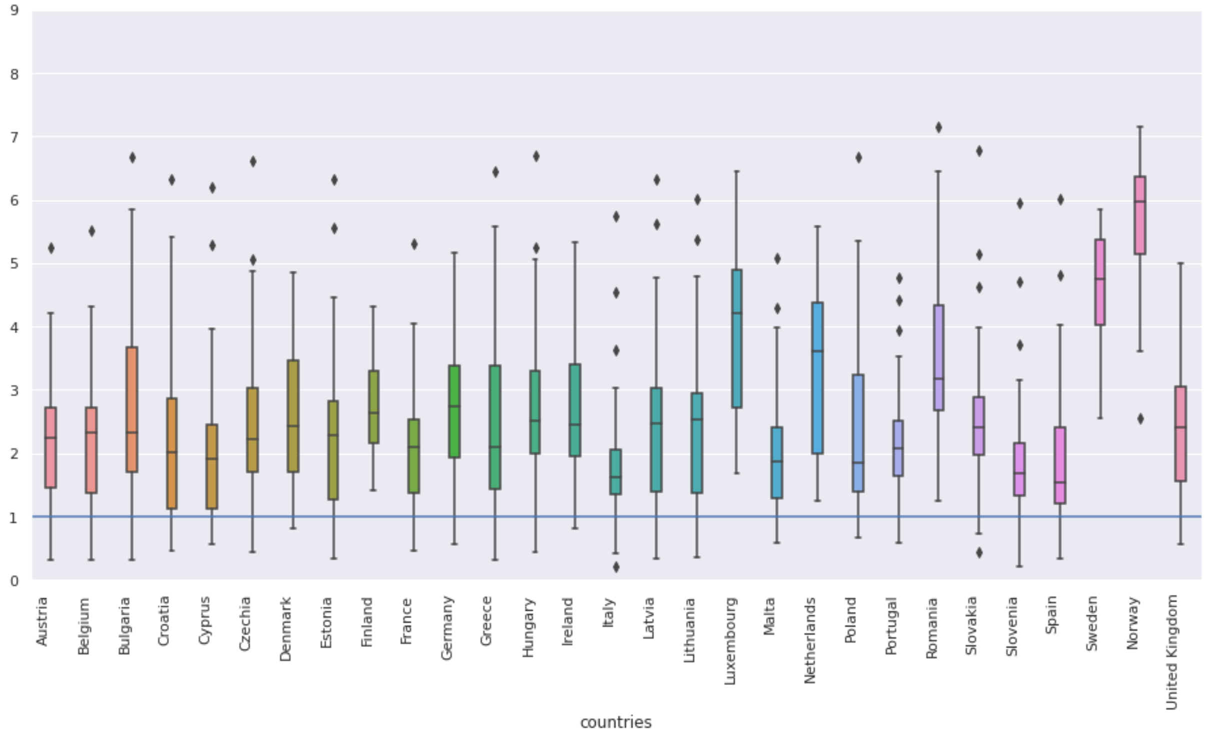


Figure 7.3: Boxplot of the distance after applying PCA (3D) between countries normalized respect to the average distance between two random points.

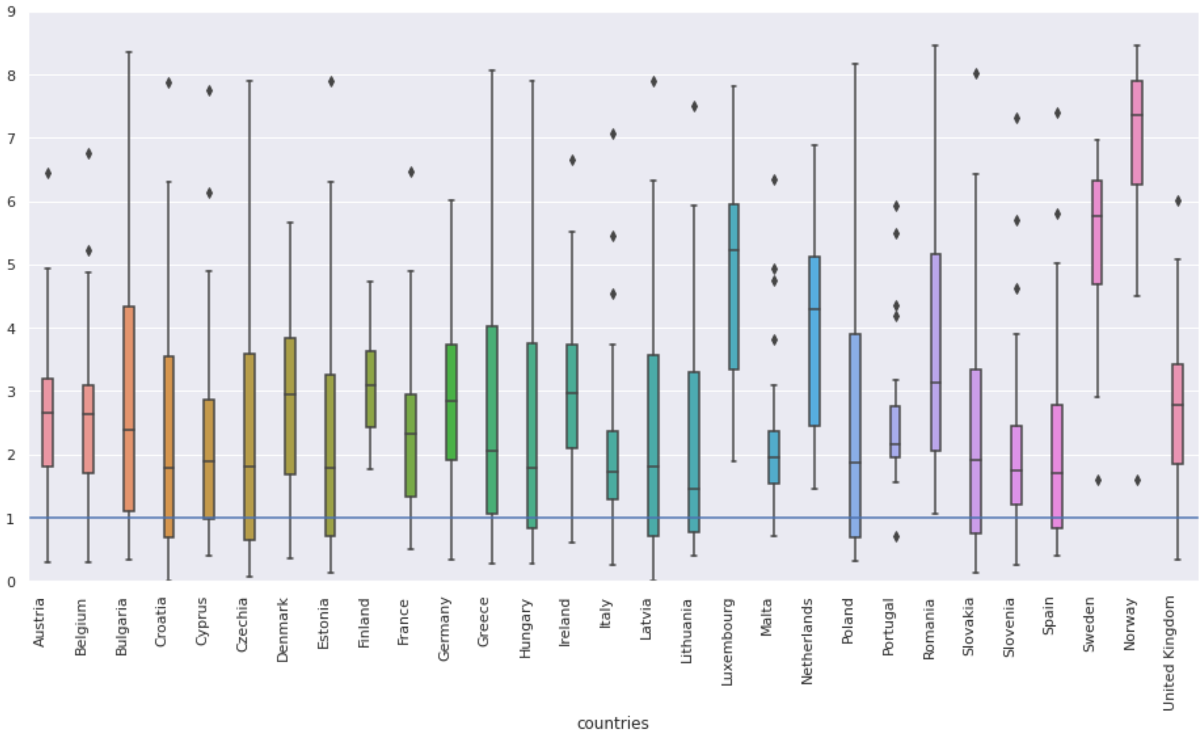


Figure 7.4: Boxplot of the distance after applying PCA (2D) between countries normalized respect to the average distance between two random points.

7.4 Hierarchical clustering additional plots

7.4.1 Hierarchical clustering for different number of clusters

In order to observe how the silhouette coefficients are distributed in each of the clusters and if there are any missclassifications, Silhouette plots are used. Here, dimensionality reduction techniques has been used and the model has 3 principal components.

This clustering has been performed using:

- method = ward
- metric = euclidean

The Silhouette plots obtained are presented in Figures 7.5, 7.6, 7.7 and 7.8

It can be observed how different mean results (red dashed line) are different for different number of clusters. For 4, 6 and 8 clusters there is one country missclassified. For 5 clusters, there are no missclasifications.

The values obtained for 4 clusters are the highest, but a higher number of groups is searched. From next clusters 6 clusters presents the highest Silhouette score,0.401 followed by 7 clusters which has a score 0.398.

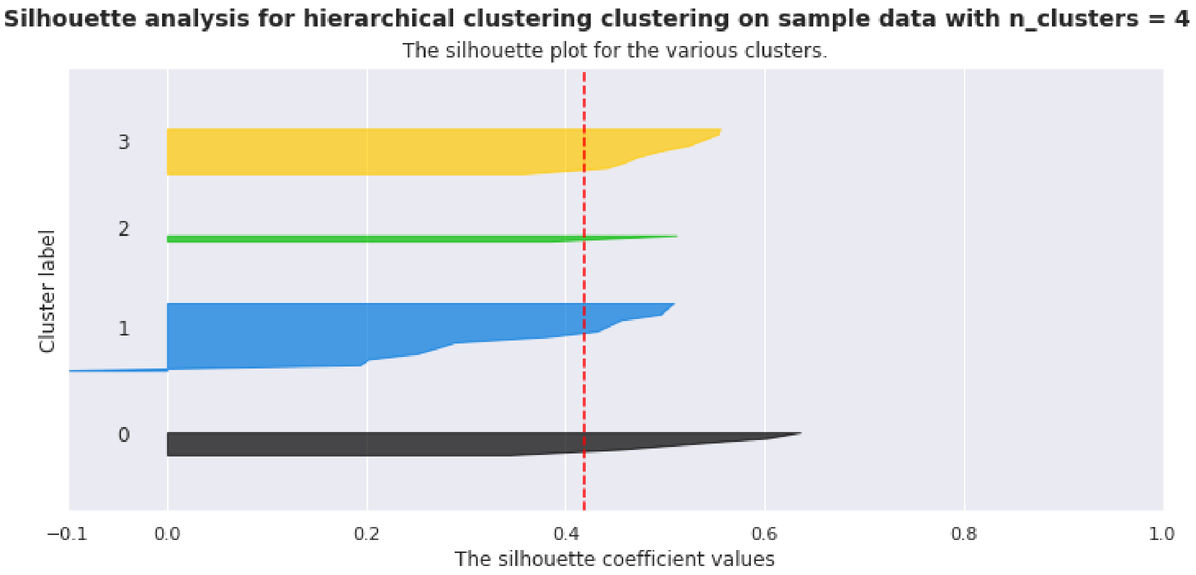


Figure 7.5: Silhouette plot for 4 clusters

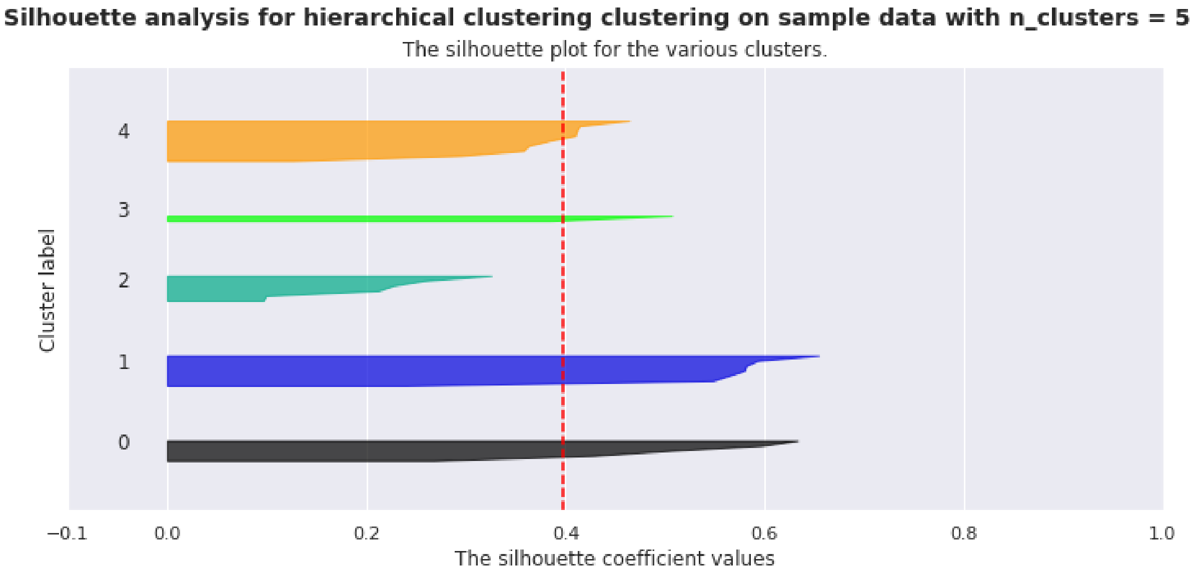
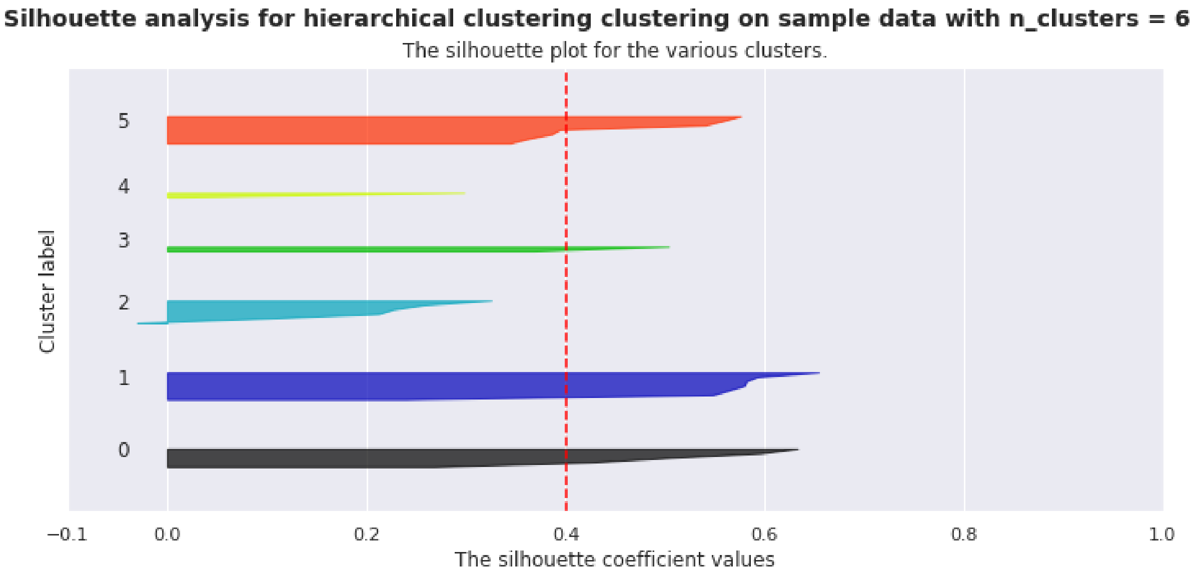


Figure 7.6: Silhouette plot for 5 clusters



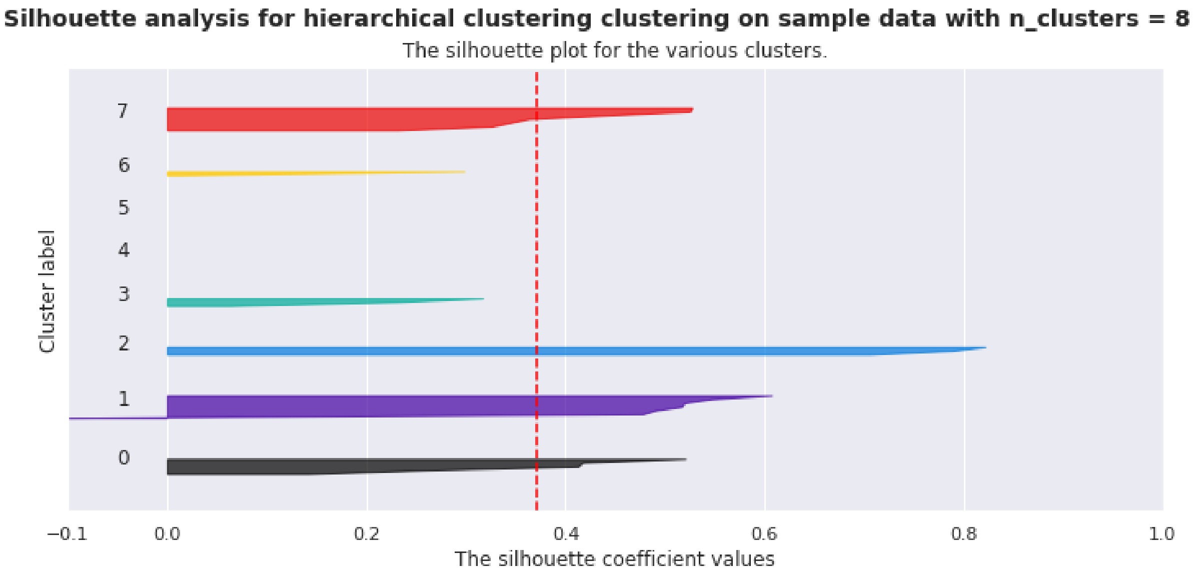


Figure 7.8: Silhouette plot for 8 clusters

7.4.2 Dendrograms for different linkage methods and similarity measurements

It is important to complement the judgement from Section 2.5.3.2 for selecting the best model not just using Similarity score, but also analyzing the dendrogram obtained:

Single link

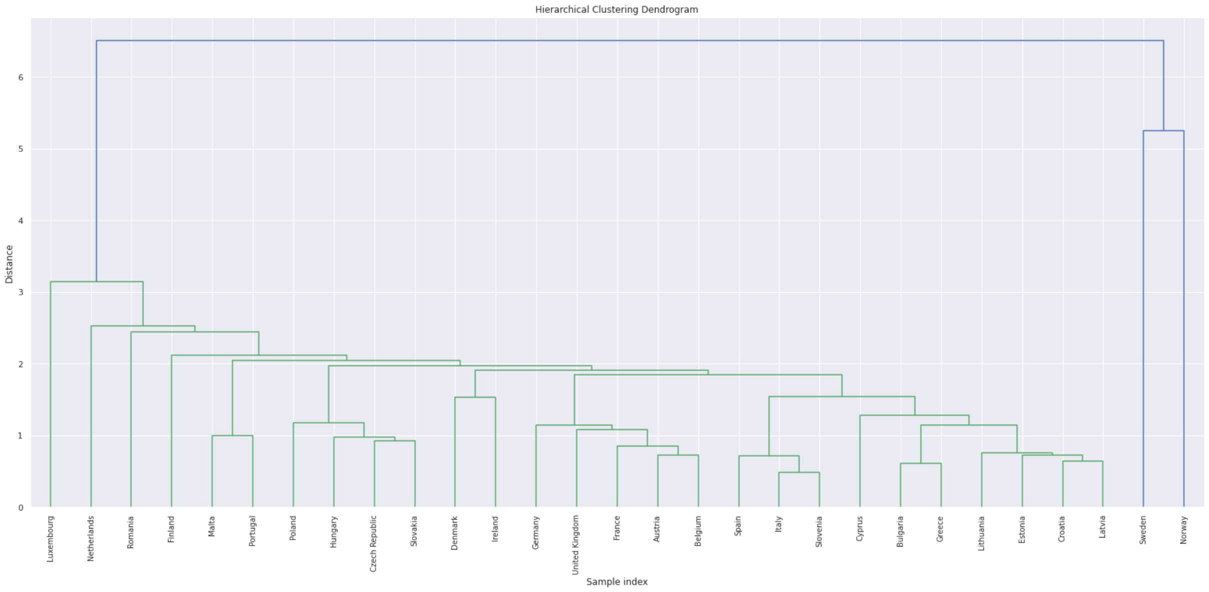


Figure 7.9: Dendrogram using single link

Complete link

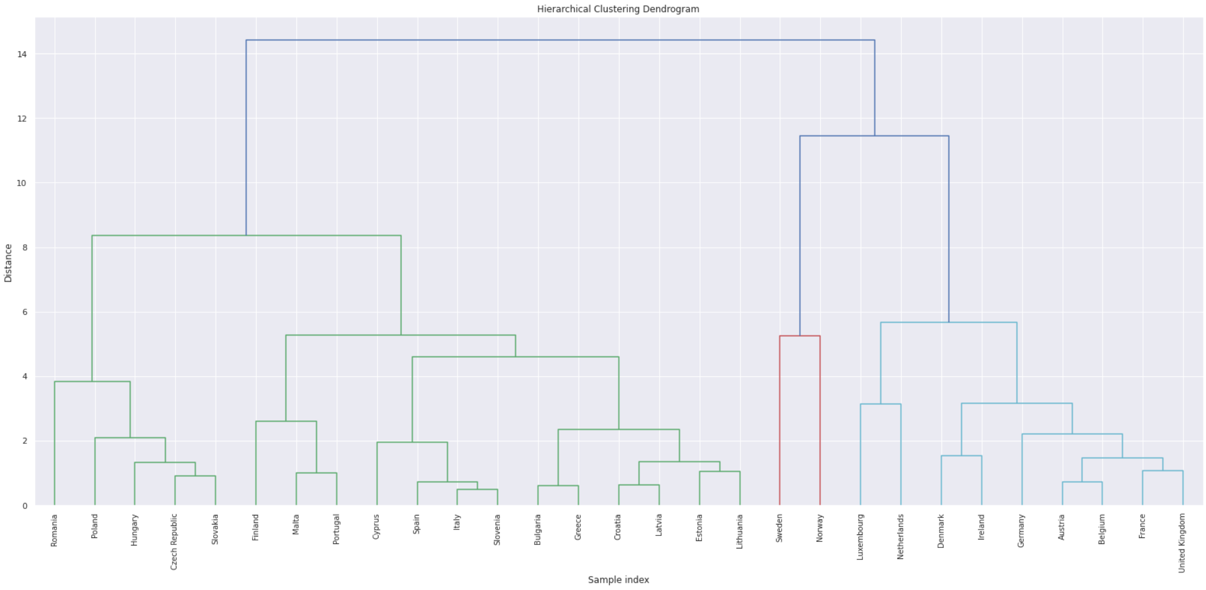


Figure 7.10: Dendrogram using complete link

Average link

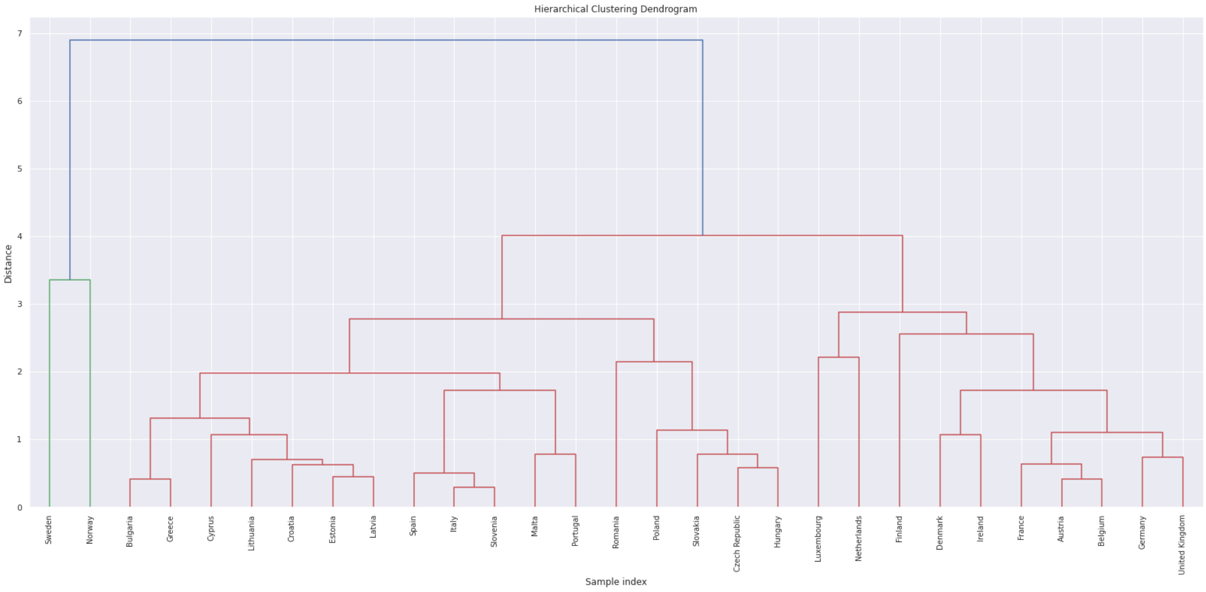


Figure 7.11: Dendrogram using average link

Weighted link

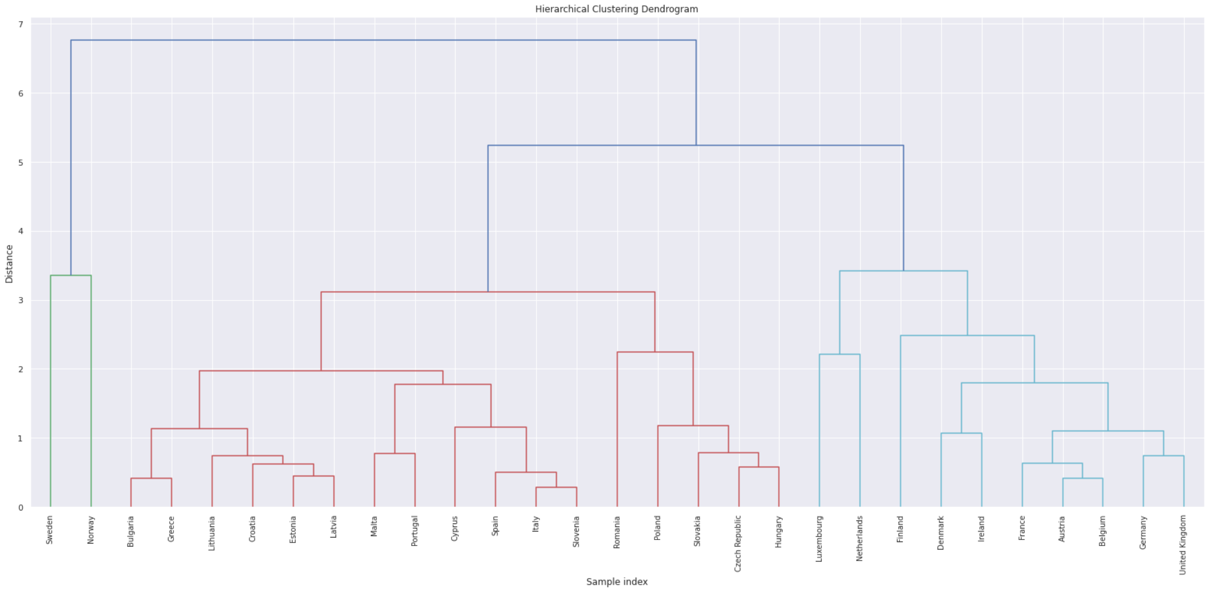


Figure 7.12: Dendrogram using weighted link

Centroid method

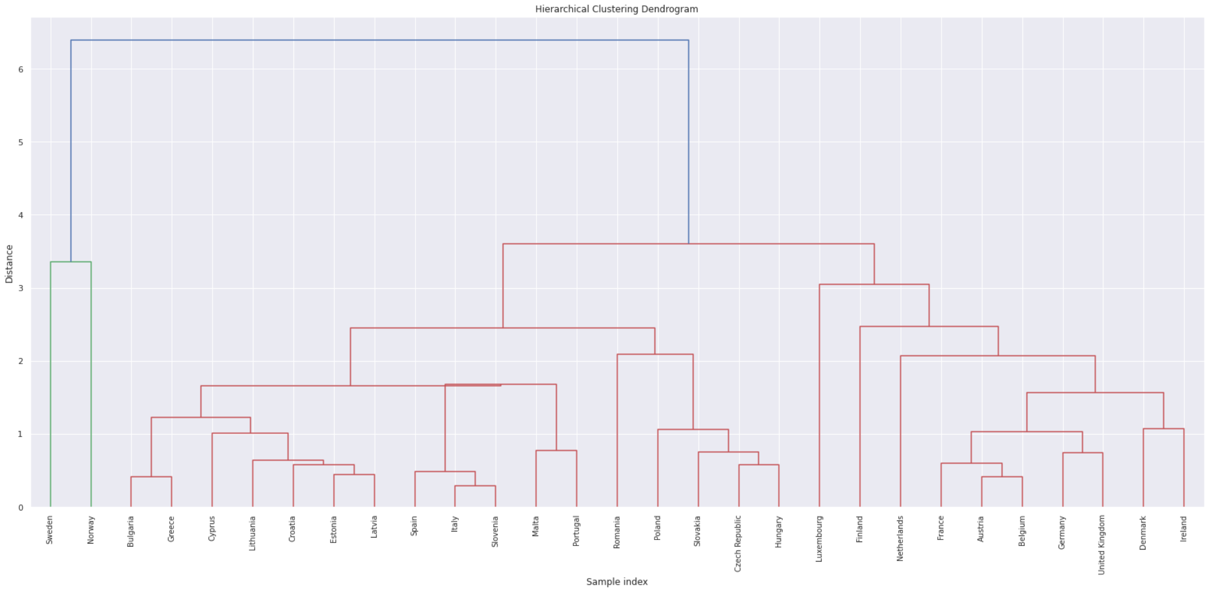


Figure 7.13: Dendrogram using centroid method

Median method

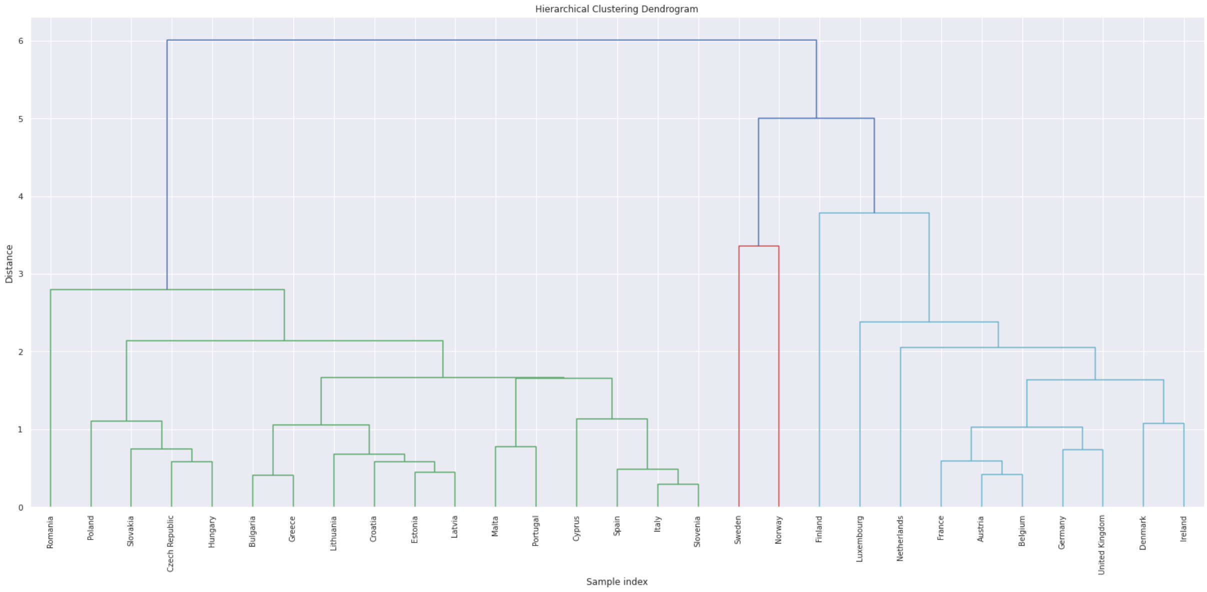


Figure 7.14: Dendrogram using median method

Ward method

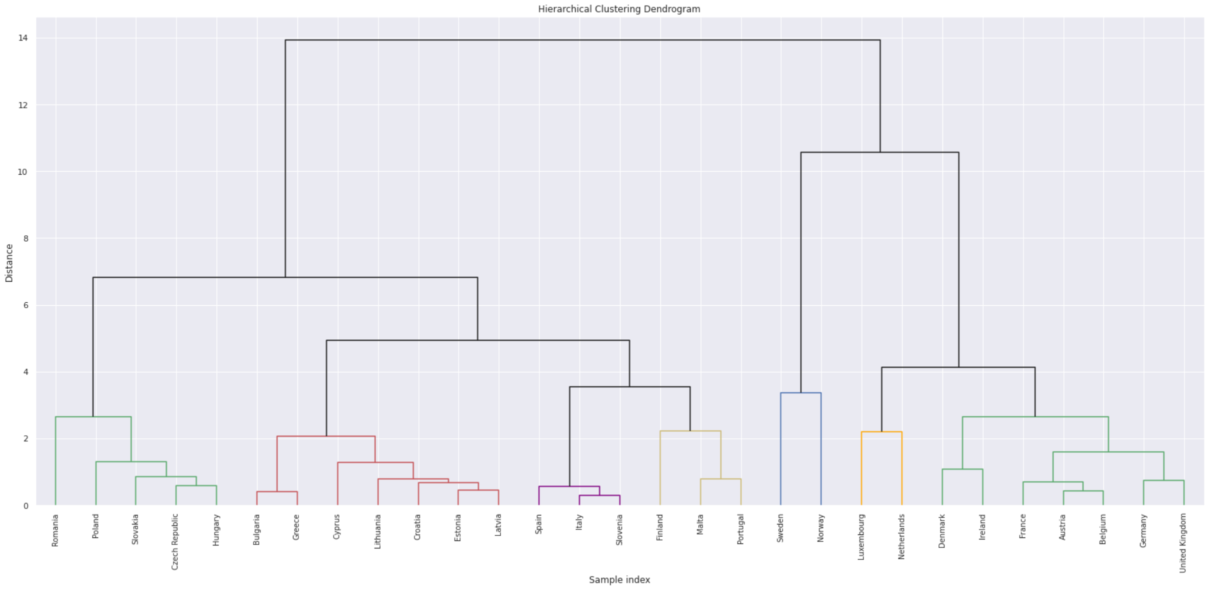


Figure 7.15: Dendrogram using ward method

7.4.3 Stability of data driven process

Apart from analyzing the Silhouette score, it is analyzed how the model varies when deleting certain features. The procedure of dimensionality reduction (3 PCs) and the selection of linkage and similarity measurement (ward and euclidean) is reused.

A major indicator for reducing CO_2 emissions in the passenger vehicle market is the new registrations of electric vehicles. It is analyzed what happens when deleting this feature.

New registrations of electric vehicles

If results obtained in the model are stable each country will remain similar and the Silhouette score would not vary much. Once dimensionality reduction techniques are used, a 2D plot can be observed in subsection of the Appendix 7.4.3.1 in Figure 7.16. It looks very similar compared with Figure 2.7. The main differences are that Norway and Sweden look much closer and Finland looks to be a bit more separated from Portugal and Malta.

The dendrogram obtained produces exactly the same clustering as without deleting the new registrations of electric cars term. This shows that the model is explaining well the intake of electric cars each cluster will have even if it is not added. The dendrogram obtained can be consulted in Subsection of the Appendix 7.4.3.1 in Figure 7.17.

Finally, the Silhouette score obtained without this feature is 0.403 for 6 clusters which is slightly higher than the Silhouette Score obtained with all features before which was 0.402. (See in section 2.5.3.2).

The results obtained are logical considering that the registration of new electric cars had a very high VIF (See in Section 2.5.1.5). This showed that this feature had a very high level of multicollinearity and could be deleted without losing almost information.

Social awareness

The term of social awareness is a parameter which can be hard to find if the model wants to be extended at an international level. It is evaluated what happens if it is deleted in order to observe if the model is very dependant from this term.

A similar 2D plot is obtained as in previous cases. It can be consulted in Appendix 7.4.3.1 in Figure 7.18.

The dendrogram obtained presents a small variation respect to the model which uses all features. Only one country is moving from the previous cluster groups. Netherlands is moving because the awareness of the habitants in the Eurobarometer regarding climate change was higher compared to Germany and France. If this feature is deleted, it gets closer to early majority group than to Luxembourg. The dendrogram obtained can be consulted in Appendix 7.4.3.1 in Figure 7.19.

In this case the Silhouette score for 6 clusters is 0.386. The score decreases slightly and it can be concluded that the term can be deleted because the changes observed are not significant. However, it is adding useful information to the model.

The social awareness indicator has as well a VIF higher than 5 which is telling that this variable is highly correlated to others and deletion of term should not generate major changes as it has been observed.

The deletion of these features is showing that if a model is created with descriptive features and it is large enough, the level of multicollinearity will be high. Thus, deleting features from a possible international model or selecting others which are similar should not be a problem for still obtaining meaningful and useful results.

7.4.3.1 Stability of data driven process plots

Results deleting new registrations of electric cars

The plot of the first two principal components (Figure 7.16) is shown once the feature of new registrations of electric cars is deleted.

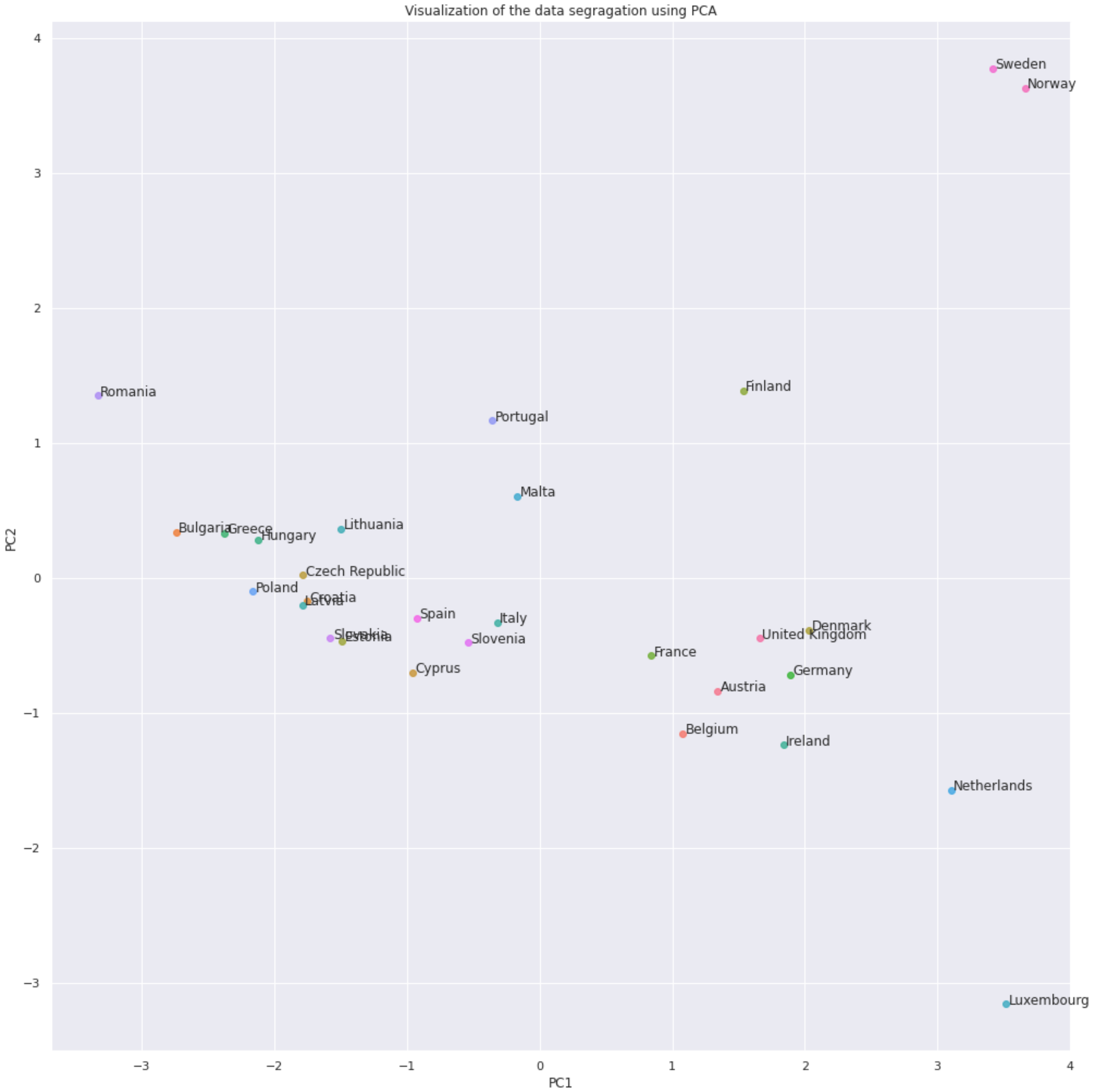


Figure 7.16: Representation of the countries in the first two principal components (PC1 and PC2) without new registrations of electric cars in 2020

Once the countries in the principal components are plotted, the final dendrogram is shown in Figure 7.17.

The same results, identical clusters, as in the final dendrogram (Figure 2.10) which considers all features are obtained

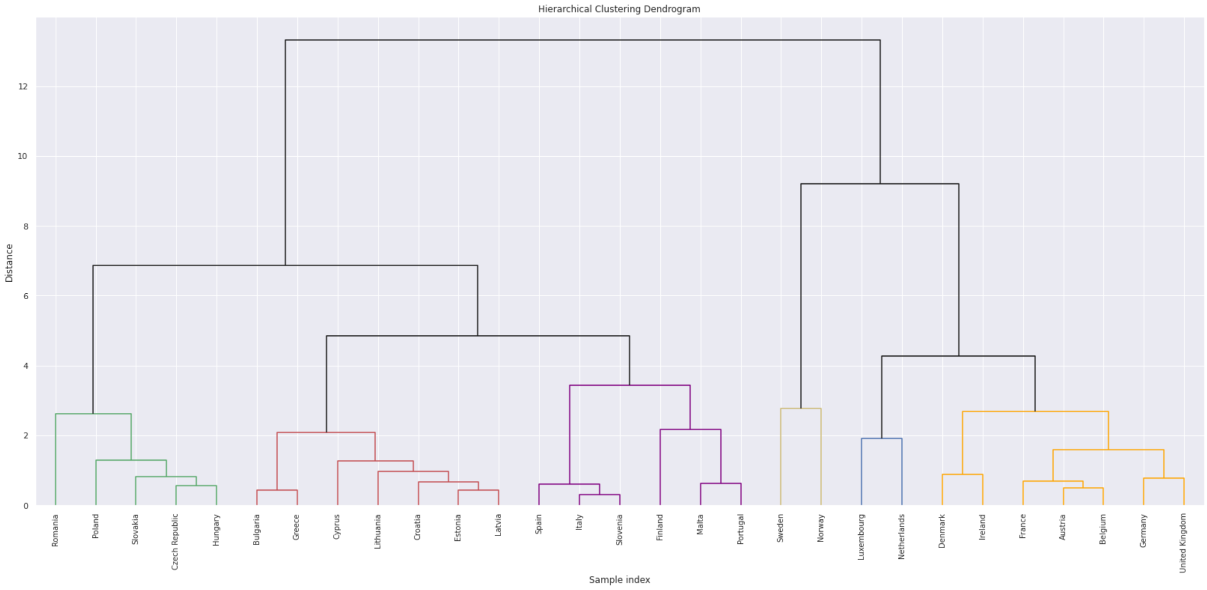


Figure 7.17: Alternative dendrogram obtained when new registrations of electric cars is deleted

Results deleting social awareness indicator

The plot of the first two principal components (Figure 7.18) is shown once the feature of climate change awareness indicator is deleted.

Once the countries in the principal components are plotted, the final dendrogram is shown in Figure 7.19.

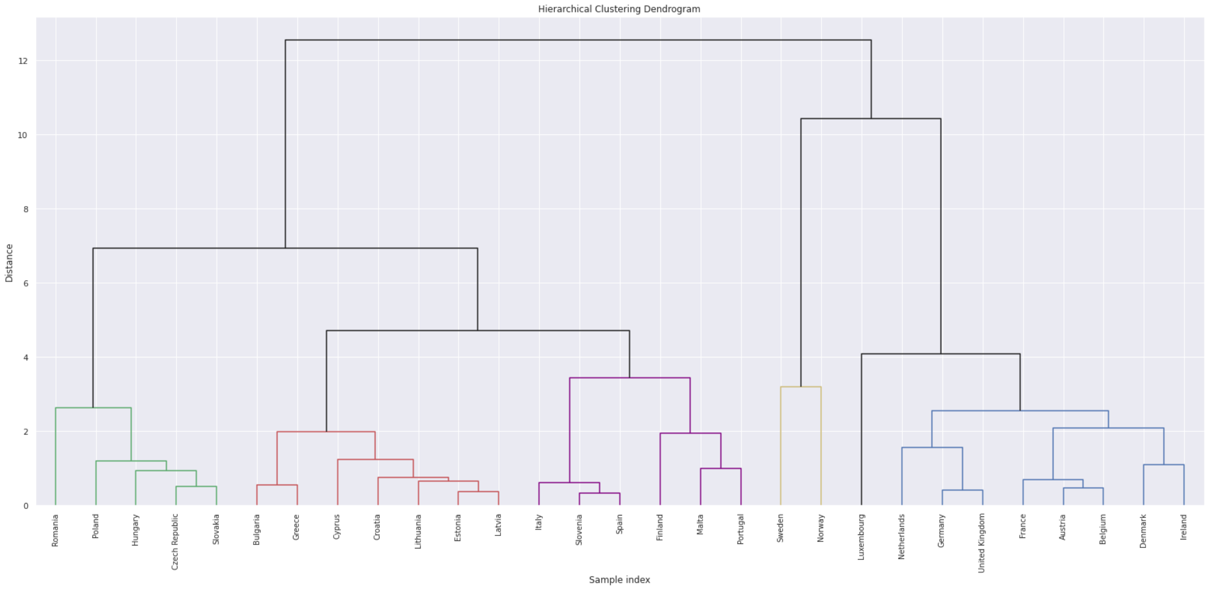


Figure 7.19: Alternative dendrogram obtained when climate change awareness indicator is deleted

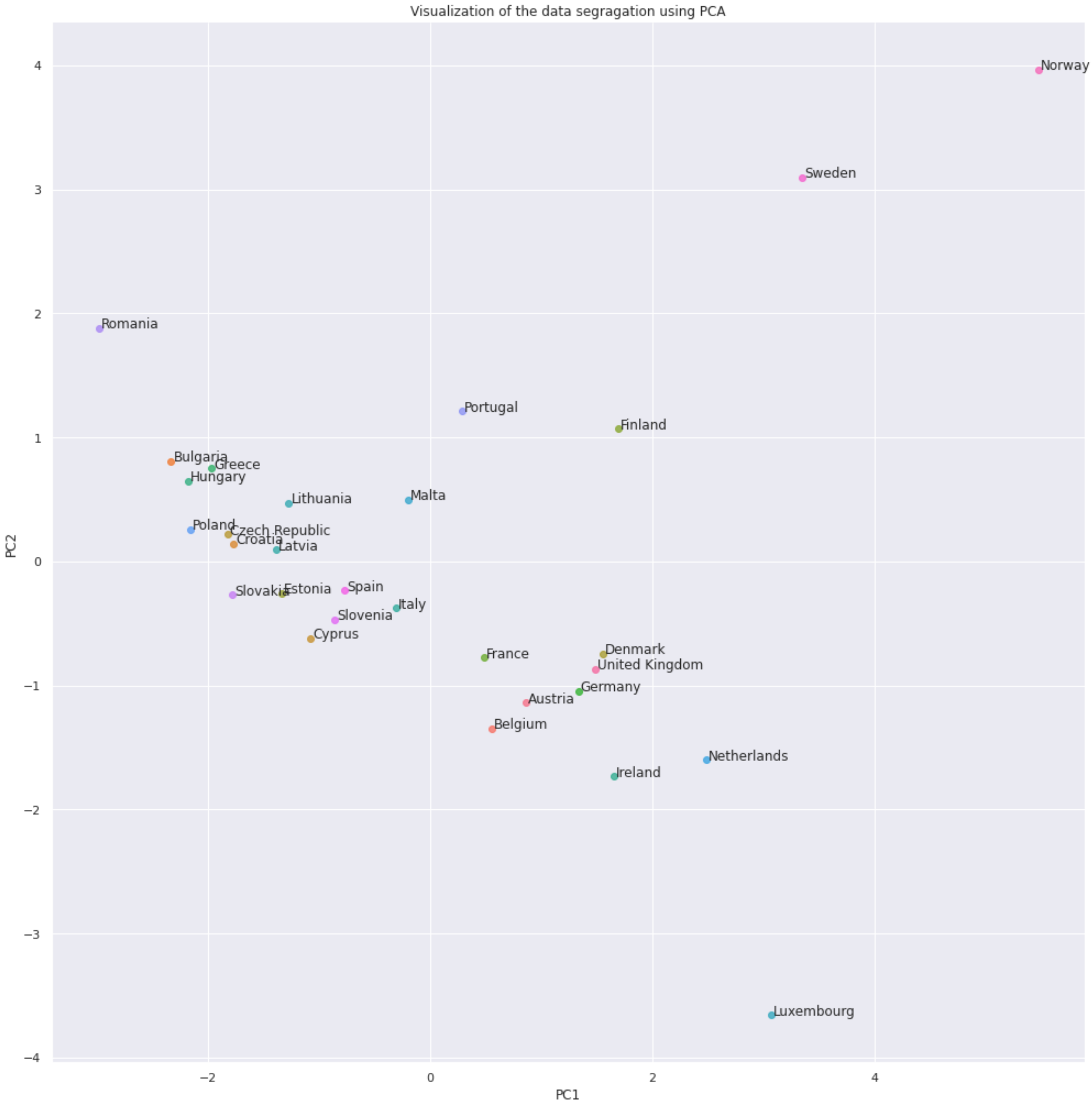


Figure 7.18: Representation of the countries in the first two principal components (PC1 and PC2) without climate change awareness indicator

When climate change awareness indicator is deleted, only one cluster is changing. Netherlands is moving with Germany and France cluster.

7.4.4 Statistical significance for model of 6 clusters

If 6 clusters are selected the intervals can be observed in the Table 7.7:

It is important to remark that obtaining results where the means of the principal components of a certain group are not statistically significant does not mean that the results are not valid (considering all principal components). Silhouette score is showing how densely and separated

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
	lower limit	**higher limit**	**lower limit**	**higher limit**	**lower limit**	**higher limit**	**lower limit**	**higher limit**	**lower limit**	**higher limit**	**lower limit**	**higher limit**
PC1	-3,31	-1,12	-2,87	-0,88	-2,11	1,84	2,81	6,69	2,81	3,41	0,41	2,25
PC2	-1,11	2,11	-0,84	1,24	-1,24	1,83	2,20	4,96	-5,40	0,2	-1,77	-0,4
PC3	0,47	2,55	-1,76	-0,52	-1,36	0,79	-3,90	4,41	-1,01	1,67	-1,50	1,75

Table 7.7: Confidence intervals for 3 principal components, confidence level 95 %

from the rest is each cluster. The mean of one principal component cannot be significant or in any, but while you increase dimensions each cluster is getting separated from the rest. Obtaining statistical significance is useful to clearly identify in which position is located each cluster inside a certain principal component axis and this is useful for further labelling.

For the confidence intervals it has been used the 2σ criteria which means including a 95.4% of all data points in the confidence interval. PC1 shows that confidence intervals of group 4 and 5 are separated from the rest but not between each other defining well the groups in this direction. Then, group 6 is next group and crosses with group which has a broad confindence interval. Group 2 and group 1 are the last ones.

For PC2, group 4 and 5 are separated. This two groups have at least one principal component which has a confidence interval that not cross with others groups. The rest of groups are not separated.

There are just two groups which are possible to separate from the rest. After observing this and the high variance of cluster 3, it is analyzed and it is observed Finland is causing this problem.

Deletion of Finland

If Finland is deleted (it represents justs 1% of the market and previously was causing the Silhouette plot to be negative), the following confidence interval groups are obtained in Table 7.8

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
	lower limit	**higher limit**	**lower limit**	**higher limit**	**lower limit**	**higher limit**	**lower limit**	**higher limit**	**lower limit**	**higher limit**	**lower limit**	**higher limit**
PC1	-3,31	-1,12	-2,87	-0,88	-1,28	0,25	2,81	6,69	2,81	3,41	0,41	2,25
PC2	-1,11	2,11	-0,84	1,24	-1,34	1,60	2,20	4,96	-5,40	0,2	-1,77	-0,4
PC3	0,47	2,55	-1,76	-0,52	-1,25	0,88	-3,90	4,41	-1,01	1,67	-1,50	1,75

Table 7.8: Confidence intervals for 3 principal components without Finland, confidence level 95 %

This allows to have 4 groups with separated confidence intervals in at least one principal component. This is a significant improvement. Groups 2 (Group of Greece) and 3 (Group of Italy, Finland) are still not separated. After observing this phenomena instead of deleting Finland, it is analyzed what happens if instead of having 6 clusters, there are 7 clusters. The new cluster appears by separating cluster of Malta, Portugal, Finland, Italy, Spain and Slovenia into two clusters, one new cluster is Malta, Portugal and Finland and the other Italy, Spain and Slovenia (See Figure 2.10 and use a distance threshold of 3.5 instead of 4).

7.4.5 Representation of confidence intervals of final model

Confidence intervals of PC1 are shown in Section 2.5.3.4 in Figure 2.11. Here it is presented the graphical representation of confidence intervals of PC2 and PC3 (See Figure XXX and Figure respectively).

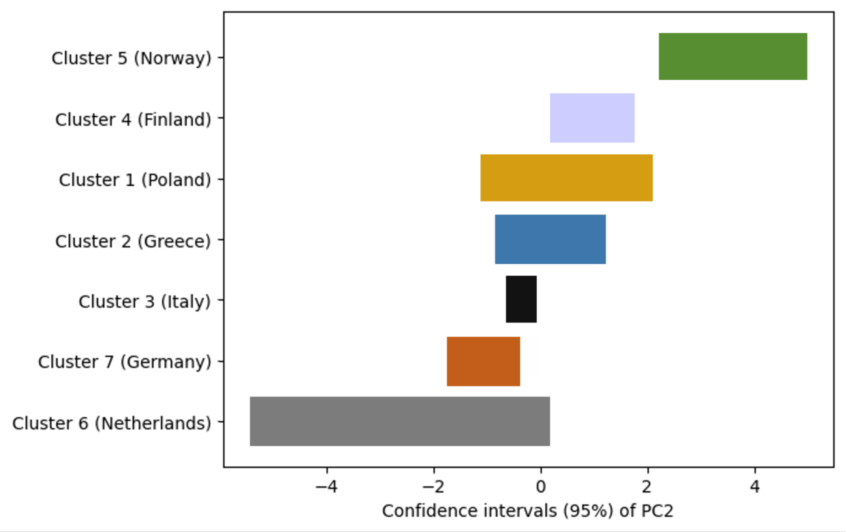


Figure 7.20: Graphical representation of confidence intervals of PC2. A country of each cluster is mentioned to identify quicker each cluster.

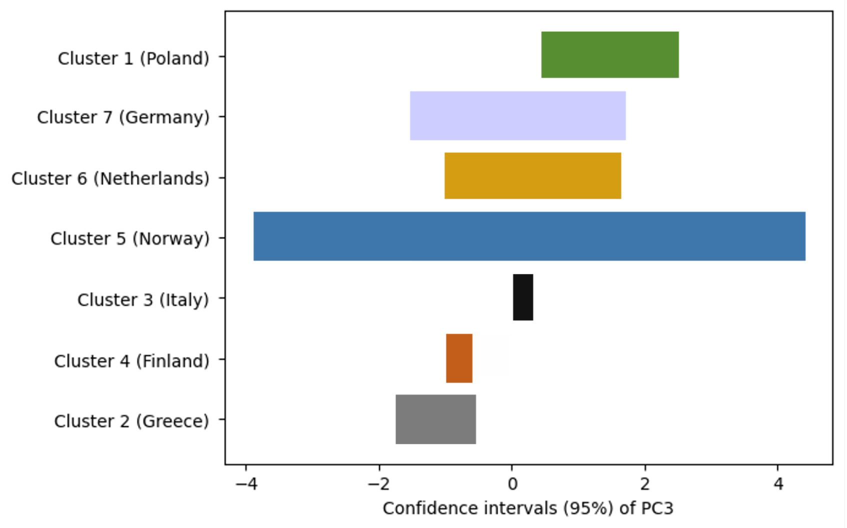


Figure 7.21: Graphical representation of confidence intervals of PC3. A country of each cluster is mentioned to identify quicker each cluster.

7.5 Ranking of original features of each countries

Here a table where how many features are ranked in the best places between 1st and 5th is shown for each country (See in Table 7.9) and as well a Table were the worst ranked countries can be seen (See in Table 7.10).

	rank 1	**rank 2**	**rank3**	**rank 4**	**rank 5**	**Ranked top 5**	**Ranked top 3**
Austria	0	1	0	0	1	2	1
Belgium	0	0	0	0	1	1	0
Bulgaria	0	0	0	0	0	0	0
Croatia	0	0	1	0	0	1	1
Cyprus	1	0	0	0	0	1	1
Czechia	0	0	0	0	0	0	0
Denmark	0	0	0	2	4	6	0
Estonia	1	0	0	0	0	1	1
Finland	0	0	2	1	0	3	2
France	0	0	0	0	0	0	0
Germany	0	1	0	1	1	3	1
Greece	0	1	0	0	0	1	1
Hungary	0	0	0	0	0	0	0
Ireland	0	1	1	0	0	2	2
Italy	0	0	0	0	0	0	0
Latvia	0	0	0	1	0	1	0
Lithuania	0	0	0	0	0	0	0
Luxembourg	2	1	2	0	0	5	5
Malta	0	0	0	0	1	1	0
Netherlands	1	1	1	2	1	6	3
Poland	0	0	0	0	0	0	0
Portugal	0	0	0	0	1	1	0
Romania	0	0	0	0	0	0	0
Slovakia	0	0	0	0	0	0	0
Slovenia	0	0	0	0	0	0	0
Spain	0	0	0	0	0	0	0
Sweden	2	2	0	1	0	5	4
Norway	2	2	2	2	0	8	6
United Kingdom	1	0	0	0	0	1	1

Table 7.9: Number of times a country feature is ranked 1st, 2nd,3rd,4th or 5th best place, the amount of times is ranked in the top 3 and ranked in the top 5

	rank 1	**rank 2**	**rank3**	**rank 4**	**rank 5**	**Ranked top 5**	**Ranked top 3**
Austria	1	0	0	0	0	1	1
Belgium	0	0	1	0	1	2	1
Bulgaria	3	1	0	0	0	4	4
Croatia	0	2	0	1	0	3	2
Cyprus	1	0	0	1	0	2	1
Czechia	0	0	1	1	0	2	1
Denmark	0	0	0	0	0	0	0
Estonia	0	0	0	2	1	3	0
Finland	0	0	0	0	0	0	0
France	0	0	0	0	0	0	0
Germany	0	0	0	0	0	0	0
Greece	1	1	2	0	1	5	4
Hungary	0	0	1	0	3	4	1
Ireland	0	0	0	0	0	0	0
Italy	0	0	0	0	1	1	0
Latvia	1	1	1	0	2	5	3
Lithuania	0	2	1	0	1	4	3
Luxembourg	0	1	0	0	0	1	1
Malta	1	0	0	0	0	1	1
Netherlands	0	0	0	0	0	0	0
Poland	0	0	1	1	0	2	1
Portugal	0	0	0	0	0	0	0
Romania	2	1	1	3	0	7	4
Slovakia	0	1	0	0	0	1	1
Slovenia	0	0	0	0	0	0	0
Spain	0	0	0	0	0	0	0
Sweden	0	0	1	0	0	1	1
Norway	0	0	0	0	0	0	0
United Kingdom	0	0	0	1	0	1	0

Table 7.10: Number of times a country feature is ranked 1st, 2nd,3rd,4th or 5th worst place, the amount of times is ranked in the top 3 and ranked in the top 5

7.6 Bloomberg study results

Here it is shown the grouping done by BloombergNEF in Figure 7.22 in the study of the electric vehicle market. For more information consult [64].

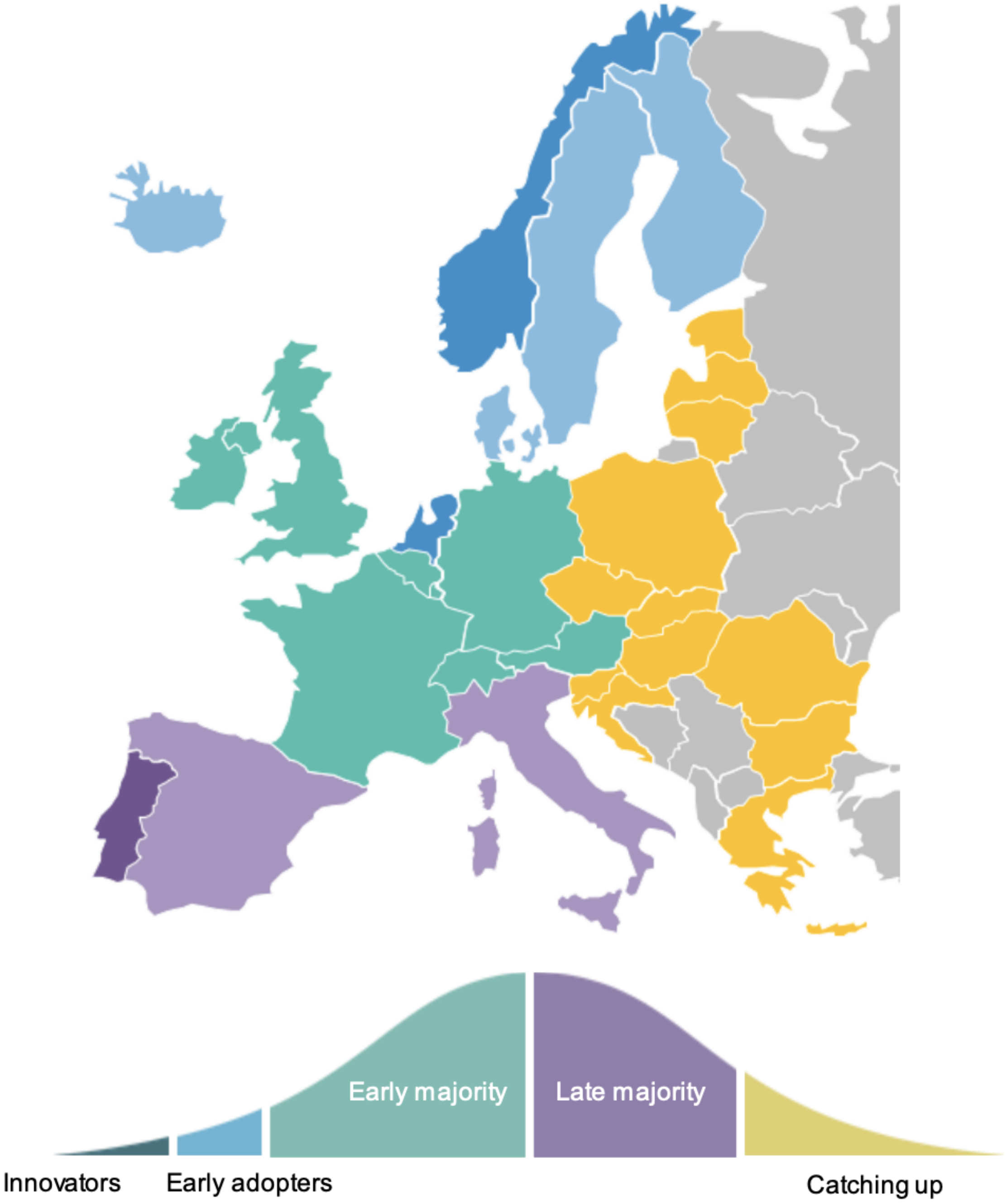


Figure 7.22: Distribution of adoption groups and country grouping. Extracted from [64]

.

8 Appendix of Chapter 4

8.1 Taxation in Germany

In 2020, a new regulation concerning the environmental bonus for the purchase of electric cars comes into effect and applies retrospectively as of 4 November 2019. The regulation was expected to apply until the end of 2025, but in 2021 an ammendment was done. In 2020, the environmental bonus defined is shown in the following tables:

Environmental bonus for BEV and FCEV in 2020					
Net list price under 40000€			**Net list price over 40000€**		
Federal share (€)	Manufacturer share (€)	**Total (€)**	Federal share (€)	Manufacturer share (€)	**Total (€)**
3000	3000	**6000**	2500	2500	**5000**

Table 8.1: Environmental bonus for battery electric vehicles and fuel cell electric vehicles in 2020 in Germany

Environmental bonus for PHEV in 2020					
Net list price under 40000€			**Net list price over 40000€**		
Federal share (€)	Manufacturer share (€)	**Total (€)**	Federal share (€)	Manufacturer share (€)	**Total (€)**
2250	2250	**4500**	1875	1875	**3750**

Table 8.2: Environmental bonus for plug-in hybrid electric in 2020 in Germany (maximum emission value of 50g of CO2/km or an electrical range of at least 40km)

BAFA (Federal government) guaranteed half of the subsidy by 2020 and the other half was guaranteed by car manufacturers. A minimum holding time of 6 months is required when purchasing and obtaining the bonus. In 2021, an amendment has been done where leasing subsidies have been included. Leasing cars are out of the scope of the study done and only purchase incentives are considered. In the course of the Corona crisis, the Federal Government's share of the environmental bonus for electric cars and plug-in hybrids was increased by the so-called innovation premium. The government doubled its bonus and the manufacturer subvention remained the same. It can be seen in the following tables:

Environmental bonus for BEV and FCEV in 2021					
Net list price under 40000€			**Net list price over 40000€**		
Federal share (€)	Manufacturer share (€)	**Total (€)**	Federal share (€)	Manufacturer share (€)	**Total (€)**
6000	3000	**9000**	5000	2500	**7500**

Table 8.3: Environmental for battery electric vehicles and fuel cell electric vehicles in 2021 and valid until 31st December 2025 in Germany

The new environmental bonus determined in the amendment in valid until end 2025. The promotion lasts for a maximum total of 400.000 cars.

Environmental bonus for PHEV in 2021					
Net list price under 40000€			**Net list price over 40000€**		
Federal share (€)	Manufacturer share (€)	**Total (€)**	Federal share (€)	Manufacturer share (€)	**Total (€)**
2250	2250	**4500**	1875	1875	**3750**

Table 8.4: Environmental bonus for plug-in hybrid electric in 2021 in Germany (Maximum emission value of 50g of CO2/km or an electrical range of at least 40km. This range requirement applies until 31 December 2021, subsequently it will be increased to 60km. From 1 January 2025 it will increase to 80km.)

8.2 Taxation in France

8.2.1 Regional component of the registration tax

Region	**Fiscal power cost by region (2022)**	**Exemption for so-called "clean" vehicles**
Île-de-France	46,15 €	100%
Centre-Val de Loire	49,80 €	50%
Bourgogne-Franche-Comté	51,00 €	100%
Normandie	35,00 €	100%
Hauts-de-France	33,00 €	100%
Grand-Est	48,00 €	100%
Pays de la Loire	48,00 €	100%
Bretagne	51,00 €	50%
Nouvelle-Aquitaine	41,00 €	100%
Occitanie	44,00 €	100%
Auvergne-Rhône-Alpes	43,00 €	100%
Provence-Alpes-Côte d'Azur	51,00 €	100%
Corse	27,00 €	100%
Guadeloupe	41,00 €	0%
Martinique	51,00 €	0%
Guyane	42,50 €	0%
La Réunion	51,00 €	0%
Mayotte	30,00 €	0%

Table 8.5: Value of the regional tax (NUTS1 level) of a CV (fiscal power unit) and percentage of exemption in 2022. Extracted from [57]

8.2.2 Ecological malus in France

The ecological malus tax rate applied in France is displayed in Table 8.6

CO_2 rate in g/km (approved under WLTP)	**Tax (€) 2021**	**Tax (€) 2022**
0	0	0
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0

38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0

79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	0	0
87	0	0
88	0	0
89	0	0
90	0	0
91	0	0
92	0	0
93	0	0
94	0	0
95	0	0
96	0	0
97	0	0
98	0	0
99	0	0
100	0	0
101	0	0
102	0	0
103	0	0
104	0	0
105	0	0
106	0	0
107	0	0
108	0	0
109	0	0
110	0	0
111	0	0
112	0	0
113	0	0
114	0	0
115	0	0
116	0	0
117	0	0
118	0	0
119	0	0

120	0	0
121	0	0
122	0	0
123	0	0
124	0	0
125	0	0
126	0	0
127	0	0
128	0	0
129	0	50
130	0	75
131	0	100
132	0	125
133	0	150
134	50	170
134	75	190
136	100	210
137	125	230
138	150	240
139	170	260
140	190	280
141	210	310
142	230	330
143	240	360
144	260	400
145	280	450
146	310	540
147	330	650
148	360	740
149	400	818
150	450	898
151	540	983
152	650	1074
153	740	1172
154	818	1276
155	898	1386
156	983	1504
157	1074	1629
158	1172	1761
159	1276	1901
160	1386	2049

161	1504	2205
162	1629	2370
163	1761	2544
164	1901	2726
165	2049	2918
166	2205	3119
167	2370	3331
168	2544	3552
169	2726	3784
170	2918	4026
171	3119	4279
172	3331	4543
173	3552	4818
174	3784	5105
175	4026	5404
176	4279	5715
177	4543	6039
178	4818	6375
179	5105	6724
180	5404	7086
181	5715	7462
182	6039	7851
183	6375	8254
184	6724	8671
185	7086	9103
186	7462	9550
187	7851	10011
188	8254	10488
189	8671	10980
190	9103	11488
191	9550	12012
192	10011	12552
193	10488	13109
194	10980	13682
195	11488	14273
196	12012	14881
197	12552	15506
198	13109	16149
199	13682	16810
200	14273	17490
201	14881	18188

202	15506	18905
203	16149	19641
204	16810	20396
205	17490	21171
206	18188	21966
207	18905	22781
208	19641	23616
209	20396	24472
210	21171	25349
211	21966	26247
212	22781	27166
213	23616	28107
214	24472	29070
215	25349	30056
216	26247	31063
217	27166	32094
218	28107	33147
219	29070	34224
220	30000	35324
221	30000	36447
222	30000	37595
223	30000	38767
224	30000	39964
225	30000	40000
226	30000	40000

Table 8.6: Ecological malus rate in 2021 and in 2022 in France. Data extracted from [43]

8.2.3 Scrapping scheme in France

The scrapping scheme (prime à la conversion) is a premium granted to car buyers for changing an old vehicle by a new one. The vehicles elegible for scrapping are:

- diesels registered before 2011
- petrol models registered before 2006

The subvention received depends on the income of the buyer and if it changes to an electric vehicle or a PHEV, a maximum of 5000€ and 2500€ respectively.

The scrapping scheme is not modelled due to the lack of information of which is the previous car the buyers of electric and PHEV had.

8.2.4 Annual malus in france

From 1 January 2009, passenger cars with high CO2 emissions registered in France were subject to €160 annual tax if they were considered to be high pollutant. The threshold was reduced until 2012 where a threshold value of 190 g/km was set. The evolution of the annual malus threshold can be seen in Table 8.7.

Year of first registration	CO2 emissions (g/km)
2009	250
2010	245
2011	245
From 2012 onwards	190

Table 8.7: Previous annual malus threshold emission value (until end of 2020)

In 2021, the annual malus was suppresed. It is no longer applied. [4]. It has been supressed due to the high increase in the registration taxes and for simplifying the ecological taxation according to the French Ministry of Public Affairs.

8.2.5 Taxation in Italy

8.2.6 Registration fees in Italy

Department of Land Transport fees	10,20 €
ACI fees	27,00 €
Stamp duty for registering the car at Automobile Public Register	32,00 €
Stamp duty for the emission of the registration certificate	32,00 €
Motor vehicle license plate	41,78 €
Total	**142,98 €**

Table 8.8: Registration fees in Italy at 2022

8.2.7 Provincial component of the registration tax in Italy

Once the national registration tax based on power is defined, different tax increase by region of the national registration is presented:

- **No increase**: Aosta , Bolzano, Trento
- **20% increase**: Arezzo, Avellino, Benevento, Regione Friuli Venezia Giulia, Grosseto, Latina, Reggio Emilia, Vicenza

- **25% increase**: Crotone, Sondrio, Ferrara
- **30% increase**: All the rest of regions

Ecological partial exemptions

Once the tax has been defined, there is an ecological exemption for certain types of cars in some regions (For more detail, consult [104] and [34]):

- Bari: A 75% of the IPT (national + provincial increase) needs to be paid for electric, hydrogen. methane and GLP cars
- Padova, Pesaro Urbino, Potenza, Ravenna and Ferma: There is a 20% provincial increase instead than a 30% for electric, hybrid, GLP and methane cars.
- Cremona and Padova: A 50% of the IPT needs to be paid for electric and hydrogen cars.
- Lodi: A 30% reduction of IPT for electric, hydrogen, methane and GLP cars.
- Salerno and Nuoro: A 25% reduction of the IPT for electric, hydrogen, methane and GLP cars.
- Lecco: A 20% reduction of the IPT for electric and hydrogen cars.
- Macerata: A 20% reduction of the IPT for electric, hydrogen, methane and GLP cars.
- Roma and Vicenza: There is no provincial increase (only national amount) for electric, hydrogen, methane and GLP cars.
- Napoli: Like Rome and Vicenza, there is no provincial increase for electric, hydrogen,methane and GLP cars. Additionally, there is no increase for hybrid cars.
- Brescia: A reduction of 50% to the provincial increase. Instead of having a 30% increase, electric, hydrogen, methane and GLP cars have a 15% increase.

8.2.8 Bonus malus scheme in Italy

National ecobonus

The funds assigned yearly depend on the level of emissions and can be seen in Table 8.9

It can be seen that funds assigned for electric vehicles and PHEV increase from 2022 to 2024 while the amount assigned for combustion engine vehicles decreases. The funds assigned are used for paying subventions in each of the categories. If the limit is reached, no more subventions are given.

Regional ecobonus

- L'Aquila (NUTS3): Capital of Abruzzo (NUTS2) has a subvention of 30% of the cost of the car (without VAT) with a maximum of 4.000€ for particulars. The subvention ends at the end of 2022. Data extracted from [13].
- Provinza Autonoma di Bolzano (NUTS2): Subvention of 2.000€ for electric vehicles and 1.000€ por PHEV. The period of the subvention is not specified. Data extracted from [89].

CO_2 (g/km)	**motorisation**	**funds 2022 (in millions of €)**	**funds 2023 (in millions of €)**	**funds 2024 (in millions of €)**
0-20	electric	**220**	**230**	**245**
21-60	plug-in hybrid electric	**225**	**235**	**245**
61-135	hybrid	**170**	**150**	**120**
	gasoline			
	diesel			
	methanol			
	GLP			

Table 8.9: Funds of the ecobonus assigned each year for period 2022 to 2024. Data extracted from [24]

- Valle d'Aosta: There are subvention for cars which cost less than 60.000€ (Without VAT and other taxes). If they emit less than 20 g/km 5.000€ subvention. If they emit between 20 g/km and 70 g/km 3.000€ subvention. The subvention ends at the end of 2022. Data extracted from [115].

8.2.9 Ownership tax in Italy

Ownership tax for new cars following Euro 6 need to pay can be consulted in Table 8.10.

Province	**Power <100 kW**	**Power >100 kW**
Valle d'Aosta/Vallée d'Aoste	2,58	3,87
Liguria	2,84	4,26
Emilia-Romagna	2,58	3,87
alt	3,12	4,69
Molise	2,76	4,14
Campania	3,12	4,69
Puglia	2,58	3,87
Basilicata	2,58	3,87
Calabria	2,84	4,26
Sicilia	2,58	3,87
Sardegna	2	2,58
Provincia Autonoma di Bolzano/Bozen	2,58	3,87
Provincia Autonoma di Trento	1,96	2,95
Veneto	2,58	3,87
Friuli-Venezia Giulia	2,58	3,87
Emilia-Romagna	2,58	3,87
Toscana	2,71	4,26

Umbria	2,58	3,87
Marche	2,79	4,18
Lazio	2,84	4,26

Table 8.10: Annual ownership tax rate in Italy defined by region and power for Euro 6 vehicles. Rate expressed in €/KW

8.3 Provincial exemptions of ownership tax in Italy

It has been shown the national exemptions and here the exemption of each region of the ownership tax is presented (All data has been collected from [101]):

- **Piemonte**

  There are no exemptions
- **Valle d'Aosta**:
  1. After the first 5 years of national exemption for electric vehicles first registration, there is an additional 75% exemption for 5 more years.
  2. After 2019, the three years after the first 5 years, there is a 100% exemption for electric vehicles. (Since January 2019 and still valid)
- **Liguria** There are no exemptions
- **Lombardia**:
  1. Hybrid cars have a 50% exemption during 5 years after first registration. (Since January 2019 and still valid)
- **Abruzzo**:
  1. 3 years of exemption for new hybrid vehicles after first registration. (Since January 2019 and still valid)
- **Molise**
  1. 75% exemption for electric vehicles after the first 5 years of complete exemption. (revised in 2016 and still valid)
- **Campania**
  1. 3 years of exemption for new hybrid vehicles after first registration. (Since 2014 and still valid)
- **Puglia**
  1. Exemption for GLP and methane vehicles during first 5 years.(from January 2013 and on)
- **Basilicata**

    - Exemption for GLP and methane vehicles during first 5 years. (from January 2013 and on)

- **Calabria**:
    1. 75% exemption for electric vehicles after the first 5 years of complete exemption. (In 2019 was valid and still valid)
- **Sicilia**
    1. No longer valid (It was valid from 2019 to 2021) 3 years of exemption for hybrid vehicles after first registration.
- **Sardegna**

  There are no exemptions
- **Provinzia Autonoma di Bolzano**
    1. 75% exemption for electric vehicles after the first 5 years of complete exemption. (Still valid)
- **Provincia Autonoma di Trento**
    1. Exemption for GLP and methane vehicles during first 5 years. (Since 2010 and still valid)
    2. 75% exemption for electric vehicles after the first 5 years of complete exemption.
    3. Hydrogen cars have a 5 year exemption after first registration (Since 2013 and still valid)
- **Veneto**
    1. 3 years of exemption for hybrid vehicles after first registration. (Since 2014)
- **Friuli-Venezia Giulia**
    1. 75% exemption for electric vehicles after the first 5 years of complete exemption. (Still valid-Article 20 of the Consolidated Law on Motor Vehicle Taxes n.39 / 195)
- **Emilia-Romagna**
    1. 75% exemption for electric vehicles after the first 5 years of complete exemption. (Still valid-Article 20 of the Consolidated Law on Motor Vehicle Taxes n.39 / 195)
- **Toscana**

  There are no exemptions
- **Umbria**

  There are no exemptions
- **Marche**
    1. 5 years of exemption for hybrid vehicles after first registration. (Valid until 2022)
    2. 75% exemption for electric vehicles after the first 5 years of complete exemption. (Still valid-Article 20 of the Consolidated Law on Motor Vehicle Taxes n.39 / 195)

- **Lazio**
    1. 3 years of exemption for hybrid vehicles after first registration. (Since 2014)

8.4 Taxation in Netherlands

8.4.1 Motor Vehicle tax in Netherlands

8.4.1.1 Provincial rates of the motor vehicle tax

The motor vehicle varies between each province and the percentage which is applied at each region respect to a fixed tariff can be seen in Table 8.11.

Drenthe	92,0%
Flevoland	82,2%
Friesland	87,0%
Gelderland	90,6%
Groningen	94,5%
Limburg	77,9%
North Brabant	79,6%
North Holland	67,9%
Overijssel	79,9%
Utrecht	77,5%
Zealand	82,3%
South Holland	91,8%

Table 8.11: Provincial rates for motor vehicle tax. Extracted from [110]

8.4.1.2 Surcharge depending on the type of fuel

Surcharge rates can be found in Article 23 in the Law Motor Vehicle Tax Act 1995 which is valid in 2022 [91]. Here a summary is presented in order to make it simplier.

8.4.1.3 Surcharge for diesel

Diesel surcharge does not depend on the region. It only depends on the weight and the following tariff can be observed in Table 8.12.

Weight (kg)	**Diesel surcharge (€)**
0 to 550	69,16
551 to 650	81,86
651 to 750	94,55
751 to 850	107,48
851 to 950	125,77
Each additional 100 kg	13,62 € extra to previous tariff

Table 8.12: Diesel surcharge depending on the weight of the vehicle

There is an additional fine dust surcharge for diesel vehicles if they accomplish any of the following conditions:

- The particulate matter emissions exceed 5 milligrams per kilometre.
- the particulate matter emissions exceed 10 milligrams per kilowatt hour.
- The particulate matter emissions referred to in the fourth paragraph are not registered in the vehicle registration register and the date of first admission, stated in the vehicle registration register, is before 1 September 2009, or if it is registered in the vehicle registration register that the particulate filter has been removed

The additional fine dust surcharge can be consulted in [6].

8.4.1.4 Surcharge for liquified petrol gas and others

As for diesel, LPG and other type of fuels surcharge does not depend on the region. It only depends on the weight and the following tariff can be observed in Table 8.13.

Weight (kg)	**LPG and others surcharge (€)**
0 to 550	81,14
551 to 650	97,26
651 to 750	113,4
751 to 850	129,49
851 to 950	141,45
Each additional 100 kg	14,98 € extra to previous tariff

Table 8.13: LPG and others surcharge depending on the weight of the vehicle

8.4.2 Ownership tax estimation in Netherlands

The lack of information in how the tax is calculated has made that tariffs have been consulted in [23] for each weight category and for 2 regions. The calculus done for obtaining the different values is presented here.

Gasoline

There is a lack of information in the ownership tax calculation. Then, tariffs have been consulted manually in [23] for each weight category and for 2 regions. Using this method, the following system of equations shown in Equation 8.1 is defined:

$$\begin{cases} x + y \cdot (province\,1\,rate\,\%) = tax\,province\,1 \\ x + y \cdot (province\,2\,rate\,\%) = tax\,province\,2 \end{cases} \tag{8.1}$$

Where x is the national tax rate for a determined weight and y the provincial tax rate for a certain provincial rate. It is possible to obtain values for all the rest of the regions using this coefficients. It has been checked that results obtained are equal and the taxation has been obtained with a precision of ±1 €.

Diesel

Diesel tariffs for each weight group can be calculated by the final gasoline price and an additional surcharge. Diesel cars pay more than triple than gasoline in the case of light cars (lower than a 1000 kg) and closer to double cars which are 3000 kg.

Additionally, there is a surcharge which affects only diesel vehicles. The fine dust surcharge is additionally included and affects normally to older diesel vehicles. Both surcharge tariffs can be consulted in Appendix 8.4.1.3.

Natural gas

It pays the same as gasoline until 850 kg. Cars which are heavier pay a surcharge of 15.65 € per each 100 additional kg.

Liquified petrol gas (LPG) and other

LPG and others pay an additional tariff respect to gasoline cars. The surcharge works equal than diesel and it is even higher.

8.5 Taxation in Norway

8.5.1 Motor Vehicle tax in Norway

The road traffic insurance tax which corresponds to the ownership tax paid in Norway has modified its amount considerably in recent years. Before March 2021, electric and hydrogen did not pay this tax. In 2021, they paid a reduced tax compared to internal combustion engine and since march 2022 they pay the same amount. The different amounts for the present year and previous years can be seen in Table 8.14

year	type of fuel	price in NOK per day	price in € per year
Mar-20	gasoline	8,12	296,53 €
Mar-21	gasoline	8,4	306,75 €
Mar-22	gasoline	8,15	297,62 €
Mar-20	diesel	8,12	296,53 €
Mar-21	diesel	8,4	306,75 €
Mar-22	diesel	8,15	297,62 €
Mar-20	others	8,12	296,53 €
Mar-21	others	8,4	306,75 €
Mar-22	others	8,15	297,62 €
Mar-20	diesel without particle filter	9,47	345,83 €
Mar-21	diesel without particle filter	9,8	357,88 €
Mar-22	diesel without particle filter	9,57	349,48 €
Mar-20	electric	0	- €
Mar-21	electric	5,85	213,63 €
Mar-22	electric	8,15	297,62 €
Mar-20	hydrogen	0	- €
Mar-21	hydrogen	5,85	213,63 €
Mar-22	hydrogen	8,15	297,62 €

Table 8.14: Road traffic insurance tax before March 2021 (Mar-20), between March 2021 and March 2022 (Mar-21) and after March 2022. Extracted from [127]

8.6 Taxation in Poland

8.6.1 Registration fees in Poland

	cost in PLN	cost in €
Registration fees	**240,50 PLN**	**52,87 €**
registration document	85,00 PLN	18,69 €
license plate fees	80,00 PLN	17,59 €
issuance of registration certificate	0,50 PLN	0,11 €
vehicle card fee	75,00 PLN	16,49 €

Table 8.15: Registration charges

8.7 Taxation in Greece

8.7.1 Registration tax in Greece

8.7.1.1 Example of calculation of registration tax in Greece

If a passenger car retail price before taxes is 24,000 € and has a level of emissions of 135 g/km (WLTP), the calculation will be the following:

- CO_2 emissions coefficient: 110% (Extracted from Table 4.17)
- calculation with tax rates for different retail price groups (Extracted from Table 4.16):

$$14,000 \cdot 4\% + 3,000 \cdot 8\% + 3,000 \cdot 16\% + 4,000 \cdot 24\% = 2,464 \; euros$$

- Using equation 4.3, the registration tax obtained can be calculated:

$$2,464 \; euros \cdot 110\% = 2,710.4 \; euros$$

8.8 Fuel and Energy prices

8.8.1 Scenarios of IEA [144]

The scenarios proposed by IEA are:

- **The net zero emissions by 2050 Scenario (NZE):** This is a normative IEA scenario that shows a narrow but achievable pathway for the global energy sector to achieve net zero CO2 emissions by 2050.

  This scenario most probably will not be achieved considering the current policies and goals announced by the different countries. This scenario is discarded.
- **The Announced Pledges Scenario (APS)**: It appears for the first time in the WEO 2021. It takes account of all of the climate commitments made by governments around the world, as well as longer term net zero targets, and assumes that they will be met in full and on time.

  This scenario assumes that countries will accomplish their goals even before it is defined what will be the path to accomplish this. This scenario is discarded.
- **Stated Policies Scenario (STEPS)**: It reflects current policy settings based on specific policies that are in place, as well as those that have been announced by governments around the world. It provides a more conservative benchmark for the future, because it does not take it for granted that governments will reach all announced goals.

  This scenario takes into account the current policies which are in use and the future policies if they have been officially announced and declared. STEPS scenario is used in the thesis to see which would be the future European automotive fleet considering the current policy settings.

- **Sustainable development scenario (SDS)**: It achieves key energy-related United Nations Sustainable Development Goals related to universal energy access and major improvements in air quality, and reaches global net zero emissions by 2070.

  In this scenario, all current net zero pledges are achieved in full and there are extensive efforts to realise near-term emissions reductions; advanced economies reach net zero emissions by 2050, China around 2060, and all other countries by 2070 at the latest.

  This scenario is more plausible than NZE because it does not consider that all the world has net zero emissions by 2050. Europe will have net zero emissions by 2050, but other regions of the world would reach it later. SDS scenario could be used because it shows how electricity and fuel prices will evolve if additional measures are taken to reach net zero emissions by 2050.

8.8.2 Correlations of gasoline and diesel prices

Gasoline and diesel prices analysis is explained in Section 4.2.2. Here, correlations of gasoline and diesel prices for the different countries are plotted just in case the reader wants to consult them.

Correlations of gasoline prices

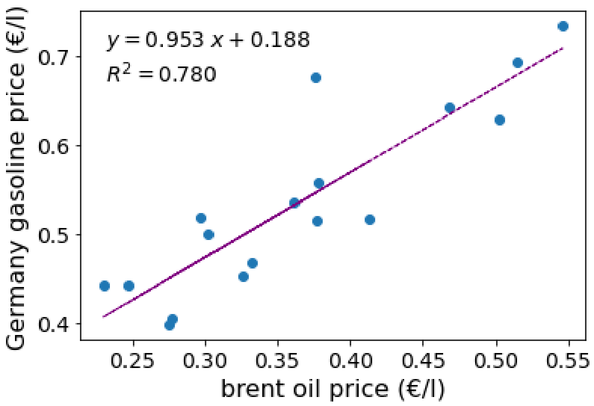


Figure 8.1: Correlation between average price of gasoline in Germany and Brent barrel oil price between 2005 and 2021

.

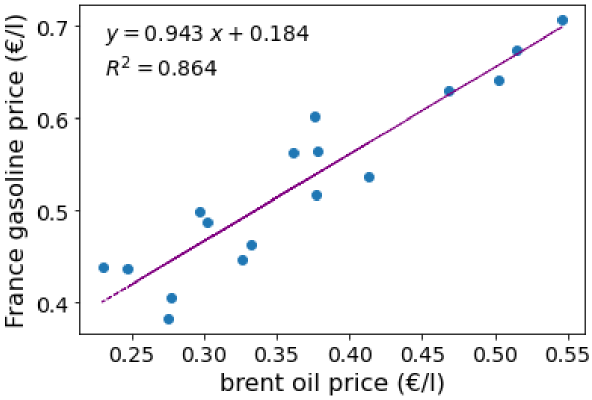


Figure 8.2: Correlation between average price of gasoline in France and Brent barrel oil price between 2005 and 2021

.

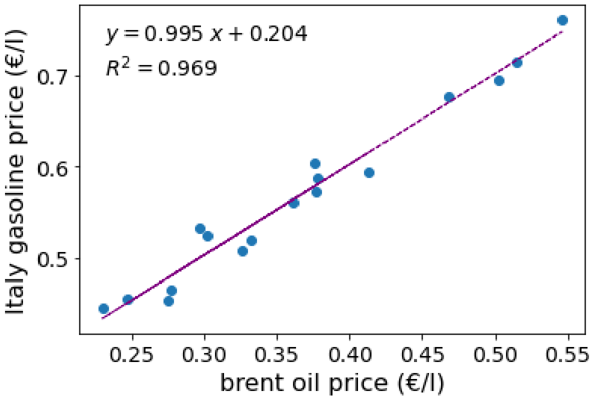


Figure 8.3: Correlation between average price of gasoline in Italy and Brent barrel oil price between 2005 and 2021

.

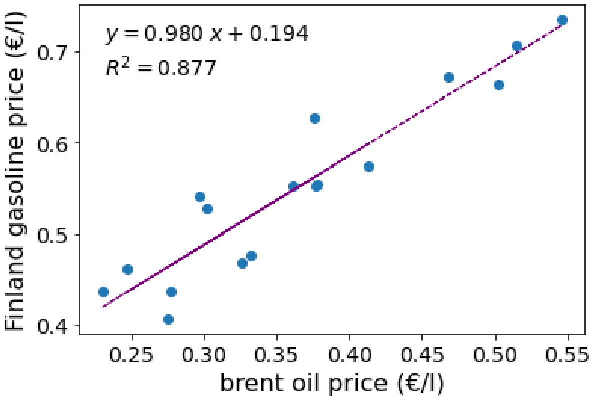


Figure 8.4: Correlation between average price of gasoline in Finland and Brent barrel oil price between 2005 and 2021

.

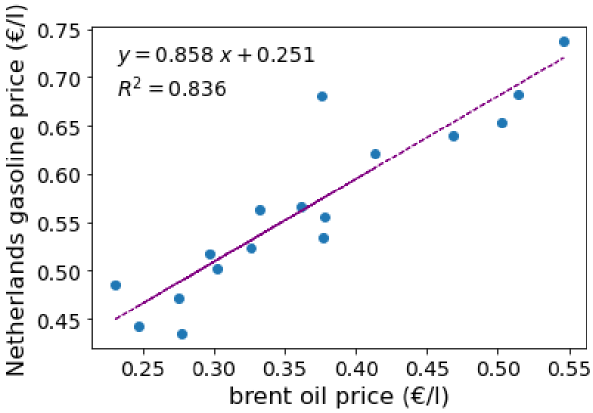


Figure 8.5: Correlation between average price of gasoline in Netherlands and Brent barrel oil price between 2005 and 2021

.

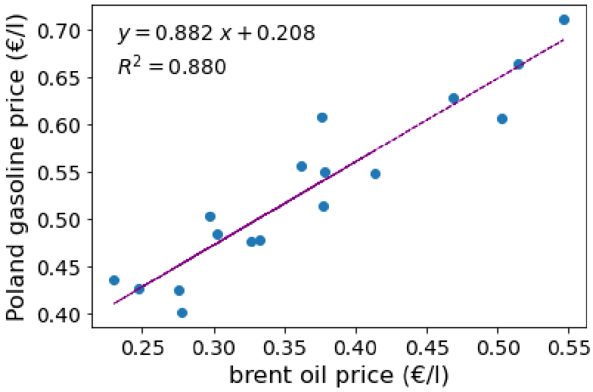


Figure 8.6: Correlation between average price of gasoline in Poland and Brent barrel oil price between 2005 and 2021

.

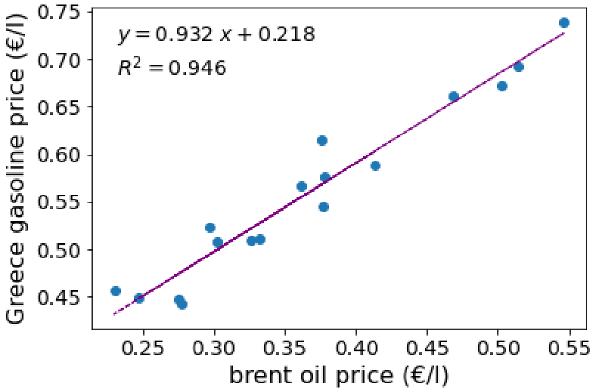


Figure 8.7: Correlation between average price of gasoline in Greece and Brent barrel oil price between 2005 and 2021

.

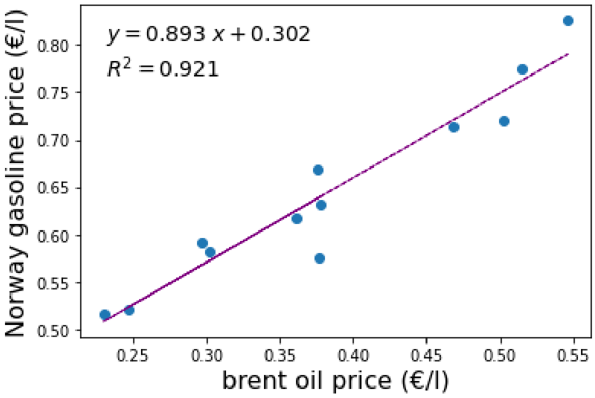


Figure 8.8: Correlation between average price of gasoline in Norway and Brent barrel oil price between 2005 and 2021

.

Once correlation between barrel price and gasoline price for each country is plotted, Table 8.16 shows the R^2 for each country and its linear regression equation.

Country	R^2	linear regression equation
Germany	0,78	y=0,953x+0,188
France	0,864	y=0,943x+0,184
Italy	0,969	y=0,995x+0,204
Finland	0,877	y=0,980x+0,194
Netherlands	0,836	y=0,858x+0,251
Poland	0,88	y=0,882x+0,208
Greece	0,946	y=0,932x+0,218
Norway	0,921	y=0,893x+0,302

Table 8.16: R^2 and linear regression equation of the price of gasoline in each country respect price of the brent barrel

Correlations of diesel prices

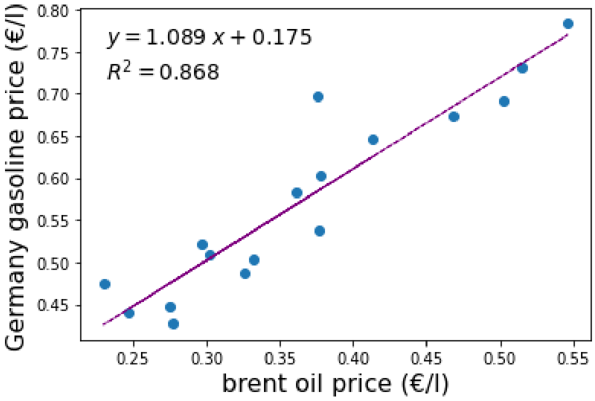


Figure 8.9: Correlation between average price of diesel in Germany and Brent barrel oil price between 2005 and 2021

.

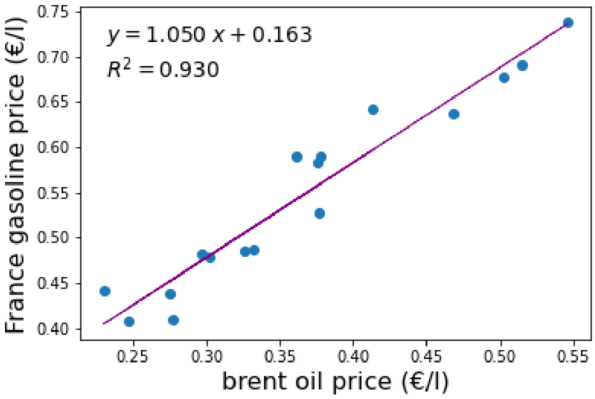


Figure 8.10: Correlation between average price of diesel in France and Brent barrel oil price between 2005 and 2021

.

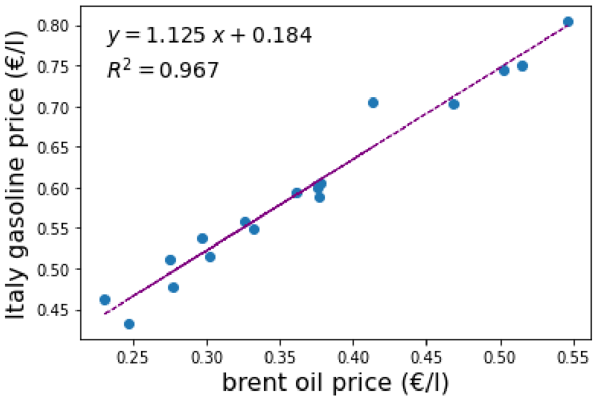


Figure 8.11: Correlation between average price of diesel in Italy and Brent barrel oil price between 2005 and 2021

.

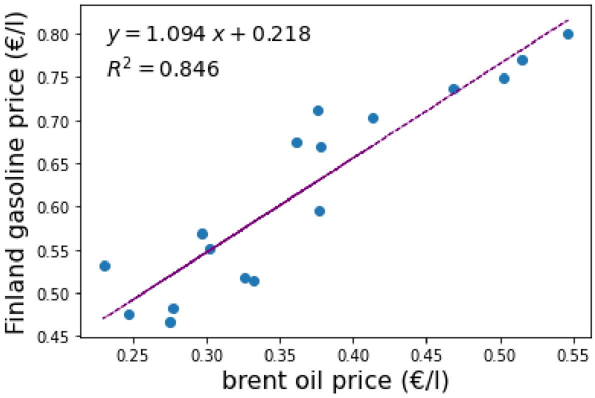


Figure 8.12: Correlation between average price of diesel in Finland and Brent barrel oil price between 2005 and 2021

.

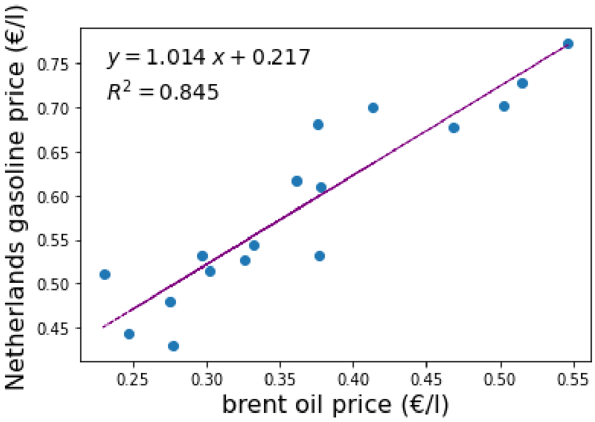


Figure 8.13: Correlation between average price of diesel in Netherlands and Brent barrel oil price between 2005 and 2021

.

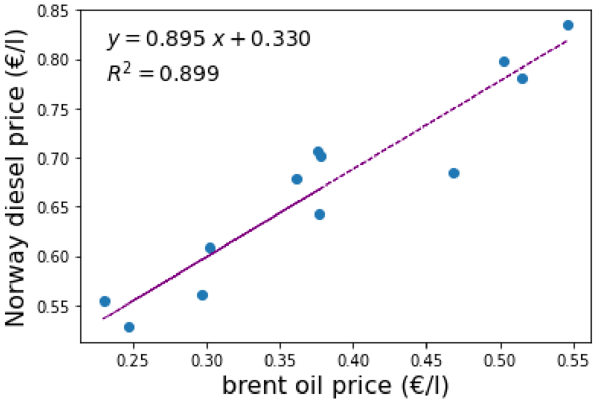


Figure 8.16: Correlation between average price of diesel in Norway and Brent barrel oil price between 2005 and 2021

.

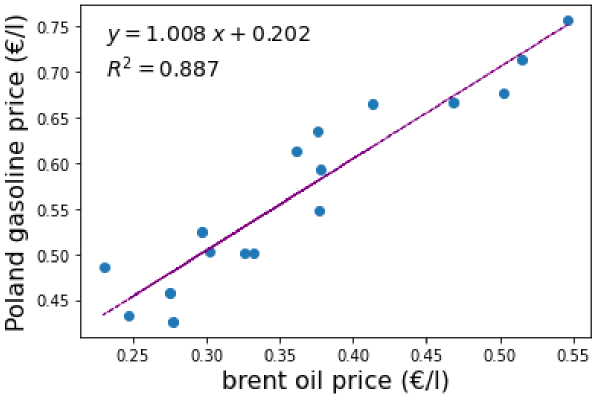


Figure 8.14: Correlation between average price of diesel in Poland and Brent barrel oil price between 2005 and 2021

.

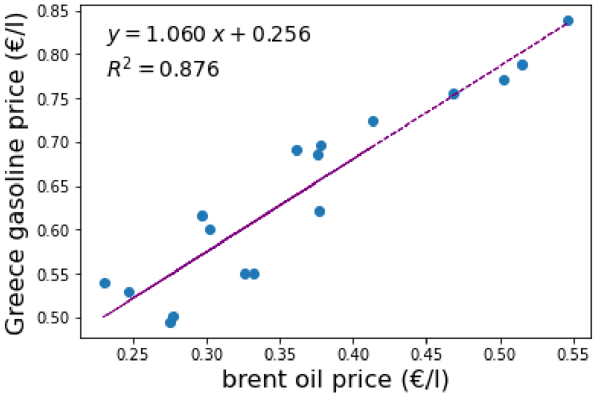


Figure 8.15: Correlation between average price of diesel in Greece and Brent barrel oil price between 2005 and 2021

.

Once correlation between barrel price and gasoline price for each country is plotted, Table 8.17 shows the R^2 for each country and its linear regression equation.

Both gasoline and diesel prices in all the countries studied present a very high R^2. The lowest is Germany gasoline price which is 0.78. A high correlation of gasoline and diesel prices with the brent barrel prices allows to use the linear regression equations to obtain the price of gasoline or fuel directly from the brent barrel price.

8.8.2.1 Time-dependence of gasoline and diesel regressions

The correlation of gasoline and crude oil price considering years can be observed in Figure 8.17.

Country	R^2	linear regression equation
Germany	0,868	y=1,089x+0,175
France	0,93	y=1,050x+0,163
Italy	0,967	y=1,125x+0,184
Finland	0,846	y=1,094x+0,218
Netherlands	0,845	y=1,014x+0,217
Poland	0,887	y=1,008x+0,202
Greece	0,876	y=1,060x+0,256
Norway	0,899	y=0,895x+0,330

Table 8.17: R^2 and linear regression equation of the price of diesel in each country respect price of the brent barrel

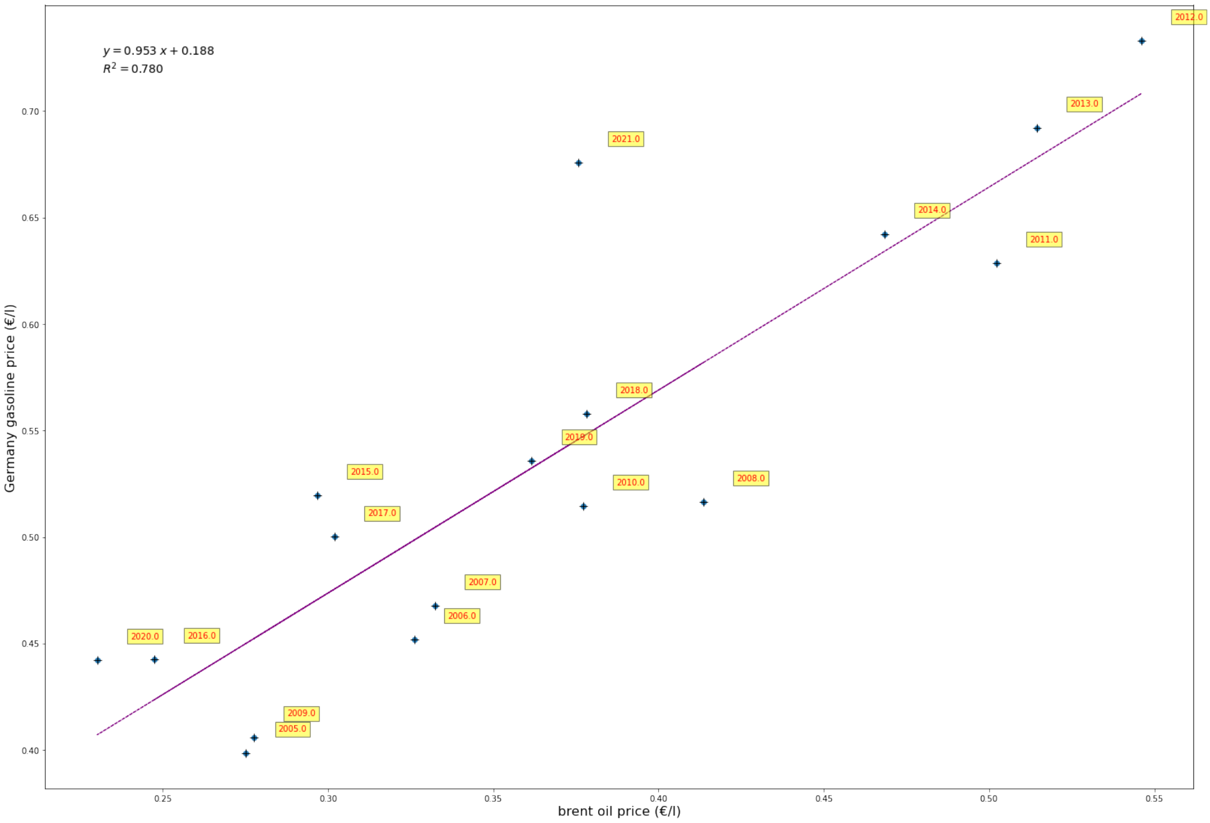


Figure 8.17: Correlation between average price of gasoline before taxes in European Union and Brent barrel oil price between 2005 and 2021 including years

8.8.3 Compressed natural gas price

Price of compressed natural gas price depends on different factors. Natural gas is used for households, for industry and for transportation. Natural gas price varies from country to country depending on the needs of natural gas each country has and the amount of electricity is able to produce with other energy sources.

It is important to remark that the amount of cars using CNG represents a 0,51% of the total EU fleet in 2021 and it has not increased in the past years; the new registrations of CNG in Europe in 2021 was 0.34% [118]. Finland and Italy are the only countries with a stock of CNG passenger cars around 2%. However, Finland presents very low new registrations of CNG, 0.37% while Italy still presents CNG passenger car new registrations above 2% (2.26% in 2021).

CNG technology is not a new technology and it is not treated as a clean technology in most countries and there are no plans of expanding the infrastructure of CNG across Europe. This makes that the relevance of CNG cars is relatively low in this study.

8.8.3.1 Components of CNG price

The price of natural gas in the EU depends on a range of different supply and demand conditions, including the geopolitical situation, the national energy mix, import diversification, network costs, environmental protection costs, severe weather conditions, and levels of excise and taxation [93].

The base price of natural gas is defined in Europe by TTF Dutch natural gas price. In U.S., for example, the base price is Henry Hub natural gas price.

However, the final price of the natural gas price before taxes in each country depends on the amount each country needs, the infrastructure for the transport and storage and many other factors.

It has not been possible to find a price breakdown of the CNG and the most similar price is the natural gas price for households which have a consumption between 20GJ and 200 GJ. Comparing the final price of CNG in fuel stations with the price of natural gas for households in the range of 20GJ-200GJ, it has been observed a similar price. Then, it is assumed that this price will be the same as the one of CNG [54].

Price of natural gas is again separated into two main components:

- **price of fuel before taxes**: Price of fuel before taxes depends first on the TTF Dutch natural gas price. It is determined in the exchange market. Future price of Dutch TTF natural gas is estimated by IEA under the different scenarios presented in 4.2.1.

  Additionally, there are other costs affecting the price of fuel before taxes. In the data provided, the natural gas price before taxes is divided into:

  1. **energy and supply**: It is related with the national energy mix and the demand each country has and the import diversification.
  2. **network costs**: It is related with costs of transportation, the infrastructure and the storage.

- **taxes**: There are different components in each country which determine the price of natural gas. Some countries have all the taxes and others do not have all of them. The most relevant groups in European Union are:

  1. **renewable taxes**

2. **capacity taxes**
3. **environmental taxes**
4. **VAT**: It is variable rate in comparison with the previous taxes which are a fixed value independent of the price of natural gas before taxes. VAT was in almost all countries constant, but in 2021 and 2022, the pandemic and the war of Ukraine have generated price spikes in natural gas and governments of Poland, Greece and Italy [76] have reduced VAT temporarily.

8.8.3.2 Correlation between Dutch TTF natural gas price and natural gas price before taxes in each country

A linear correlation between the TTF natural gas price and the natural gas price before taxes is considered. There is data of average yearly Dutch TTF natural gas price [56] and there is natural gas price before taxes from 2008 to 2021 available for 6 countries [53].

There is no data for Finland. but there is data for Sweden. It has not been possible to find data for Finland. It has been found the price of natural gas for Sweden and Finland and in 2021 they were very similar. The countries are close to each other geographically and any of both present natural gas resources. Then, it has been assumed the price of natural gas and of CNG is the same in Finland and Sweden.

There is no data available for Norway. Norway has a 0.00% of its fleet which uses CNG and new sales of CNG cars are 0.00%. It is assumed the price of Germany which it is lower than the price of Norway after applying taxes. Even with a lower price of CNG than in reality, it is expected a 0% of the new sales of CNG in Norwegian market.

Once all data is gathered, regressions are performed.

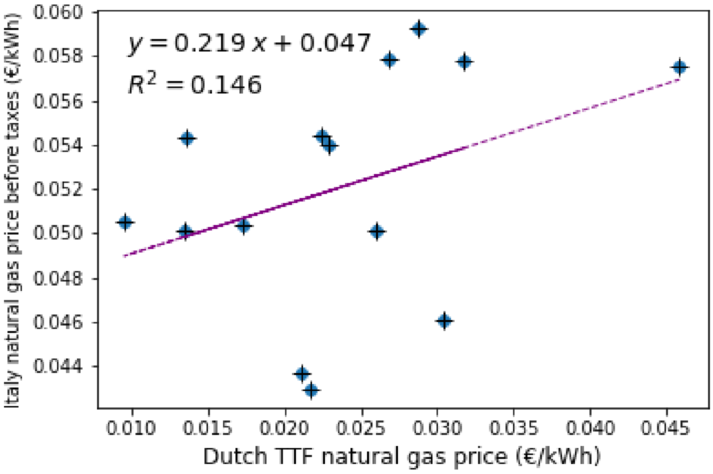


Figure 8.18: Correlation between average price of natural gas before taxes in Italy and Dutch TTF natural gas price between 2008 and 2021

In Figure 8.18, Italy natural gas price before taxes is analyzed. It is observed very spread data points and the R^2 is very low. However, with the data available it was difficult to obtain anything better. R^2 for all countries and its regression can be consulted in Appendix 8.8.4. R^2 obtained is low for all countries.

A deeper study should be done, but the complexity of the estimation, the lack of data and the relative low importance of CNG in European market allows to accept this results.

Price of natural gas before taxes can be obtained up to 2050 using the Dutch TTF natual gas price estimated under IEA scenarios and the regression parameters obtained for each country. Then, fixed taxes (renewable taxes, capacity taxes, environmental taxes and others) are assumed to remain constant with the value of 2020 or 2021. Finally, VAT is applied to the price of natural gas before taxes and the fixed taxes. Prices of CNG are estimated from 2020 to 2050.

8.8.4 Correlations of natural gas prices

Natural gas prices analysis is explained in Section 8.8.3. Here, correlations of natural gas prices for the different countries are plotted just in case the reader wants to consult them.

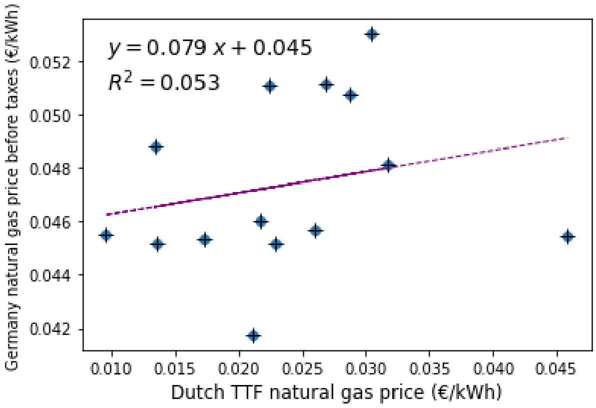


Figure 8.19: Correlation between average price of natural gas before taxes in Germany and Dutch TTF natural gas price between 2008 and 2021

.

Once correlations for each country are plotted, Table shows the R^2 for each country and its linear regression equation.

Natural gas prices before taxes presents a very low R^2. A different procedure could be done, but the complexity of the estimation, the lack of data and the relative low importance of CNG in European market allows to accept this results.

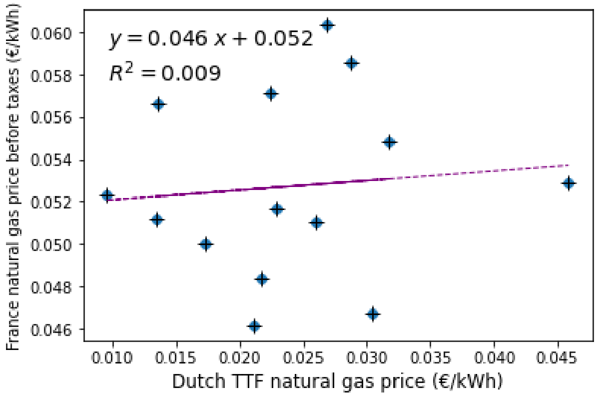


Figure 8.20: Correlation between average price of natural gas before taxes in France and Dutch TTF natural gas price between 2008 and 2021

.

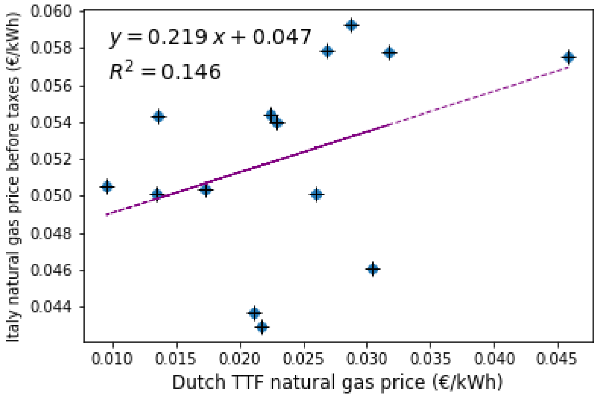


Figure 8.21: Correlation between average price of natural gas before taxes in Italy and Dutch TTF natural gas price between 2008 and 2021

.

Country	R^2	**linear regression equation**
Germany	0,009	y=0,046x+0,052
France	0,529	y=0,6501x+0,039
Italy	0,149	y=0,1529x+0,040
Finland	0,053	y=0,079x+0,045
Netherlands	0,08	y=0,089x+0,035
Poland	0,36	y=0,8651x+0,047
Greece	0,146	y=0,2191x+0,047
Norway	0,053	y=0,079x+0,045

Table 8.18: R^2 and linear regression equation of the price of natural gas price before taxes in each country respect price of Dutch TTF natural gas

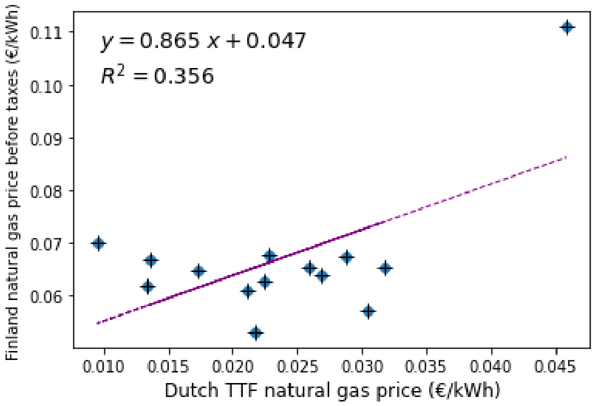


Figure 8.22: Correlation between average price of natural gas before taxes in Finland and Dutch TTF natural gas price between 2008 and 2021

.

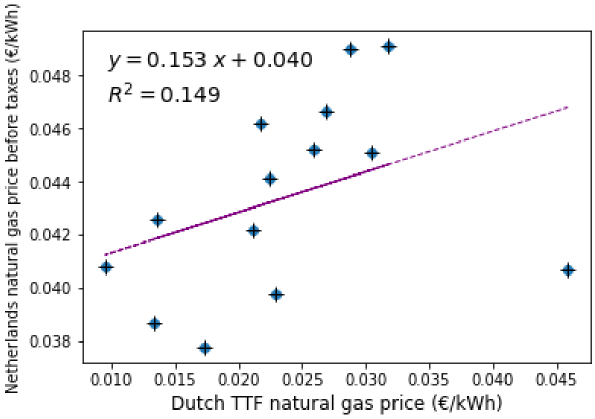


Figure 8.23: Correlation between average price of natural gas before taxes in Netherlands and Dutch TTF natural gas price between 2008 and 2021

.

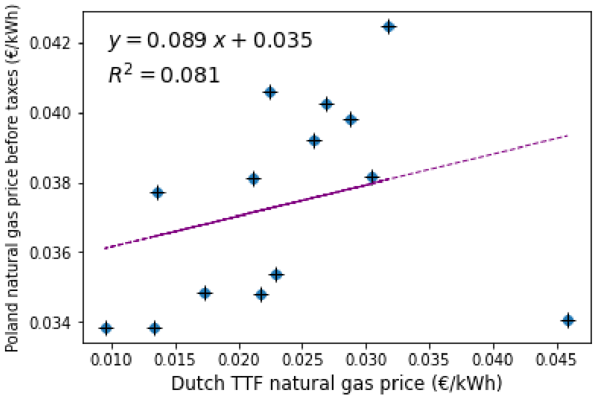


Figure 8.24: Correlation between average price of natural gas before taxes in Poland and Dutch TTF natural gas price between 2008 and 2021

.

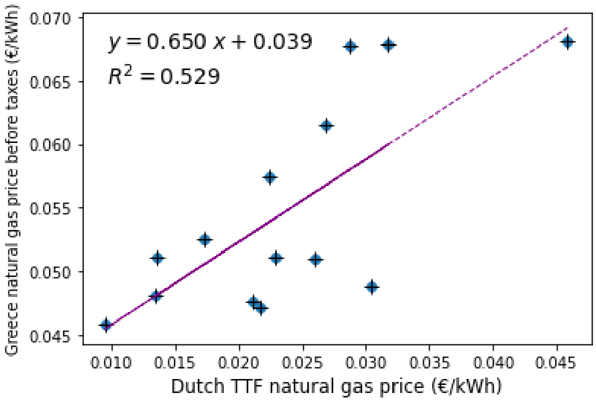


Figure 8.25: Correlation between average price of natural gas before taxes in Greece and Dutch TTF natural gas price between 2008 and 2021

.

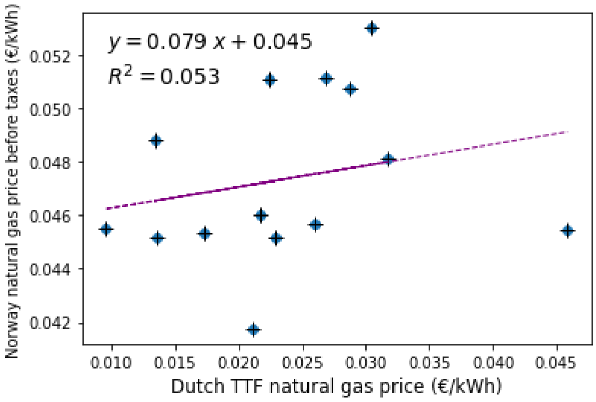


Figure 8.26: Correlation between average price of natural gas before taxes in Norway and Dutch TTF natural gas price between 2008 and 2021

.

8.9 Income distribution

8.9.1 Income distribution first decile, quantiles and last decile

Here it is presented the median of the quantiles of the national disposable income respect to average disposable income. These ratios are considered to be constant at a NUTS2 and NUTS3 level. In order to obtain the disposable income distribution value it is only necessary to multiply by the correspondant regional (NUTS2 or NUTS3) average disposable income value by the deciles and quantiles percentage of the table.

Country	first decile	first quintile	second quintile	third quintile	fourth quintile	fifth quintile	tenth decile
Finland	29%	35%	60%	86%	113%	162%	192%
France	31%	40%	62%	82%	106%	161%	200%
Germany	23%	30%	60%	85%	114%	171%	204%
Greece	20%	30%	62%	80%	109%	174%	209%
Italy		35%	59%	81%	117%	183%	
Netherlands	29%	37%	60%	84%	115%	168%	201%
Norway		34%	54%	78%	112%	164%	
Poland	24%	32%	60%	81%	111%	172%	201%

Table 8.19: Median of quantile disposable household income distribution respect to average disposable income

Norway and Italy data was not found in Eurostat source [123]. They have been found in national statistic institutes. Data for Norway can be consulted in [1] and for Italy in [130]

8.9.2 mean disposable income by degree of urbanisation

The mean equivalised disposable income by degree of urbanisation is obtained. The total mean equivalised disposable income of each country is given as well. Using both terms, the relative disposable income by degree of urbanisation is calculated and presented in Table 8.20

8.9.3 Population by degree of urbanisation at a national level

The population by degree of urbanisation at a national level can be calculated and the following Table which is split into Table 8.21 and Table 8.22 is shown.

Year 2020	cities	towns and suburbs	rural
NUTS nomenclature	1	2	3
Finland	108%	97%	93%
France	104%	97%	98%
Germany	100,3%	100,5%	98,4%
Italy	105%	100%	91%
Netherlands	98%	104%	100%
Norway	104%	99%	96%
Poland	116%	102%	86%
Greece	109%	104%	86%

Table 8.20: Relative mean disposable income by degree of urbanisation respect to total mean disposable income

degree of urbanisation	Finland	Germany	France	Netherlands
urban	30%	44%	35%	74%
towns and suburbs	30%	41%	37%	25%
rural	39%	16%	28%	1%
total population	100%	100%	100%	100%

Table 8.21: Population by degree of urbanisation relative to total population (First group of countries). Extracted from [14]

degree of urbanisation	Greece	Norway	Italy	Poland
urban	45%	13%	48%	25%
towns and suburbs	24%	66%	42%	39%
rural	31%	21%	10%	36%
total population	100%	100%	100%	100%

Table 8.22: Population by degree of urbanisation relative to total population (Second group of countries). Extracted from [14]

8.9.4 Methodology proposed for obtaining income distribution at a NUTS3 level

The mean disposable income can be obtained by degree of urbanisation at a national level (Extracted from [86]). It is separated using metropolitan areas, urban and rural areas. This are the 3 categories that exist in NUTS3 (NUTS system is explained in Section 3.1.1). An average of all regions is given at a national level (See in Appendix 8.9.2). For most countries, metropolitan areas have higher disposable income than urban areas and urban areas have higher disposable income than rural areas. For Netherlands and Germany this is not accomplished.

Once this ratios are obtained, there could be applied for obtaining NUTS3 average disposable income from NUTS2 average disposable income.

For ITC1 (NUTS2 Italy region code), there are 8 smaller subregions (at a NUTS3 level), which some are metropolitan, some urban and some rural. The NUTS3 disposable income could be calculated by applying the coefficients by degree of urbanisation (See in Appendix 8.9.2), but this would not be correct. It is not correct because if there is a NUTS2 region only composed of rural subregions multyplying by the rural coefficient would reduce the average disposable income of the whole NUTS2 region. In the case of Italy, this would diminish more than it should all regions which are mostly rural and increase the regions which are mostly metropolitan. If a region is mostly rural or mostly metropolitan, the average value of the region is already considering this phenomena, so it should not be accounted twice.

For solving this issue, it is decided to calculate the population by degree of urbanisation at a national level and for each NUTS2 region. (Data of population at a NUTS3 level is extracted from [14]).

Once the population by degree of urbanisation is obtained, if the composition of a region is exactly the same than the composition by degree of urbanisation at a national level, then, the relative disposable income by degree of urbanisation shown in Appendix 8.9.3 is directly used.

However, normally this is not the case. Then, in order to decide which disposable income would have each subregion (NUTS3 region), a line is adjusted for each degree of urbanisation (metropolitan, urban and rural). The conditions of the adjusted line are the following:

- If there is the same % of population at regional level that at national level for a type of urbanisation, then the % of disposable income respect to the total is the same as the one at the national level.
- If there is 100% of a certain type of urbanisation (for example, urban), then the disposable income at NUTS2 is the same one as for the subregions.

One line is adjusted for each type of urbanisation (one for metropolitan, one for urban and one for rural). In order to make it more visually, an example is done for the urban case in Italy. Using the data for Italy Figure 8.27 is obtained.

The coefficients of the adjusted line m and n are:

$$m = \frac{100 - national\ disp.\ income\ by\ type\ of\ urb}{100 - \%\ of\ national\ population\ by\ type\ of\ urb.} \tag{8.2}$$

$$n = 100 \cdot (1 - m) \tag{8.3}$$

The equations for calculating the coefficients have been defined. In the case of Italy, the coefficients are presented in Table 8.23.

Once the coefficients are defined, the disposable income at a NUTS3 level for any of the countries can be calculated for each region using the equation of the line.

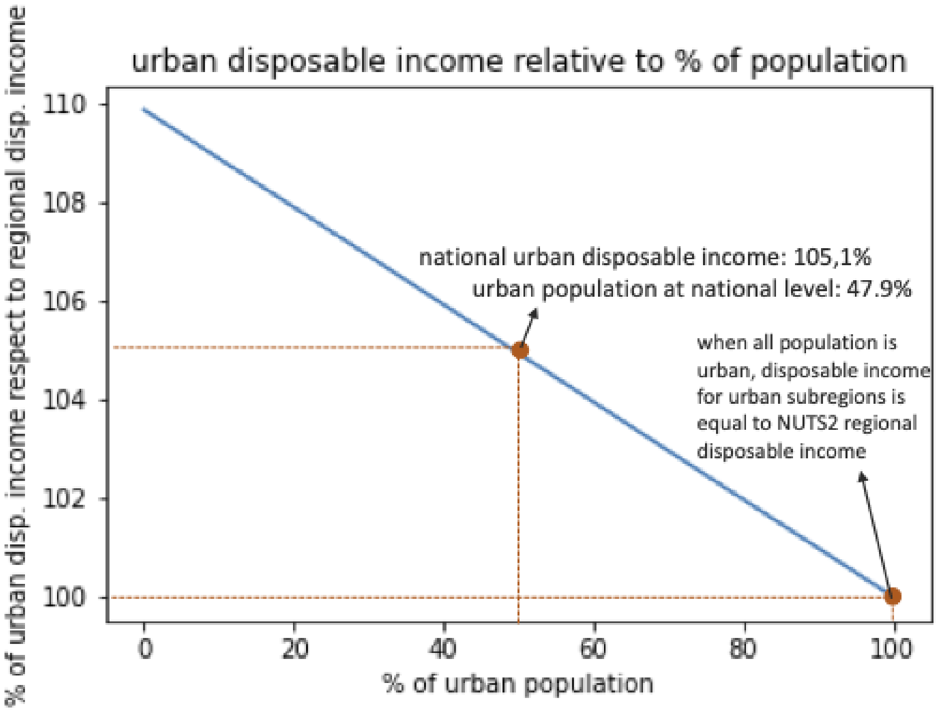


Figure 8.27: Adjusted line for calculating the metropolitan disposable income at a NUTS3 level in Italy depending on the degree of urbanisation of a NUTS2 region.

	m	n
metropolitan	- 0,10	1,10
urban	0,01	0,99
rural	0,10	0,90

Table 8.23: Coefficient of adjusted lines

$$disp.\ income\ by\ type\ of\ urb. = m \cdot (\%\ of\ population\ by\ type\ of\ urb. + n) \tag{8.4}$$

After the % of disposable income for each subregion (NUTS3 region) respect to the NUTS2 regional disposable income is obtained, a final step is required. The sum of total NUTS3 incomes is calculated and it can not be bigger or smaller than original NUTS2 value (For example, when there are urban and rural areas only the % of both cases is smaller than 100%). Then, it is normalized to guarantee the average disposable income of NUTS2 is being used.

8.10 Yearly mileage distribution

8.10.1 Yearly mileage sources and characteristics by country

- **Germany**: The average yearly mileage in Germany has been found in [27]. It is data provided by Enerdata. Enerdata is co-funded by the European Union. It is not specified how yearly mileage is calculated. However, data of France is as well provided by Enerdata. Enerdata values and National Statistics of France values are the same. National Statistics of France uses odometer reading. Then, it is concluded Germany and other countries provided by Enerdata are provided using odometer reading.
- **France**: The most updated report of traffic presented by official authorities is *Annual Report on Transport in 2019* which was presented in December 2020 [10]. Data of average yearly mileage can be found. As it has been said before, it is measured using odometer reading data recorded by road authorities.
- **Italy**: The average yearly mileage is provided by Enerdata [27]. Average yearly mileage in Italy is very low compared to France and Germany showing clearly that country is a relevant source of differences in terms of yearly mileage.
- **Netherlands**: The average yearly mileage is provided by Enerdata [27]. Additionally, yearly mileage distribution is provided by CBS (Official National Statistics Office of Netherlands) [78]. CBS specifies that the vehicle kilometers are estimated based on odometer readings. The average yearly mileage obtained by CBS is very close to the one provided by Enerdata, less than 1% difference (See Table 4.27). It is the only country where yearly mileage distribution is provided for year 2019.
- **Norway**: The average yearly mileage of Norway is available by Statistics Norway in 2019 [119]. Values are measured using odometers reading.
- **Poland**: Values of average yearly mileage in Poland for year 2019 are obtained in Enerdata [27].
- **Greece**: Values of average yearly mileage in Greece for year 2019 are obtained in Enerdata [27].
- **Finland**: Average yearly mileage in Finland in 2018 is given by Traficom (Official Statistics Office of Finland). [126] It uses odometer reading data. There is no direct data of year 2019.

  Additionally, there is data of the vehicle kilometers in another source of Traficom [136]. Vehicle kilometers stayed constant from 2018 to 2019. The value was 40,718 million km. According to statistics Finland, there is an increase of 2,3% of the fleet. Then, it is considered that the yearly mileage will be reduced to a 97.75% of what it was in 2018.

8.10.2 Data of average yearly mileage distribution of countries selected

Data of average yearly mileage and its source is provided for all the countries. Data can be consulted in Table 8.24

Country	**average yearly mileage (source used)**	**Enerdata**	**OECD**	**Source Used**
Finland	13483,635	-	10329	National Statistics Office
France	12222,98789	12223	12741	National Statistics Office
Netherlands	12851	12849	12313	National Statistics Office
Norway	11883	-	11415	National Statistics Office
Italy	8464	8464	-	Enerdata
Germany	13602	13602	13175	Enerdata
Greece	10786	10786		Enerdata
Poland	8607	8607	7803	Enerdata

Table 8.24: Average yearly mileage in kms for different sources

8.10.3 Distribution fitting

8.10.3.1 Normality test

First, it is evaluated if data is following a normal distribution. Just watching Figure 4.10, it is quite clear it is not following a normal distribution. However, there are tools like QQ plot which allow to extract different conclusions of the distribution.

QQ plot

If the data follows a normal distribution, the quantiles of the data must be perfectly in line with the “theoretical” normal quantiles: a straight line on the QQ Plot defines a normal distribution. The QQ plot can be seen in Figure 8.28.

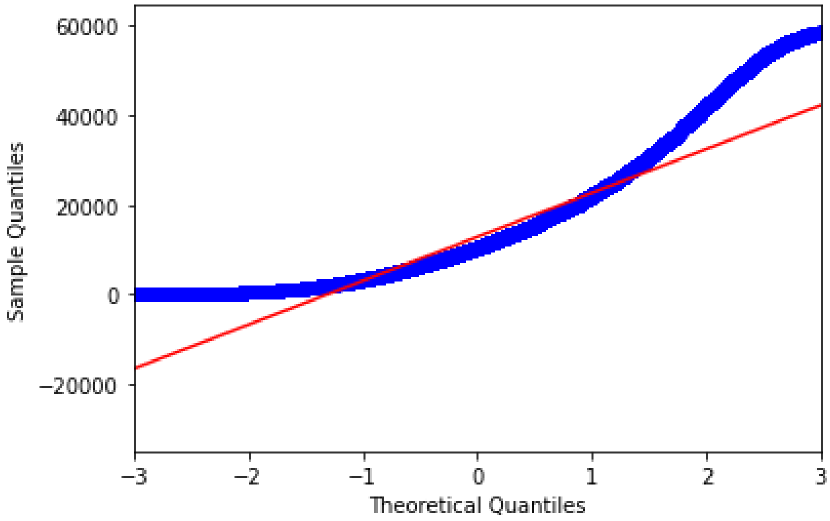


Figure 8.28: QQ plot of yearly mileage in 2019 in Netherlands

In QQ plot data does not look like a normal distribution. Normal distribution presents 99,9% of the data between -3 and 3.

Observing how tails separates in the extreme it can be concluded that it does not follow a normal distribution. A similar example can be seen in [111]. The shape indicates is right skewed and maybe an exponential function could fit this dataset.

Kolmogorov Smirnov test and Shapiro Wilk test

If the observed data perfectly follow a normal distribution, the value of the Kolmorogov Smirnov test statistic or as well the Shapiro Wilk test statistic will be 0. The P-Value is used to decide whether the difference is large enough to reject the null hypothesis. If the P-Value of the KS Test is larger than 0.05, we assume a normal distribution.

It is obtained statistic higher of 0.85 for both cases and a p-value of 0.0. It is confirmed data is not following a normal distribution.

8.10.3.2 Distribution fitting: beta distribution example

The Beta distribution has an extremely flexible shape, much more versatile than the normal distribution. Its default domain is the interval [0,1] for its random variates of x. The domain can be extended to much wider intervals or reallocated by adding location and scale parameters to the two share parameters a and b.

Beta distribution using a=1 and b=8 is plotted and the following distribution of data can be seen in Figure 8.29.

The parameters chosen create a positive-skewed distribution, with a drawn-out tail on the right. This is more similar to the yearly mileage distribution than a normal distribution. Here, the properties of the distribution are shown in Table 8.25.

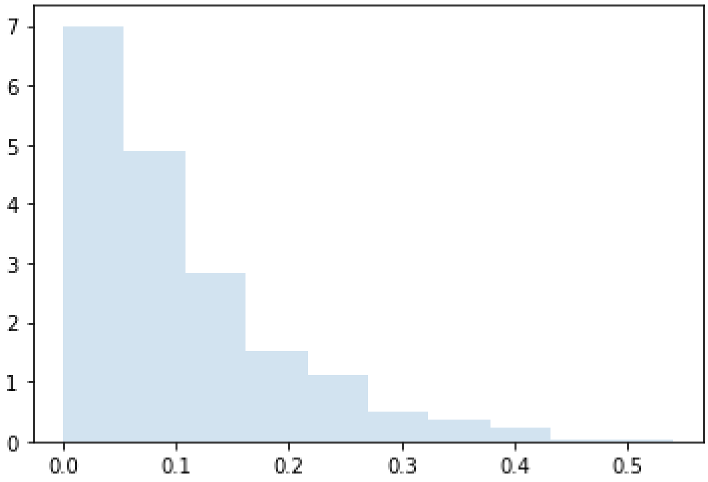


Figure 8.29: Histogram of beta distribution with parameters a=1 and b=8

.

skewness	1,42
kurtosis	2,28
mean	0,11
standard deviation	0.1

Table 8.25: Properties of beta distribution with parameters a=1 and b=8

The beta distribution has almost the same skewness and kurtosis than the distribution of yearly mileage data (See Table 4.30). However, the shape (See Figure 8.29) of beta distribution is still a bit different (See Figure 4.10).

Probability distribution function

In order to see better how beta distribution is in the domain, the probabilistic distribution function is plotted between the 1% and 99% quantiles.

It can be observed that this is not the shape that the yearly mileage distribution data of Netherlands is following. It is not able to capture the initial increase until reaches a maximum and then, a decrease.

Rescaling and shifting the distribution

Location and scale allow to reallocate the distribution and extend the domain. The location parameter loc moved the minimum of its support along the x-axis to the mean value; the sum of location and scale moved its maximum to the maximum value considering where a 99% would be reached.

In Figure 8.31, beta distribution has been shifted and rescaled. It has much more similar dimension to what would be expected for the yearly mileage distribution. Mean and variance have changed, but the kurtosis and skewness is the same meaning the shape of the distribution has not changed.

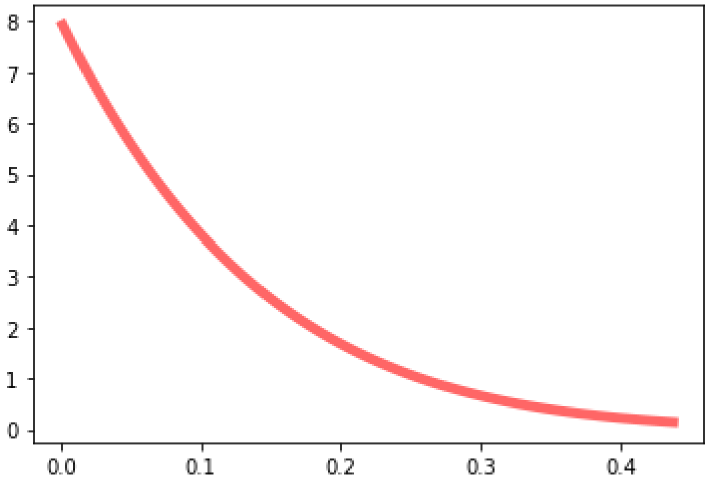


Figure 8.30: Probabilistic density function of beta distribution with parameters a=1 and b=8 between the 1% and 99% quantiles.

.

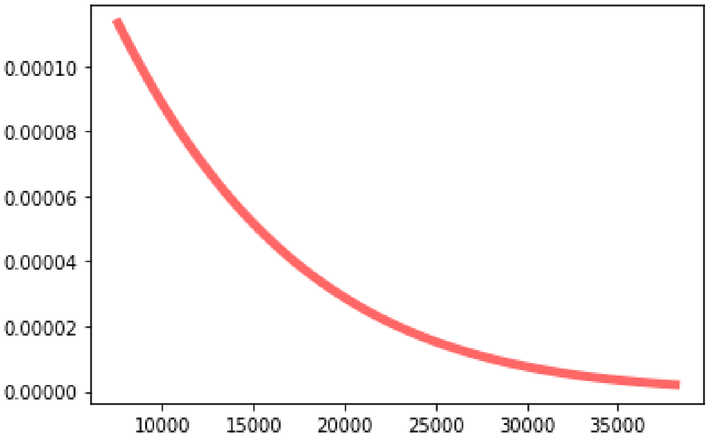


Figure 8.31: Probabilistic density function of beta distribution with parameters a=1 and b=8 between the 1% and 99% quantiles shifted and reallocated

.

Shifting and rescaling is used when fitting any distribution function to adapt the distribution to the specific dimensions of the data.

8.10.3.3 Criteria of selection of distribution functions

The candidate distributions to fit to the observational date should be chosen based on the following criteria:

- **The nature of the random process**: There are many distribution functions which are tailored to describe specific types of random processes. For example, to model the time till failure of equipment, a Weibull distribution should be among the candidates chosen for the fitting process. → **It is hard to define a specific distribution function which is specificaly designed for yearly mileage distribution.**

- **The domain or support of the distribution**: Most natural phenomena and most technical processes, for instance, cannot assume negative values. Some distributions have a non-negative domain inherent in their mathematical definition. Others can undergo transformations → **Yearly mileage has only positive domain.**
- **The shape of the observational data**: The shape of the histogram, skewness and kurtosis can help to select the distribution function. → **Yearly mileage data is right-skewed and has a positive kurtosis.**
- **Multi-model validations**: The investigation is not limited to a single model. Rather, several alternative distributions are used in the fitting process. → **80 distributions are used as candidates and passed to the fitter. This is done to ensure it is not obtained a suboptimal model.**
- **Goodness-of-fit tests**: It is used to compare the results among distributions and select the most optimal distribution model → **Sum squared error is used for evaluating distribution models.**

8.10.3.4 Selection of adjusted distribution function

It is studied if distribution functions selected kappa3 and burr, pass the Kolmogorov-Smirnov test and which is the value of the statistic. Genhalflogistic is used as well to compare results with the other two functions. The Kolmogorov–Smirnov test is used to serve as a goodness of fit test.

The Kolmogorov-Smirnov Test is affected by the number of points in the data. It is shown in Table 8.26 for a different number of points the p-value obtained.

number of points	**kappa3 distribution p-value**	**burr distribution p-value**	**gen halflogistic p-value**
10.000	0,095	0,089	0,0009
20.000	0,0021	0,0039	0,0000
50.000	0,0000	0,0000	0,0000
1.000.000	0,0000	0,0000	0,0000

Table 8.26: Kolmogorov Smirnov test for selected distribution functions and different number of points

For only 10,000 points burr and kappa3 pass the KS test because they have a p-value higher than 0,05. Genhalflogistic which it is not capable of having an inflection point or a maximum between 0 and 10.000 does not pass the KS test.

However, when increasing number of points, the p-value decreases. This is normal due to how the intervals of data are defined and supposing that they follow a uniform distribution. In Figure 8.32, the distribution for 1.000.000 points can be seen.

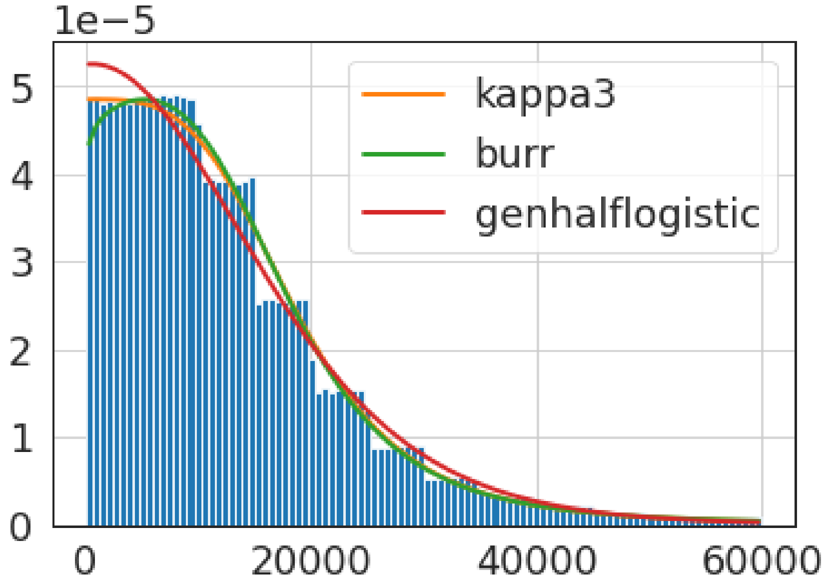


Figure 8.32: Histogram of data using 1 million points with adjusted distribution functions

The jumps between intervals are making that it is not possible to have a perfect fit and that KS test is failed. The solution for this would be having a higher number of intervals. For example, every 1,000 km or even every 500 km. Not passing KS test does not mean this are not following correctly the shape of the data.

Given the intervals that can be obtained from CBS (National Statistics of Netherlands) [78], this is the best that can be done.

Regarding the selected distribution function, kappa3 gives slightly better results than burr when checking statistics obtained (p-value, minimum squared error). However, the shape does not present a defined maximum after 0. Considering other study of yearly mileage distribution of Netherlands in 2017 [83], it can be observed that there is a maximum between 5,000 and 10,000 km. Then, burr function is describing better yearly mileage distribution.

Burr function is selected as the function to best fit yearly mileage distribution.

8.10.3.5 Error of selected distribution function

In Table 8.27, it can be observed that the error of the cumulative distribution function is maximum at genhalflogistic. kappa3 and burr present relatively low error presenting the highest error at the beginning where is most difficult to the function to adjust.

If the average error at all the points selected (kms done) is calculated, burr presents an error of 0.65%, kappa3, 0.58% and genhalflogistic, 1.18%. A slightly higher error in burr can be expected because of the shape it has. It is not following 100% the intervals of data, but it is following a typical yearly mileage distribution.

kms done	**cdf data (%)**	**cdf burr (%)**	**cdf kappa3 (%)**	**cdf genhalflogistic (%)**	**error burr (%)**	**error kappa3 (%)**	**error genhalflogistic (%)**
5000	24,14	23,80	24,20	25,85	1,43	0,25	7,08
10000	48,55	47,48	47,46	48,79	2,19	2,25	0,49
15000	68,18	67,41	67,09	66,76	1,14	1,60	2,08
20000	80,96	81,01	80,78	79,47	0,07	0,23	1,83
25000	88,61	88,99	88,92	87,79	0,43	0,35	0,92
30000	93,06	93,41	93,43	92,94	0,37	0,40	0,13
35000	95,72	95,87	95,93	96,00	0,16	0,22	0,29
40000	97,33	97,29	97,36	97,78	0,04	0,03	0,46
45000	98,40	98,15	98,22	98,78	0,25	0,18	0,39
50000	99,11	98,69	98,75	99,34	0,42	0,36	0,23
55000	99,58	99,05	99,10	99,65	0,54	0,48	0,07
60000	100,00	99,29	99,33	99,81	0,71	0,67	0,19

Table 8.27: Cumulative distribution function values at different points and relative error respect to original data.

9 Appendix of Chapter 5

9.1 Purchase price after taxes and incentives

Purchase price including taxes and incentives for large segment is shown in Figure 9.1.

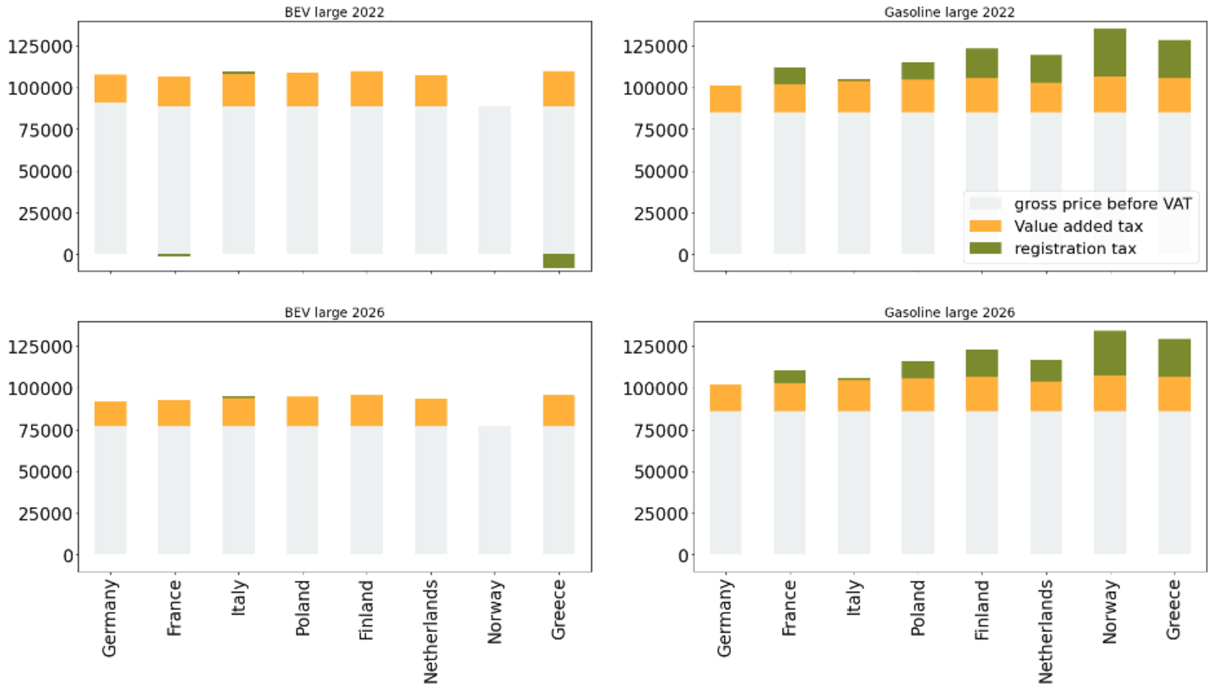


Figure 9.1: Purchase price including taxes and incentives for gasoline vehicles and battery electric vehicles of large segment in 2022 and 2026

Purchase price including taxes and incentives for small segment is shown in Figure 9.2.

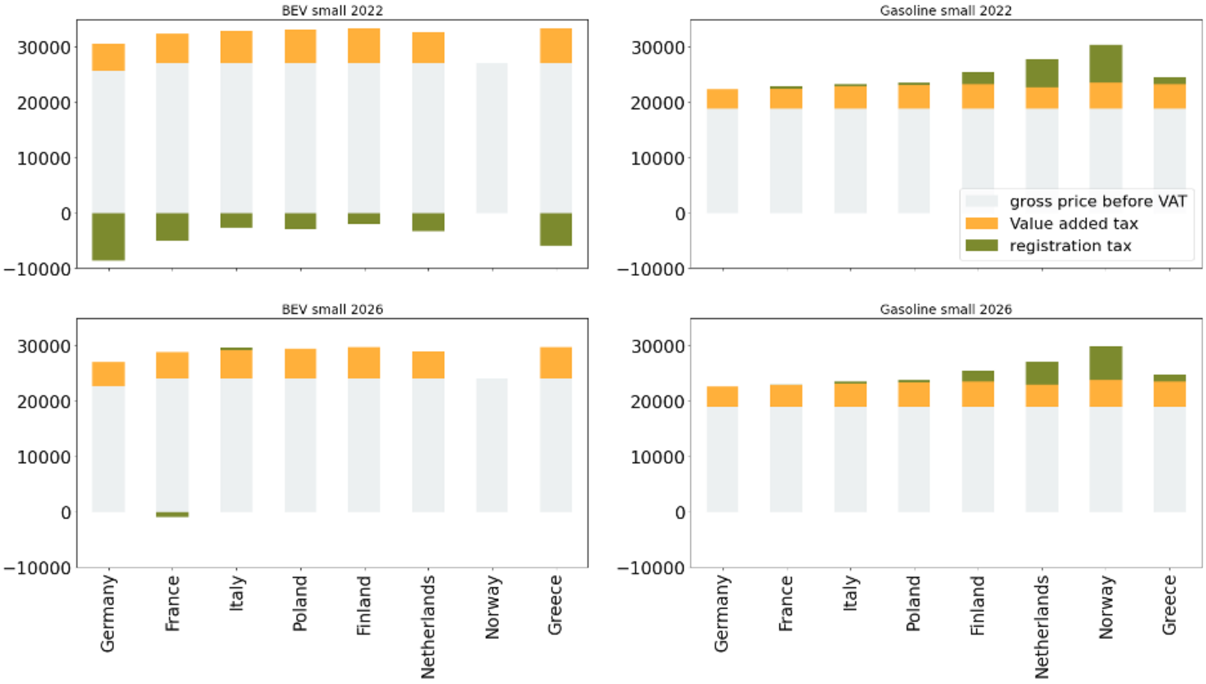


Figure 9.2: Purchase price including taxes and incentives for gasoline vehicles and battery electric vehicles of small segment in 2022 and 2026