HAL
open science

Worker mobility and the purchase of low CO2 emission vehicles in France: a datamining approach

Raphaël Homayoun Boroumand, Stéphane Goutte, Thomas Péran, Thomas Porcher

To cite this version:

Raphaël Homayoun Boroumand, Stéphane Goutte, Thomas Péran, Thomas Porcher. Worker mobility and the purchase of low CO2 emission vehicles in France: a datamining approach. The European Journal of Comparative Economics, 2019. halshs-01968001

HAL Id: halshs-01968001

https://shs.hal.science/halshs-01968001v1

Submitted on 1 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Worker mobility and the purchase of low CO2 emission vehicles in France: a datamining approach

Raphaël Homayoun BOROUMAND *, Stéphane GOUTTE †,
Thomas PERAN * and Thomas PORCHER *.

April 24, 2017

Abstract

This paper seeks to pattern a non-driven geographical classification of French departmental territorial units based on both mobility behavior and passenger car fleet composition. With no mathematical regression analysis but applying datamining methodology to behavior, consumption and geography variables, we have grouped French territorial units into 8 clusters with similar characteristics. The main results reveal that commuters' behavior with respect to the choice of transport mode varies substantially across clusters (rural and highly rural, urban and highly urban clusters, ...). Conversely, the structure of the French vehicle fleet and French car purchases in terms of engines, tax horsepower and CO2 emissions are similar. this finding should enable state organizations to implement differentiated public policies for environmental and industrial sectors. Alternatively, our paper should help industrial groups to better deploy their economic strategies in line with environmental concerns.

Keywords: CO2 emissions; Datamining; Cluster; Worker mobility; Passenger cars; France.

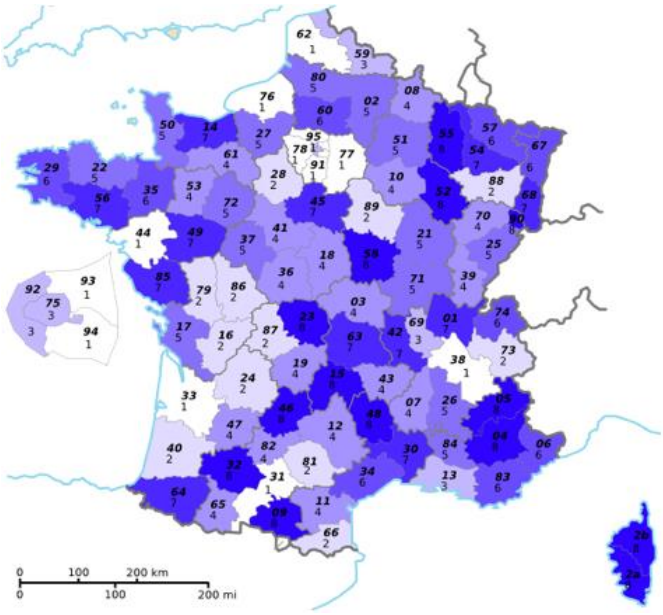
JEL classification: O18; P18; R11; R40.

*Associate Professor, Department of Applied Economics, PSB Paris School of Business, 59 rue Nationale 75013 Paris France.

†Université Paris 8 (LED), 2 rue de la Liberté, 93526 Saint-Denis Cedex, France.

Graphical Abstract

Heterogeneity
of mobility in France*



*Bolded figure indicates the administrative number of the territorial unit
Second figure indicates the cluster to which the territorial unit belong

Homogeneity
of car purchases† in France

Low CO2 emissions

Clusters	Car purchases [101;120] g
1	35,92%
2	36,26%
3	35,73%
4	35,98%
5	36,16%
6	36,33%
7	36,38%
8	36,43%

High CO2 emissions

Clusters	Car purchases [166;200] g
1	7,49%
2	7,35%
3	7,55%
4	7,44%
5	7,33%
6	7,26%
7	7,22%
8	7,18%

†Cluster average over 2006-2014 period

1 Introduction

The emissions of greenhouse gases (GHG) from transport have grown substantially in recent years, and projections for 2020-2030 are pessimistic (Pérez-Suárez and López-Menéndez, 2015). In many countries of the European Union (EU) such as France, this growth may even thwart the aim of complying with the 3x20 environmental rule[1]; it is a central component of the EU's growth strategy for 2020 (EU, 2010). Moreover, it is a long-established finding that CO2 emissions from the most polluting vehicles are directly responsible for a growing number of cancers and respiratory diseases (Dockery et al., 1993; Fullerton et al., 2008). During the year 2012 alone, 7 million premature deaths were attributed to air pollution exposure (WHO, 2014). Particular attention must also be devoted to the emission of fine particles from diesel vehicles (Longhin et al., 2016). Studies linking health and air pollution report that particulate matter ($PM_{2.5}$) and black carbon emissions from incomplete fuel combustion are responsible for long-term mortality (Bond et al., 2013), and cindynics[2] has provided insights on this matter that can be incorporated into risk analysis models (Fann et al, 2012; Smith et al., 2013).

Designed as a component of the new environmental science, the study of territories and, specifically, the negative externalities related to car use in urban areas have already generated good results and helped to provide strong conclusions (Parry et al., 2007). Since the 1990s[3], the environmental protection of territories has fully included new binding practices for reducing waste, protecting the environment and maintaining an optimal quality of life, which includes the preservation of air quality (IPCC, 2014).

The dominant stream of literature focuses on describing the positive externalities associated with control over transport. Indeed, while industry, agriculture and services are all responsible for the growth of CO2 emissions, the most important component of air quality degradation comes from transport (Curiel-Esparza et al., 2016; Rustemoglu and AndrŐs, 2016; Davydova-Belitskaya and Skiba, 2003).

Studies that go deeper into identifying the causes of carbon emissions stress the key role played by factors such as population change (Schellnhuber and Svirejeva-Hopkins, 2008), economic growth per capita, regional energy intensity, the contribution of regional fuel mix and energy and carbon intensity (GonzǦlez et al., 2014; Remuzgo and Sarabia, 2015). Other specific papers over the 1995-2015

[1]Based on the EU's 1990 pollution level, this rule calls on Member States to reduce GHG emissions by 20%, have renewable energy represent 20% of the energy mix and achieve a 20% energy savings.

[2]The science of risk analysis.

[3]From the adoption of the United Nations Framework Convention on Climate Change at the Earth Summit in Rio de Janeiro (June 1992) to the inclusion, for example, of the Environmental Charter (2004) into the French Constitutionality Corpus, a strong chain of hierarchically organized national and international legal standards devoted to the preservation of the environment has been established. In France, the French constitutional judge, since Constitutional Act No. 2005-205 of March 1, 2005 (Article 2), can invalidate national law that contravenes an environmental obligation and fully apply the precautionary principle.

period strictly demonstrate the strong relationship between the per capita income of urban territorial units and the pollution burden they generate4: Boyce, 1994; Torras and Boyce, 1998; Scruggs, 1998; Boyce et al., 1999; McGranahan and Satterthwaite, 2002; Hedenus and Azar, 2005; White, 2007; Kovacs et al., 2013; Zwickl et al., 2014; Holian and Kahn, 2015.

If we restrict our attention to the analysis of air pollution factors in territorial units (regions or sub-regional units), in addition to socioeconomic data collection to establish a typography of transport users' polluting practices in urban areas (Buchs and Schnepf, 2013; Chancel, 2014), it is crucial to have data on the composition of the private car fleet (Agyemang-Bonsu et al., 2010; Kholod et al, 2016). This difficulty is not always easy to overcome because state censuses of private parking facilities, such as the COPERT 4[4] databases for Europe and parts of America and Asia, are heterogeneous and incomplete due to their practice of compiling public data that are not published on a regular basis. Alternative approaches proposed in various studies such as photographs and videos of roads and parking lots (Kholod et al., 2016) provide valuable additional information, but their accuracy is insufficient if the aim is to evaluate the level of pollution caused by a country's passenger car fleet. For example, data from the COPERT 4 file distinguish among passenger cars, light commercial vehicles, trucks, buses, motorcycles and mopeds. However, the file contains no data on the tax horsepower or engine types of passenger cars. However, several studies claim that among the factors affecting pollution from transport emissions in urban areas, it is most important to consider the composition of the fleet and parking restrictions (Kholod et al., 2016), if possible through composite indicators (Kilkis, 2016) or multidisciplinary studies (Venkatesh et al., 2014).

Only a public census over a long period of time enables the effective application of data relating to fleets and behavior to measure progress in addressing air pollution from transport. Recall that a 2010 estimate indicated that the share of transport in total global greenhouse gas emissions was 23%, with the corresponding figure for the US being 25% (Sims et al., 2014). Automobile manufacturers in the EU have already implemented innovations to significantly reduce the fuel consumption of passenger cars (Fontaras and Dilara, 2012). However, the overall volume of fuel combustion and corresponding pollution from transport may continue to grow if forecasts indicating a substantial increase in the number of vehicles per resident of a territorial unit, city, region or state continue to hold (IEA/OECD, 2010). Indeed, forecasts from the IPCC indicate that the number of vehicles will double by 2030 and triple by 2050, bringing the total number of light-duty vehicles worldwide to 2 billion (Kahn Ribeiro et al., 2007). Given these uncertainties, two political interventions can be recommended: first, to encourage lower-emissions transport choices related to individual mobility as carpooling, widespread public transportation use, rail system use, and alternative scenarios (Vermote et al., 2013); and, second, promote technological innovation to optimize vehicles and make them less polluting (Palencia

[4]The COPERT 4 program (Computer Program to Calculate Emissions from Road Transport) is a global model for estimating air pollution from road transport. Developed in coordination with the European Environment Agency, it calculates gas emissions estimates.

et al., 2012). However, in the 1990s, the EU initiated a new policy: moving from a petrol-driven[5] car market to a diesel-driven one. Studies have demonstrated that the potential gains from this industrial policy choice are small and overcompensated by, first, the massive CO2 emissions attributed to the associated supply chain and, second, the resulting increase in black carbon particles because the first generation of diesel vehicles were not equipped with particulate filters (Cams and Helmers, 2013). Worse, the dieselization of the European fleet allowed buyers of passenger cars with diesel engines to affect future savings on their vehicle purchases. With an average of -35% in fuel combustion efficiency for a type-A diesel vehicle compared to gasoline vehicles of the same type, buyers can afford to buy more powerful and therefore more polluting cars (Schipper and Fulton, 2009).

On a purely behavioral level, studies seeking to better understand the determinants of the choice of commuter transport modes are numerous. Some incorporate into model inputs regarding the possible degree of leisure during the trip, comfort, ease of postponing travel or the cost related to the daily commute to work (Wang et al. 2013; Chowdhury et al., 2015; Peer et al., 2016). Others identify the quality and the security of route changes as levers that could encourage the use of public transport instead of private cars (Kingham et al., 2001). It appears that commuters' age and income influence passenger car purchasing behavior and, therefore, their choice of transport mode (Liu et al., 2016). However, if the ease of access to transport infrastructure and the route practicability are clearly identified as factors affecting the satisfaction of workers with mobility requirements (Wang and Wang, 2016), the literature rarely includes the exact distribution of daily commuting methods used by a population based on geographic data at the level of the territorial unit. There are studies demonstrating the difficulties in commuting to work related to geographical factors (Mackett and Thoreau, 2015) and others that offer a high degree of detail on the modes of transport used daily by commuters (St-Louis et al., 2014). However, few studies are able to combine behavioral data, national vehicle fleet data and geographical data.

In an original approach using datamining and clustering, justified in section 2.2, our paper fills a major gap in the literature by providing a highly detailed study and results regarding the level of pro-environmental behavior by French commuters when making a car purchase. Global and national studies highlighting the cause and effect relationship between fuel combustion and the environmental and health problems are now sufficiently numerous and precise to promote environmental awareness among consumers. However, few of these studies are directly applicable to France, where the administrative division of the territory is an intricate and historical legacy with high impact in the implementation of energy policies. Our working hypothesis is that in spite of the great geographical and behavioral heterogeneity of the French territorial units, it is possible to establish a new basic system to implement environmental and industrial policies respectful of the diversity of the population (i.e., departmental units in our paper). The first objective of this research is to document and highlight the specific

[5] In 1990, the proportion of diesel vehicles in the total European fleet did not exceed 10%.

behaviors and trends among the French in terms of labor mobility, that is to say, mobility linked to job requirements (distance, frequency, infrastructure, etc.). The second objective of this paper is to shape a non-driven geographical classification of French departmental territorial units, based on both mobility behavior and passenger car fleet composition.

The paper is structured as follows. Section 1 discusses the level of detail available in extant studies on worker mobility. Section 2 describes the conceptual framework and specifies the data and methodology. Section 3 develops the empirical results and economic discussions. Section 4 concludes and provides policy recommendations.

2 Data and Methods

2.1 Data sources

Studies on workers' mobility have increased in France since the 1970s, when the government decided to create a Ministry of Environment[6]. Therefore, the government and the National Statistics and Economic Studies Service[7] (INSEE) have conducted relevant investigations and qualitative analyses. The production of behavioral data on commuters was institutionalized in the late twentieth century, and thus general behavioral facts have been highlighted and widely reported in the academic literature. This research has established that the most disadvantaged workers are residents of medium-sized cities (Villeneuve, 1970). These cities are large enough to ensure a significant number of jobs and small enough to ensure proximity to employment. However, estimates indicate that in the late 1980s, one out of two commuters leaves his town of residence to go to work, meaning that commuting is longer and entails higher speeds (Terrier, 1986).

Since that time, given the understanding of what the application of statistical approaches to workers' mobility can provide in terms of defining and strengthening public policies, institutional data and French government data on "mobility and work" have become abundant. Such data can now be used to highlight new commuter behaviors or reveal the acceleration of existing trends. However, one difficulty is that many of these data are produced and published by public departments that do not communicate in a coordinated manner. The aim of the present study is to combine data from various public institutional sources into a single dataset to restate and study them more consistently. In total, the database that we construct and use in this article contains more than 3.1 million observations distributed across 354 variables to study their evolution over time. Specifically, we present three

[6] The Ministry of Environment (otherwise known as the Ministry of Ecology or the Sustainable Development Ministry) and the Ministry of Equipment coexisted in France until 2007, when the two were merged.

[7] INSEE is the department responsible for the production and publication of official statistics under the Ministry of Economy.

categories of data: behavioral data, consumption data and geographical data[8].

2.1.1 Commuter behavioral data (2006-2012)

Since 2004[9], INSEE has performed an annual national population census. This regular census should enable us to accurately count the French population to determine its demographic and social characteristics and, ultimately, develop a profile of French housing and work. This census uses municipalities as the territorial units. Under the terms of the 2002 law on local democracy, for municipalities with a population under 10,000 residents, censuses are comprehensive and are conducted annually over a five-year rolling window. For the other municipalities, a sample survey is conducted annually and the entire territory of these municipalities is surveyed at the end of the same five-year period. The institutional behavioral data we include in our database are exclusively mobility flow data for the active workforce. According to the definition employed by the censuses conducted by the State authorities since 2004, the active workforce is defined as any person 15 years of age or older reporting a profession (salaried or not) working full or part time, helping a person in his work as an apprentice or a paid trainee, being unemployed or engaged in reduced working time, or finally, being student or retired but still employed.

Mobility flows for year n are available in year $n+3$ and based on the population census from year n and the geographical division of year $n+2$. Furthermore, the average age of private cars included in the French fleet as of January 1, 2012, is 8.2 years[10]. For some datasets, including those related to commuter consumption and technological advances in vehicle engines, we considered the years 2013 and 2014. We therefore selected 2006-2012 as our period of study. The data studied are generally available from the institutional websites of the French government[11] or available on request from the relevant government services and departmental directorates. The first step in constructing the database was to establish, for each of the 36,688[12] French municipalities, the working population (consisting of individuals over 15 years of age). For the latter number, we report the numbers of individuals using - in their daily commute to work with a starting point of the town of residence considered - a private car, public transportation, two-wheeled vehicles and walking. We then made our observations and drew conclusions based on our study of the variation in the municipal-level

[8] See the Appendix for the main municipal and departmental variables under consideration.

[9] In accordance with French law no. 2002-276 of February 27, 2002, on local democracy.

[10] http://www.ccfa.fr. The French automobile manufacturers committee (CCFA) is the leading French trade association for car manufacturers.

[11] http://www.insee.fr; http://www.statistiques.developpement-durable.gouv.fr; http://www.observatoire-des-territoires.gouv.fr

[12] This number represents the maximum number of territorial units over the period 2006-2012. In the initial construction of this database, we include the 112 municipalities in the French overseas communities, namely, Martinique, Guadeloupe, French Guyana and Reunion.

utilization rates of each means of transportation, including the numbers and rates of individuals working outside their municipality of residence. Of course, it will be necessary, first, to explain the variation in all of these rates over the period considered for a given territorial unit and second, to compare the differences across territorial units during the period of reference or within a single year.

2.1.2 Consumption and mobility (2006-2014)

To avoid aggregation bias[13], we applied our datamining and clustering methodology at the departmental level. Consequently, another important step in the construction of the final database was to reassemble the 100 French departments[14] by aggregating municipal data. The French departments, administered by General Councils, are the territorial reference units, both politically and geographically. This provides us with the departmental utilization rates of different means of transportation (private cars, public transport, two-wheeled vehicles and walking) and the per department rates of workers working outside their municipality of residence. We then inserted our "consumption" data. First, we include the department-level composition of private cars under 15 years old. This composite comprises variables on motor vehicles. We employed two databases in this regard: one detailing the engine power source and tax horsepower [15], with the other only providing (highly detailed) information on the tax horsepower of vehicles. Second, we also included the registration volumes of new private cars. Overall department-level volumes are supplemented by department-level volumes based on CO2 emissions of vehicles registered[16]. Note again that the public institutional data are published at an inconsistent rate and via heterogeneous channels. For example, the entries in the departmental fleet of gasoline + super ethanol powered vehicles and gasoline + electricity powered vehicles only begin in 2001; diesel + electricity powered vehicles have entries beginning in 2012. These categories are certainly marginal in terms of vehicle numbers, but their publishing schedule and difficulties, logically related to technological advances, clearly demonstrate the difficulty for French ministerial and departmental statistical services[17] in compiling new series.

[13] As the basic administrative level of the database is the municipality, we chose to aggregate at the departmental territorial unit. We assume that geographical heterogeneity is minor but meaningful at this level. In France, administrative levels in order of importance (ascending order) are municipality, department and region.

[14] We exclude Mayotte because of the unreliability of the associated data due to the special status of the community. Mayotte has been a French department and an overseas region since 2011, following the referendum of March 29, 2009.

[15] Otherwise called "tax rating".

[16] Refer here to the Statistical Directory of Road Vehicles (RSVERO) published by the Ministry of Environment.

[17] The departmental directorates of equipment, for example.

2.1.3 Geographical data

To better reflect the geographical and topographical (or geological) realities in interpreting our results, we also reconstructed areas by initially aggregating data at the municipality level. In this regional sub-database, we excluded data from communities overseas because their geographical positions, as islands or on other continents, make them outliers. Furthermore, we established two sub-databases, a departmental one and a regional one, in which variables are periodic. The reference period 2006-2012 is maintained. When data are accessible and comprehensible, the 2006-2014 period is considered. The figures for gasoline and diesel consumption in thousands of cubic meters are available for the period from 1980 to 2013, but only for the regional territorial units. To prepare the results and facilitate their interpretation, we expanded our database to include other "geographical" indicators. To that end, we specified when a city, state or region borders another country. We also constructed territorial meta-units based on telephone area codes defined by the French regulatory authority for electronic communications and post (ARCEP). This independent administrative authority, created in 1997, identifies five key areas associated with local telephone codes: the Paris region, northwest region, northeast region, southeast region and southwest region. The aggregation of larger-scale data allows us to employ the previously described variables for each of these areas. Finally, we created four categories of city and indicated, for each department and each region, the number of cities that belong to each category. Category 1 consists of Paris (>2 million inhabitants); category 2 includes cities with populations between 200,000 and 1,999,999; category 3 includes cities with populations between 100,000 and 199,999; and category 4 includes cities with fewer than 100,000 residents.

2.2 Datamining and Clustering

2.2.1 Principal Component Analysis (PCA)

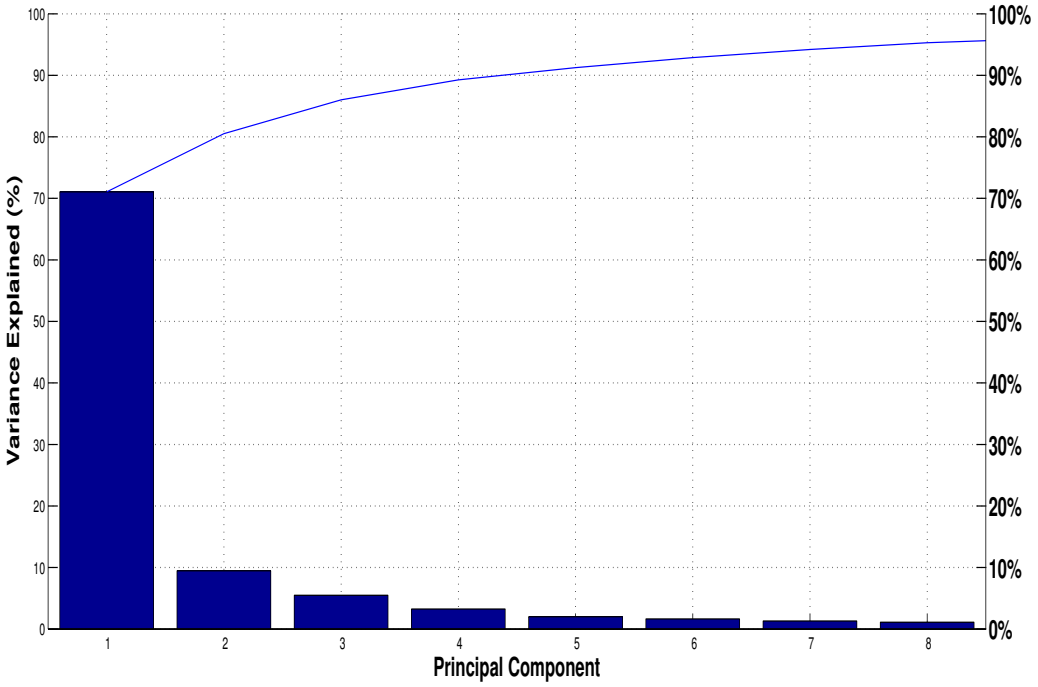
We begin with a datamining methodology, Principal Component Analysis (PCA), to highlight the clustering structures in our variables. This methodology allows us to identify the so-called principal components. They represent the underlying structure of the data and reflect the dimensions along which there is the highest variance, or where the data are most diffuse. We do this because in the presence of a large number of variables (as in our case), the dispersion matrix may be too large to study and interpret properly. There would be too many pairwise correlations between the variables to consider. A graphical depiction of the data may also not be of particular use if the dataset is very large.

To interpret the data in a more meaningful way, it is necessary to reduce the number of variables to a few, interpretable linear combinations of data. Each linear combination will correspond to a

principal component.

To measure the effectiveness of reducing the number of variables, we plot in Figure 1 the explained variance for the components obtained via PCA. We observe that the first component (first axis) explains 71% of the total variance, and with a 2D plot, we conserve and explain 81% of the total variance.

Figure 1: Explained variance for the components obtained via Principal Component Analysis



2.2.2 Hierarchical Clustering Analysis (HCA)

Cluster analysis is a data exploration (mining) tool for dividing a multivariate dataset into "natural" clusters (groups). We use this method to explore whether previously undefined clusters (groups) may exist in the dataset.

Cluster analysis is used when there is reason to believe that the sample units come from an unknown number of distinct populations or sub-populations. We also assume that the sample units come from a number of distinct populations, but there is no a priori definition of those populations. Our objective is to describe those populations (i.e., our departments) using the observed data.

The use of datamining offers major leverage in many disciplines. The aim is to take advantage of a "non-driven" analytical approach exempt from any reading bias. The enrichment of science by datamining appears mainly in biological (Nguyen et al., 2016), environmental (Su et al., 2004) or information systems studies (Altuntas et al., 2016). Representing knowledge in an explicit form with datamining is highly recommended in geographical or spatial papers in order to avoid the shortcomings

of other existing methods, such as a "lack of interactiveness in the objective space, inability to handle discrete variables and inability to generate explicit knowledge" (Bandaru et al., 2015). In particular, mathematical regression analyses are numerous and useful in the economic fields to demonstrate correlations or to predict asset responses to condition variations (Sharif B et al., 2017). However, taking a completely different scientific approach, our paper aims to pattern a classification of similar geographical units based on a quite prominent database without any supervised information or criteria. Our aim is to identify a statistical structure that is based on the French department territorial units for our data. The use of regression or correlation will give us only a quantification of corresponding results and the parameter impacts on a dependent variable, but they will not provide us with a specific structure for our data set.

In the agglomerative hierarchical approach, we begin by defining each data point as a cluster and combine existing clusters in each step. We employ Ward's method. This method does not directly define a measure of distance between two points or clusters but is an ANOVA-based approach. Each stage merges the two clusters that generate the smallest increase in the combined error sum of squares from a one-way univariate ANOVA for each variable, with groups defined by the clusters at that stage of the process.

This cluster analysis can be regarded as an analysis of variance problem, instead of using distance metrics or measures of aggregation.

This method involves an agglomerative clustering algorithm. Ward's method begins with n clusters of size 1 and continues until all observations are included in a single cluster. This method is most appropriate for quantitative variables, as in our case. Let X_{ijk} denote the value for variable k in observation j belonging to cluster i. Furthermore, we define

Error Sum of Squares:

$$ESS = \sum_i \sum_j \sum_k |X_{ijk} - \bar{x}_{i.k}|^2$$

We sum over all variables and all units within each cluster. We compare individual observations for each variable against the cluster means for that variable. Note that when the error sum of squares is small, this suggests that our data are close to their cluster means, implying that we have a cluster of like units.

Total Sum of Squares:

$$TSS = \sum_i \sum_j \sum_k |X_{ijk} - \bar{x}_{..k}|^2$$

The total sum of squares is defined as usual. Here, we compare the individual observations for each variable against the mean for that variable.

R-Squared:

$$R^2 = \frac{TSS - ESS}{TSS}$$

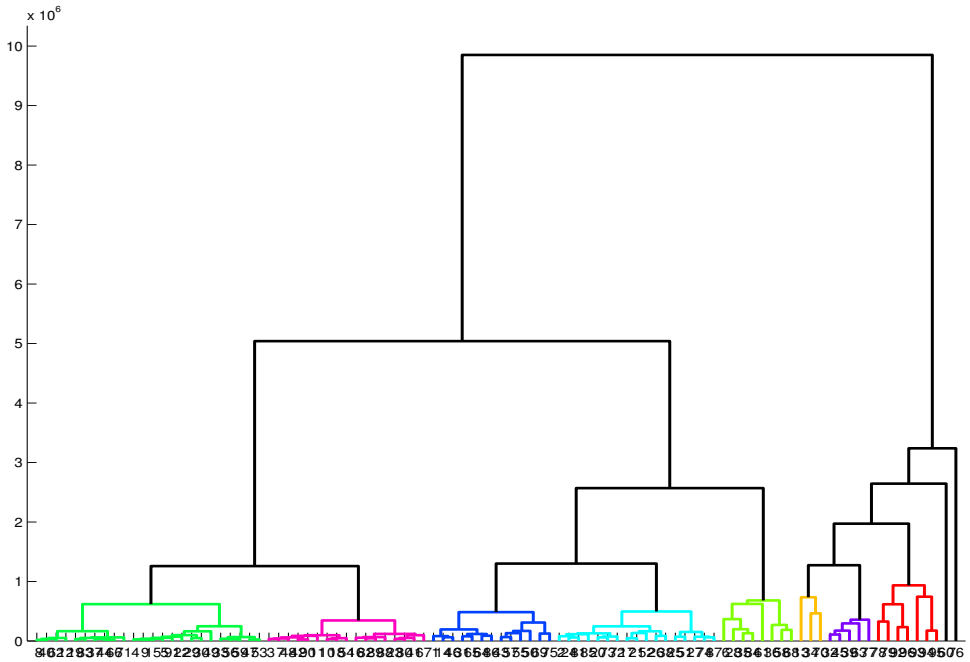
This R^2 value is interpreted as the proportion of variation explained by a particular clustering of the observations.

Using Ward's method, we begin with all sample units in n clusters of size 1 each. In the first step of the algorithm, $n-1$ clusters are formed, one of size two and the remainder of size 1. The error sum of squares and R^2 values are then computed. The pair of sample units that yield the smallest error sum of squares or, equivalently the largest R^2 value, will form the first cluster. Then, in the second step of the algorithm, $n-2$ clusters are formed from the $n-1$ clusters defined in step 2. These may include two clusters of size 2 or a single cluster of size 3 that includes the two items clustered in step 1. Again, the value of R^2 is maximized. Thus, at each step of the algorithm, clusters or observations are combined to minimize the error sum of squares or, alternatively, to maximize the R^2 value. The algorithm stops when all sample units are combined in a single large cluster of size n.

2.2.3 Dendrogram

The results of a cluster analysis are best summarized using a dendrogram. In a dendrogram, distance is plotted on one axis, while the sample units are plotted on the other axis. The tree shows how sample units are combined into clusters, the height of each branching point corresponding to the distance at which two clusters are joined. Our dendrogram is plotted in Figure 2.

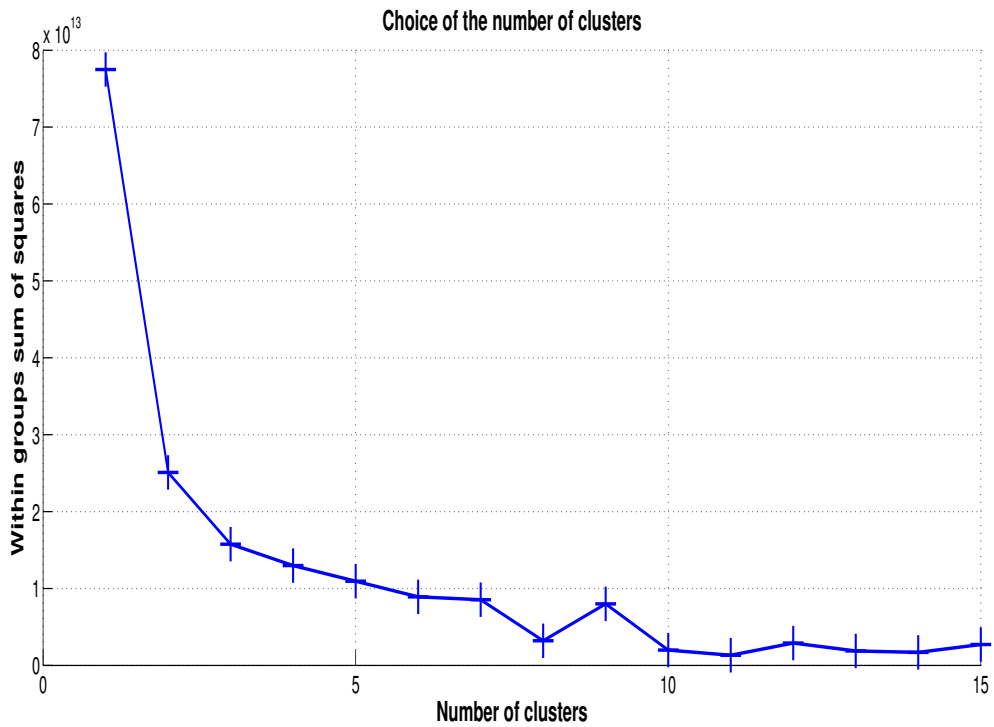
Figure 2: Dendrogram



2.2.4 Optimal number of clusters

It is necessary to determine the optimal number of clusters. To do so, we plot in Figure 3 the within-group sum of squares (i.e., error sum of squares (EES)), which we would like to be close to zero.

Figure 3: Within groups sum of squares



The result indicates that the choice of $K = 8$ clusters (or more than 10) seems to minimize the ESS and thus to be a good number of clusters. We will choose $K = 8$ instead of greater than 10 because it will be too difficult to capture, identify and interpret the structure of the departments in each cluster if we take too many groups. Of course, when $K \to N$, where N is the total number of departments, the ESS converge to zero.

Another statistical test to choose the optimal number of clusters based on the dendrogram tree is the Duda-Hart (2001, sec. 10.10) Je(2)/Je(1) index. This rule is as follows: larger values indicate more distinct clustering. Presented with the Duda-Hart, Je(2)/Je(1) values are pseudo-T-squared values. Smaller pseudo-T-squared values indicate more distinct clustering.

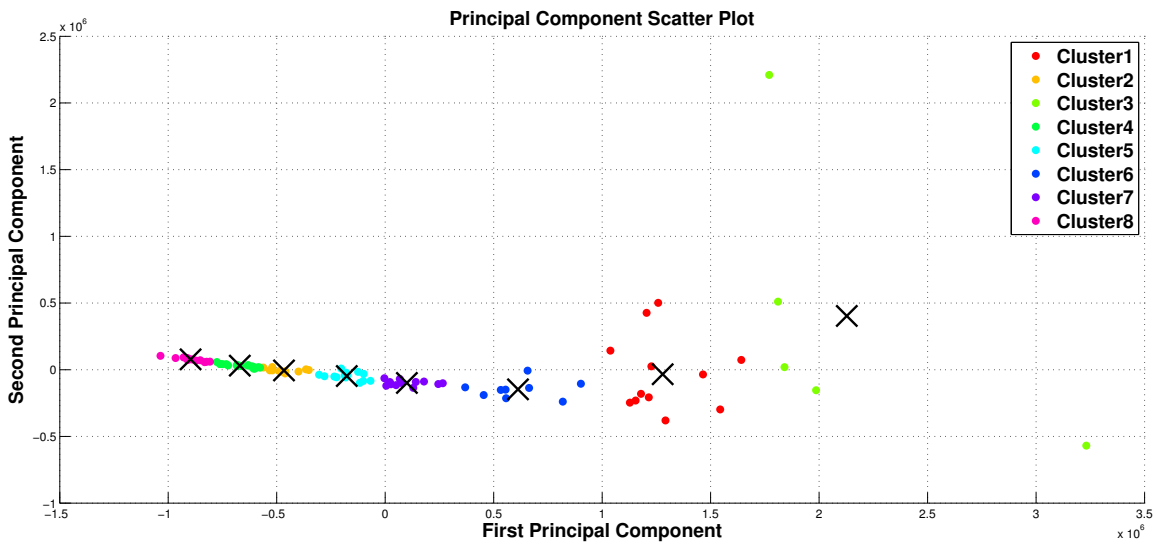
	Duda/Hart	
Number of clusters	Je(2)/Je(1)	pseudo T-squared
1	0.3494	175.04
2	0.2670	189.38
3	0.7409	8.04
4	0.3936	67.79
5	0.5967	10.14
6	0.5490	11.50
7	0.5433	10.93
8	0.7313	9.54
9	0.3840	36.89
10	0.3005	13.97
11	0.5624	3.89
12	0.5959	4.07
13	0.0517	18.33
14	0.3190	36.29
15	0.3301	6.09

We clearly see that this test gives $K = 3$ or $K = 8$ as the optimal number of clusters. We choose $K = 8$ instead of $K = 3$ in the sequel because with only 3 clusters, the segmentation is not accurate enough to allow good economic and policy recommendations.

2.2.5 Clustering projection on principal component

We plot in Figure 4 the departmental cluster projection with $K = 8$ in the principal components obtained via PCA. We observe that the non-supervised segmentation divides the projected departments into well separated clusters. This means that our clustering is efficient in separating the data in our dataset.

Figure 4: Cluster segmentation for $K = 8$



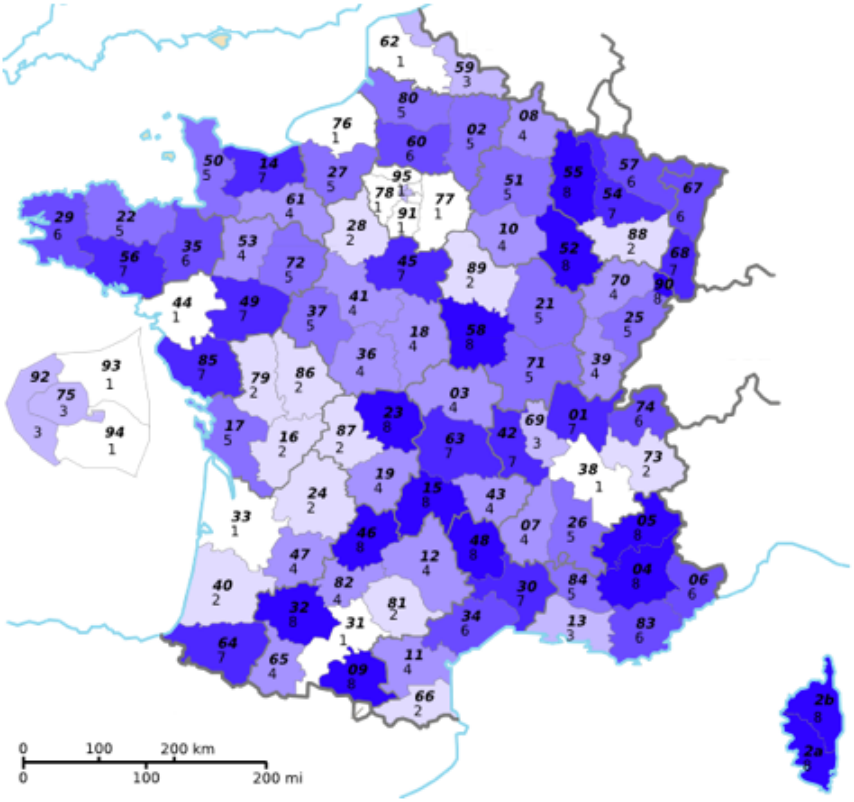
We put the clustering results obtained with $K = 3$ and $K = 5$ in the Appendix.

2.2.6 Clustering projection on French map

We plot in Figure 5 the departmental cluster projection with $K = 8$ on the French geographical map.

We put the clustering map results for each cluster separately in the Appendix.

Figure 5: Cluster segmentation for $K = 8$



3 Results and Discussion

Before evaluating the determinants of and the links between French mobility and consumer behaviors, a geographical map of possible territorial unit combinations should be made. Table 1 reports the composition of the 8 clusters and the departments associated with them. We present a description of geographical realities highlighted by clustering and the behavioral tendencies of the active workforce over 15 years old in daily commuting to work. Furthermore, we depict a typical private car purchasing structure in each cluster. The discussion section aims to connect these figures and offers an explanation of the environmental awareness of French workers.

Table 1: Composition of the 8 clusters

Geographical clustering content (departmental based)

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Essonne	Charente	Bouches-du-Rhône	Allier	Aisne	Alpes-Maritimes	Ain	Alpes-de-Haute-Provence
Gironde	Deux-Sèvres	Hauts-de-Seine	Ardèche	Charente-Maritime	Bas-Rhin	Calvados	Ariège
Haute-Garonne	Dordogne	Nord	Ardennes	Côte-d'Or	Finistère	Gard	Cantal
Isère	Eure-et-Loir	Paris	Aube	Côtes-d'Armor	Haute-Savoie	Haut-Rhin	Corse-du-Sud
Loire-Atlantique	Haute-Vienne	Rhône	Aude	Doubs	Hérault	Loire	Creuse
Pas-de-Calais	Landes		Aveyron	Drôme	Ille-et-Vilaine	Loiret	Gers
Seine-et-Marne	Pyrénées-Orientales		Cher	Eure	Moselle	Maine-et-Loire	Haute-Corse
Seine-Maritime	Savoie		Corrèze	Indre-et-Loire	Oise	Meurthe-et-Moselle	Haute-Marne
Seine-Saint-Denis	Tarn		Haute-Loire	Manche	Var	Morbihan	Hautes-Alpes
Val-d'Oise	Vienne		Haute-Saône	Marne		Puy-de-Dôme	Lot
Val-de-Marne	Vosges		Hautes-Pyrénées	Saône-et-Loire		Pyrénées-Atlantiques	Lozère
Yvelines	Yonne		Indre	Sarthe		Vendée	Meuse
			Jura	Somme			Nièvre
			Loir-et-Cher	Vaucluse			Territoire de Belfort
			Lot-et-Garonne				
			Mayenne				
			Orne				
			Tarn-et-Garonne				

3.1 Geographical, topographical and urbanization description of the clusters

Clustering and datamining processes clearly reveal two types of territorial units: urban and rural units. Due to its high level of detail, our study also allows us to distinguish highly urban clusters and highly rural clusters. The following findings are based on the number of cities located in the territorial units and their population density. Specifically, cluster 1 and cluster 3 are urban clusters, but the degree of urbanity is higher for cluster 1. The departments belonging to cluster 1 include most of what we call the "major French regional cities," including Toulouse, Bordeaux, Grenoble, Nantes, Rouen and Versailles. The three largest French cities with the highest population density (Paris, Marseille and Lyon) are in cluster 3. Cluster 4 and cluster 8 are rural clusters, and the degree of rurality is higher for cluster 8. Cluster 1 and cluster 3 are at the top of the cluster ranking, both in terms of urban density and the total number of large cities. Specifically, 58.33% of cluster 1 departments and 100% of cluster 3 departments contain cities of category 2 or 3. The only city in category 1 (Paris) is located in cluster 3. By contrast, 0% of cluster 4 and cluster 8 departments contain cities of categories 1, 2

and 3. If one classes all French departments according to the total workforce population[18], Paris[19], the Nord, the Bouches-du-Rhône, the Rhône and the Hauts-de-Seine are the leading departments and are in cluster 3.

The geographical and topographical features of the departments contained in cluster 1 and cluster 3, on the one hand, and in cluster 4 and cluster 8, on the other, are also identifiable. Urban clusters are the least mountainous in France[20], with only 16.67% of cluster 1 departments containing mountains of over 1,500 meters against 0% of cluster 3 departments. In contrast, 50% of cluster 8 departments (22.22% of cluster 4 departments) contain mountains over 1,500 meters. Cluster 8, the most rural, is thus distinguished by the presence of mountains, an isolating factor. Note also that the French notion of isolation also requires identifying opportunities for maritime openings and waterways. In this regard, it is clear that the most rural clusters also have the least coastal geography[21]. Among the 8 clusters, cluster 4 (5.56%) and cluster 8 (14.29%) are again at the bottom of the cluster ranking with respect to coastal departments. Logically, cluster 1 (33.33%) and cluster 3 (40%) are at the top of the same ranking.

Finally, the results of our study should also allow us to assess the impact of borders on French mobility and consumption behaviors. While cluster 1 (bordering Spain) and cluster 3 (bordering Belgium) include 8.33% and 20% of border departments, respectively, cluster 4 (bordering Belgium, Switzerland and Spain) and cluster 8 (bordering Italy, Spain, Andorra, Belgium and Switzerland) comprise 16.67% and 35.71% of border departments, respectively. Cluster 6 includes the largest share of border departments at 44.44%.

3.2 Behavioral outcomes and mobility practices in France

Interesting results regarding mobility practices in France emerge from the analysis of our 8 clusters. We will present these results with respect to the use of private vehicles for commuting, the rate of public transport use, the rate of walking, the rate of two-wheeled vehicle use, and the share of employees who work outside of their home municipality. The classification of clusters based on the rate of private vehicle use for commuting and the public transport utilization rate exhibits near-perfect asymmetry. Table 2 shows that cluster 1 and cluster 3 have among the lowest rates of private vehicle use for commuting at 64.34% and 68.68%, respectively, while simultaneously having higher transport utilization rates, at 22.17% for cluster 1 and 16.08% for cluster 3. Conversely, in cluster 8, the rate

[18] Taking, for example, 2012 figures.

[19] The French capital is both a department and a city.

[20] The city of Grenoble, which exceeds 160,000 residents (since 2013) although lying in the mountainous region of the Alps is an exception.

[21] Metropolitan France has more than 3,400 kilometers of coastline (this figure increases to over 4,000 if one includes overseas communities).

of private vehicle use for commuting is the highest (75.39%) while the rate of transit use is the lowest (7.99%). Cluster 2 has a fleet utilization rate of 72.18%, which is relatively close to that of cluster 4 (72.89%), but has a public transit usage rate (13.11%) well above that of cluster 4 (10.94%)[22].

Table 2:

Utilization rates of means of transport for daily commuting (2016-2012)*

Clusters	Private car	Public transportation	Walking	Two wheels	Work outside municipality of residence
1	64,34%	22,17%	6,16%	3,79%	72,63%
2	72,18%	13,11%	6,87%	3,53%	67,04%
3	68,68%	16,08%	7,22%	3,80%	63,55%
4	72,89%	10,94%	7,70%	3,60%	62,35%
5	74,08%	9,65%	7,71%	3,65%	62,30%
6	74,29%	9,46%	7,68%	3,76%	62,28%
7	75,02%	8,91%	7,58%	3,73%	62,76%
8	75,39%	7,99%	8,01%	3,59%	61,31%

* Cluster average over the period

Our results show that when individuals choose between public transport and private vehicles, the decision is explained by the availability of transport infrastructure. The two most urban clusters show a lower use of private vehicles in commuting and greater use of public transport. Meanwhile cluster 8, the most rural cluster, exhibits the highest rate of vehicle use in commuting to work and the lowest rate of transport use.

Importantly, the share of commuters walking to work has an obvious relationship with the utilization rates for private vehicles and public transport. The rates of walking range from 6.16% (cluster 1) to 8.01% (cluster 8). Cluster 1 exhibits the lowest rate of walking, as it uses fewer cars and more public transport. Conversely, in cluster 8, residents walk more, meaning that there is less use of public transport and more car use. This suggests a negative correlation between public transport use and walking. However, other clusters do not support this correlation. Cluster 3 exhibits greater public transport use than cluster 2 (16.08% and 13.11%, respectively) and a higher walking rate (7.22% and 6.87%, respectively). Clusters 4, 5, 6 and 7 are similar in the majority of our findings (with respect to the rates of private vehicle use, public transport use and walking). The rates of two-wheeled vehicle use for commuting are very similar, ranging from 3.53% (cluster 2) to 3.80% (cluster 3) and there-

[22] To neutralize the numbering effect in the interpretation of our results, behavioral data are normalized by relating the figures to the total working population of the territorial units over time; consumption data are normalized by relating them to the total vehicle fleet for the categories and periods studied.

fore have no connection with the differences in the use of private vehicles (excluding motorcycles) or public transport in commuting. Regarding the rate of workers working outside their municipality of residence, cluster 1 and cluster 2 have rates of 72.63% and 67.04%. Clusters 3, 4, 5, 6, 7, and 8 have fairly similar rates, ranging between 61.31% and 63.55%.

3.3 The structure of the vehicle fleet and the choice of cars purchased

Despite substantial differences in mobility behavior (Table 2), Tables 3 and 4 report very small differences across clusters in the structure of the fleet and in the choice of cars purchased. In terms of gasoline engines, the proportion of cars purchased with engines of less than 6 horsepower per cluster ranges as follows: 19.26% for cluster 5 to 20.30% for cluster 1. The figure for engines between 6 and 7 horsepower ranges from 13.52% (cluster 8) to 14.22% (cluster 1). For vehicles with more than 8 horsepower, the figure ranges from 5.28% (cluster 2) to 5.79% (cluster 3). There is also little difference in the proportion of diesel engines for engines under 6 horsepower, with the share ranging from 22.91% (cluster 1) to 24.07% (cluster 8) and diesels with more than 6 horsepower, ranging from 36.35% (cluster 1) to 37.31% (cluster 5).

Table 3:

Fleet of private cars under 15 years (on january 1st of each year) by energy source and administrative power class (2006-2014)*

Clusters	Gasoline <6 horsepower	Gasoline [6;7] horsepower	Gasoline ≥8 horsepower	Diesel <6 horsepower	Diesel ≥6 horsepower	Gasoline LPG <6 horsepower	Gasoline LPG ≥6 horsepower	Electric cars	Undetermined energy source and/or adminitrative power class
1	20,30%	14,22%	5,60%	22,91%	36,35%	0,27%	0,31%	0,02%	0,03%
2	19,69%	13,71%	5,28%	23,67%	37,04%	0,26%	0,32%	0,02%	0,03%
3	19,99%	14,00%	5,79%	23,21%	36,38%	0,26%	0,32%	0,02%	0,03%
4	19,51%	13,65%	5,53%	23,74%	36,95%	0,26%	0,32%	0,02%	0,03%
5	19,26%	13,48%	5,31%	24,02%	37,31%	0,25%	0,32%	0,02%	0,03%
6	19,48%	13,61%	5,45%	23,87%	37,00%	0,24%	0,30%	0,02%	0,03%
7	19,40%	13,57%	5,36%	23,98%	37,10%	0,24%	0,30%	0,02%	0,03%
8	19,34%	13,52%	5,34%	24,07%	37,14%	0,24%	0,30%	0,02%	0,03%

* Cluster average over the period

Fleet of private cars under 15 years (on january 1st of each year) by administrative power class (highly detailed, 2006-2014)*

Clusters	[1;4] horsepower	5 horsepower	6 horsepower	7 horsepower	8 horsepower	9 horsepower	[10;11] horsepower	≥12 horsepower	Undetermined horsepower
1	18,92%	24,78%	23,07%	16,88%	6,26%	3,75%	3,47%	2,86%	0,01%
2	18,78%	25,04%	23,53%	16,56%	6,21%	3,65%	3,43%	2,79%	0,01%
3	18,79%	24,87%	23,21%	16,53%	6,20%	3,73%	3,56%	3,09%	0,01%
4	18,64%	25,07%	23,60%	16,37%	6,16%	3,66%	3,50%	2,99%	0,01%
5	18,59%	25,16%	23,83%	16,31%	6,16%	3,60%	3,44%	2,91%	0,01%
6	18,62%	25,18%	23,64%	16,22%	6,18%	3,65%	3,49%	3,00%	0,01%
7	18,57%	25,26%	23,75%	16,20%	6,17%	3,62%	3,46%	2,95%	0,01%
8	18,55%	25,30%	23,79%	16,15%	6,17%	3,61%	3,47%	2,95%	0,01%

* Cluster average over the period

Table 4:
Total number of new cars registered in the year by CO2 emissions (2006-2014)*

Clusters	≤100 g	[101;120] g	[121;130] g	[131;160] g	[161;165] g	[166;200] g	[201;250] g	>250 g	Undedermined carbon emissions
1	6,85%	35,92%	11,22%	34,50%	1,44%	7,49%	2,05%	0,51%	0,02%
2	6,91%	36,26%	11,21%	34,27%	1,45%	7,35%	2,02%	0,51%	0,01%
3	7,14%	35,73%	11,29%	34,07%	1,45%	7,55%	2,14%	0,61%	0,02%
4	7,09%	35,98%	11,28%	34,05%	1,45%	7,44%	2,10%	0,59%	0,02%
5	7,08%	36,16%	11,29%	34,05%	1,45%	7,33%	2,06%	0,57%	0,02%
6	7,08%	36,33%	11,34%	33,88%	1,51%	7,26%	2,02%	0,56%	0,01%
7	7,04%	36,38%	11,34%	33,95%	1,50%	7,22%	2,00%	0,56%	0,01%
8	7,07%	36,43%	11,33%	33,93%	1,50%	7,18%	1,99%	0,55%	0,01%

* Cluster average over the period

Finally, for uncommon engine types such as LPG vehicles [23] with less than 6 horsepower, the proportions by cluster are equally similar, ranging from 0.24% (clusters 6, 7 and 8) to 0.27% (cluster 1). For LPG vehicles over 6 horsepower, the figures are between 0.30% (cluster 6, 7 and 8) and 0.32% (clusters 2, 3, 4 and 5). For electric cars, the share is the same in all clusters at 0.02%.

A more detailed analysis of the share of vehicles by engine type yields similar proportions across clusters. For example, for vehicles with 1 to 4 horsepower, the share ranges from 18.55% (cluster 8) to 18.92% (cluster 1). The pattern is similar for the purchase of vehicles with 7 or 9 horsepower, where the shares are between 16.15% (cluster 8) and 16.88% (cluster 1) and 3.61% (cluster 8) and 3.75% (cluster 1), respectively. Table 4 presents the details of purchases by horsepower and cluster. We find that the shares of purchases for the various vehicle types across clusters are very similar.

An analysis of vehicle purchases based on CO2 emissions yields similar results. For vehicles emitting less than 100g of CO2, purchase shares by cluster range from 6.85% (cluster 1) to 7.14% (cluster 3). For vehicles emitting 161g to 165g, the shares range from 1.44% (cluster 1) to 1.51% (cluster 6). A similar pattern is observed for vehicles emitting more than 250g, where the shares range from 0.51% (clusters 1 and 2) to 0.61% (cluster 3).

Consumer choices with respect to the vehicle type change very little across clusters. This result relates to the substantial differences in mobility behavior across clusters.

3.4 Discussion

An individual using public transport does not necessarily choose less powerful engines or those that emit less CO2. Such decision should therefore be explained through other factors, including tax

[23] Liquefied petroleum gas

incentives, industrial policy and vehicles offered.

Concerning tax incentives in France, the tax on diesel is much lower than that on gasoline (a difference of 0.15 euros per liter), and historically, the two French brands of Renault and Peugeot have specialized in diesel. This tax difference explains the high proportion of diesel vehicles purchased in France, ranging from 70.48% (cluster 1) to 71.38% (cluster 2), but it shows that the distance (employees working outside their home municipality) and the need to drive to work (vehicle use rate for commuting) has little influence on the choice of this type of engine.

Tax incentives to purchase vehicles, i.e., "bonus- malus," do not seem to explain the dissimilarity in vehicle purchases. Electric cars (-20g CO2 emissions rate) are eligible for a bonus of 6,300 euros but represent only 0.02% of purchases in all clusters. Clusters 1 and 3, as urban areas, have the same rate of electric vehicle purchases as do other clusters.

Penalties begin to apply at 131g of CO2 emissions (a malus of 150 euros). The French are buying more new cars with emissions ranging from 131g to 160g (entailing a penalty between 150 and 2,200 euros) than cars of the lower emissions class (not subject to a penalty, between 121g and 130g). For the 131g to 160g CO2 category, the shares of purchases vary from 33.9% (cluster 7) to 34.5% (cluster 1), while those of the lower emission class range from 11.2% (cluster 2) to 11.34% (cluster 6).

This difference calls into question the limited effectiveness of tax incentives in affecting vehicle purchasing decisions, especially given that the share of new cars purchased with CO2 emissions between 101g and 120g varies between 35.7% (cluster 3) and 36.4% (cluster 8). We therefore find that in terms of CO2 emissions, the differences across clusters in the choice of vehicle types are small but the differences among vehicle categories are important and do not respond to tax incentives implied by the "bonus-malus".

Differences in the mobility of the French, as well as tax incentives for purchasing vehicles, appear to have little influence on vehicle choice. Therefore, one should consider the question of the supply of vehicles in France and the industrial policies of France and the EU with respect to the automotive industry.

4 Conclusion

This paper analyzes the impact of individual mobility differences in France on the structure of the vehicle fleet and purchasing behavior for new cars with respect to CO2 emissions. Our database and analytical method allow us to obtain interesting results. The first result is that mobility behavior such as the use of private vehicles and public transport in daily commutes varies significantly from one cluster to another. This reflects the specific characteristics of each cluster, with some being urban

and with developed transport infrastructure while others are rural. The second result is that, despite these differences in mobility, the fleet structure changes very little across clusters, as captured by vehicle purchases.

This result is interesting because it shows that even if private vehicle use differs depending on the ability to use public transport, the choice of vehicle types purchased is very close.

Encouraging the purchase of vehicles with engines emitting little CO2 is a challenge in the fight against global warming. In the case of France, we show that despite regional differences (urban, rural, public transport, etc.) and a tax advantage for low CO2 emissions, French vehicle buyers' choices are similar wherever they live.

This raises the issues of industrial policy and product offering. Environmental policies focus primarily on price signals and consumer adjustment via tax incentives to encourage certain behaviors, but they offer very little in terms of supply incentives. To implement an effective environmental policy, it is important to invest heavily in transport infrastructure in order to create incentives to orientate consumer behavior. It is also crucial to provide direct incentives for the production and supply of more efficient and cleaner vehicles.

Acknowledgements

The authors would like to thank Prof. Thomas Stocker (University of Bern, Former Co-Chair of Working Group 1 of the UN IPCC), Prof. Julien Chevallier (University Paris 8) and Dr. Simon Porcher (University Paris 1 Sorbonne) for enriching discussions on decarbonizing policies in France.

References

[1] Agyemang-Bonsu K, Dontwi I, Tutu-Benefoh D et al. (2010) Traffic-data driven modelling of vehicular emissions using COPERT III in Ghana: a case study of Kumasi. *American Journal of Scientific Industrial Research* 1(1): 32-40.

[2] Altuntas S, Dereli T and Kusiak A (2016) Assessment of corporate innovation capability with a data-mining approach: industrial case studies. *Computers & Industrial Engineering* 102: 58-68.

[3] Bandaru S, Ng A AHC and Deb K. (2017) Data mining methods for knowledge discovery in multi-objective optimization: Part B - New developments and applications. *Expert Systems With Applications* 70: 119-138.

[4] Bond T, Doherty S, Fahey D et al. (2013) Bounding the role of black carbon in the climate system: a scientific assessment. *Journal of Geophysical Research Atmospheres* 118(11): 5380-5552.

[5] Boroumand RH, Goutte S and Porcher T (2015) *20 idées recues sur l'énergie*. Ed. De Boeck.

[6] Boyce J (1994) Inequality as a cause of environmental degradation. *Ecological Economics* 11: 169-178.

[7] Boyce J, Klemer A, Templet P and Willis, C (1999) Power distribution, the environment, and public health: a state-level analysis. *Ecological Economics* 29: 127-140.

[8] Buchs M, Schnepf S (2013) Who emits most? Associations between socio-economic factors and UK households' home energy, transport, indirect and total CO2 emissions. *Ecological Economics* 90: 114-123.

[9] Cames M and Helmers E (2013) Critical evaluation of the European diesel car boom - global comparison, environmental effects and various national strategies. *Environmental Sciences Europe* 25(1): 15-37.

[10] Chancel L (2014) Are younger generations higher carbon emitters than their elders?: Inequalities, generations and CO2 emissions in France and in the USA. *Ecological Economics* 100: 195-207.

[11] Chevallier J and Goutte S (2016) Operation Management of the Power Company in Presence of Carbon Costs. *Commodity Markets*, edited by Balcilar, M. and Hammoudeh, S. Emerald Group Publishing.

[12] Curiel-Esparza, Mazario-Diez JL, Canto-Perello J et al. (2016) Prioritization by consensus of enhancements for sustainable mobility in urban areas. *Environmental Science & Policy* 55: 248-257.

[13] Chowdhury S, Ceder A and Schwalger B (2015) The effects of travel time and cost savings on commuters' decision to travel on public transport routes involving transfers. *Journal of Transport Geography* 43: 151-159.

[14] Davydova-Belitskaya V and Skiba YN (2003) On the estimation of impact of vehicular emissions. *Ecological Modelling* 166: 169-184.

[15] Dockery D, Pope C, Xu X et al. (1993) An association between air pollution and mortality in six U.S. cities. *The New England Journal of Medicine* 329: 1753-1759.

[16] EU (2010) *Europe 2020: A European strategy for smart, sustainable and inclusive growth*. EU. Available at: http://ec.europa.eu/europe2020

[17] Fann N, Lamson A, Anenberg S et al. (2012) Estimating the national public health burden associated with exposure to ambient PM2.5 and ozone. *Risk Analysis* 32(1): 81-95.

[18] Fontaras G and Dilara P (2012) The evolution of European passenger car characteristics 2000-2010. *Energy Policy* 49: 719-730.

[19] Fullerton D, Bruce N and Gordon S (2008) Indoor air pollution from biomass fuel smoke is a major health concern in the developing world. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 102(9): 843-851.

[20] González PF, Landajo M and Presno MJ (2014) The driving forces behind changes in CO2 emission levels in EU-27. Differences between member states. *Environmental Science & Policy* 38: 11-16.

[21] Hedenus F and Azar C (2005) Estimates of trends in global income and resource inequalities. *Ecological economics* 55: 351-364.

[22] Holian M and Kahn M (2015) Household carbon emissions from driving and center city quality of life. Ecological Economics 116: 362-368.

[23] IEA/OECD (2010) *Energy technology perspectives 2010: scenarios & strategies to 2050*. Paris: International Energy Agency and Organization for Economic Cooperation and Development.

[24] IPCC (2014) *Climate Change 2014: Synthesis Report*. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

[25] Kahn Ribeiro S, Kobayashi S, Beuthe M et al. (2007) Transport and its infrastructure. In: Metz B, Davidson O, Bosch P et al. *Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

[26] Kholod N, Evans M, Gusev E et al. (2016) A methodology for calculating transport emissions in cities with limited traffic data: Case study of diesel particulates and black carbon emissions in Murmansk. *Science of the Total Environment* 547: 305-313.

[27] Kilkis S (2016) Sustainable development of energy, water and environment systems index for Southeast European cities. *Journal of Cleaner Production* 130: 222-234.

[28] Kingham S, Dickinson J and Copsey S (2001) Travelling to work: will people move out of their cars. *Transport Policy* 8: 151-160.

[29] Kovacs K, Haight R, Jung S et al. (2013) The marginal cost of carbon abatement from planting street trees in New York City. *Ecological economics* 95: 1-10.

[30] Liu Y, Hong Z and Liu Y (2016) Do driving restriction policies effectively motivate commuters to use public transportation? *Energy Policy* 90: 253-261.

[31] Longhin E, Gualtieri M, Capasso L et al. (2016) Physico-chemical properties and biological effects of diesel and biomass particles. *Environmental Pollution* 215: 366-375.

[32] Mackett R and Thoreau R (2015) Transport, social exclusion and health. *Journal of Transport & Health* 2: 610-617.

[33] McGranahan G and Satterthwaite D (2002) The environmental dimensions of sustainable development for cities. *Geography* 87(3): 213-226.

[34] Nguyen PM, Haghverdi A, De Pue J et al. (2017) Comparison of statistical regression and data-mining techniques in estimating soil water retention of tropical delta soils. *Biosystems engineering* 153: 12-27.

[35] Palencia J, Furubayashi T and Nakata T (2012) Energy use and CO2 emissions reduction potential in passenger car fleet using zero emission vehicles and lightweight materials. *Energy* 48: 548-565.

[36] Parry I, Walls M and Harrington W (2007) Automobile externalities and policies. *Journal of Economic Literature* 45(2): 373-399.

[37] Peer S, Knockaert J and Verhoef E (2016) Train commuters' scheduling preferences: Evidence from a large-scale peak avoidance experiment. *Transportation Research Part B* 83: 314-333.

[38] Pérez-Suárez R and López-Menéndez AJ (2015) Growing green? Forecasting CO2 emissions with Environmental Kuznets Curves and Logistic Growth Models. *Environmental Science & Policy* 54: 428-437.

[39] Remuzgo L and Sarabia JM (2015) International inequality in CO2 emissions: A new factorial decomposition based on Kaya factors. *Environmental Science & Policy* 54: 15-24.

[40] Rustemoglu H and Andrés AR (2016) Determinants of CO2 emissions in Brazil and Russia between 1992 and 2011: A decomposition analysis. *Environmental Science & Policy* 58: 95-106.

[41] Schellnhuber HJ and Svirejeva-Hopkins A (2008) Urban expansion and its contribution to the regional carbon emissions: Using the model based on the population density distribution. *Ecological Modelling* 216: 208-216.

[42] Schipper L and Fulton L (2009) Dazzled by diesel? The impact on carbon dioxide emissions of the shift to diesels in Europe through 2009. *Energy Policy* 54: 3-10.

[43] Scruggs LA (1998) Political and economic inequality and the environment. *Ecological Economics* 26: 259-275.

[44] Sharif B, Makowski D, Plauborg F et al. (2017) Comparison of regression techniques to predict response of oilseed rape yield to variation in climatic conditions in Denmark. *European Journal of Agronomy* 82: 11-20.

[45] Sims R, Schaeffer R, Creutzig F et al. (2014) Transport. In: Edenhofer O, Pichs-Madruga R, Sokona Y et al. Climate *Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

[46] Su F, Zhou C et al. (2004) A data-mining approach to determine the spatio-temporal relationship between environmental factors and fish distribution. *Ecological Modelling* 174: 421-431.

[47] Smith T, Axon C and Darton R (2013) The impact on human health of car-related air pollution in the UK, 1995-2005. *Atmospheric Environment* 77: 260-266.

[48] St-Louis E, Manaugh K, Van Lierop D et al. (2014) The happy commuter: A comparison of commuter satisfaction across modes. *Transportation Research Part F* 26: 160-170.

[49] Terrier C (1986) Les déplacements domicile-travail en France : évolution de 1975 ï£¡à 1982. *Espace, populations, sociétés* 2: 333-342.

[50] Torras M and Boyce J (1998) Income, inequality, and pollution: a reassessment of the environmental Kuznets Curve. *Ecological Economics* 25: 147-160.

[51] Venkatesh G, Chan A and Bratteb H (2014) Understanding the water-energy-carbon nexus in urban water utilities: comparison of four city case studies and the relevant influencing factors. *Energy* 75: 153-166.

[52] Vermote L, Macharis C, Hollevoet J and Putman K (2014) Participatory evaluation of regional light rail scenarios: A Flemish case on sustainable mobility and land-use. *Environmental Science & Policy* 37: 101-120.

[53] Villeneuve A (1970) Les déplacements domicile-travail. *Economie et statistique* 17: 3-16.

[54] Wang L, Li L, Wu B et al. (2013) Private Car Switched to Public Transit by Commuters, in Shanghai, China. *Procedia - Social and Behavioral Sciences* 96: 1293-1303.

[55] Wang L and Wang D (2016) Geography of urban life satisfaction: An empirical study of Beijing. *Travel Behaviour and Society* 5: 14-22.

[56] WCED (1987), Our Common Future. Oxford: Oxford University Press

[57] White T (2007) Sharing resources: the global distribution of the ecological footprint. *Ecological Economics* 64: 402-410.

[58] WHO (2014) 7 Million Premature Deaths Annually Linked to Air Pollution. *World Health Organization*. Available at: http://www.who.int/mediacentre/news/releases/2014/air-pollution/en/

[59] Zwickl K, Ash M and Boyce J (2014) Regional variation in environmental inequality: Industrial air toxics exposure in U.S. cities. *Ecological Economics* 107: 494-509.

Appendix

Table 5:
Behavioral main variables' table

Variables available for communal, departmental and regional territorial units*
Total population of the territorial unit
Total active workforce population aged 15 or more (2006-2012)
Number of workers using the private car to get to the workplace (2006-2012)
Private car utilization rate to get to the workplace (number reported to the active workforce population, 2006-2012)
Number of workers using public transportation to get to the workplace (2006-2012)
Public transportation utilization rate to get to the workplace (number reported to the active workforce population, 2006-2012)
Number of workers using walking to get to the workplace (2006-2012)
Walking utilization rates to get to the workplace (number reported to the active workforce population, 2006-2012)
Number of workers using two wheels to get to the workplace (2006-2012)
Two wheels utilization rate to get to the workplace (number reported to the labor force, 2006-2012)
Number of workers working outside the municipality of residence (2006-2012)
Rates of workers working outside the municipality of residence (number reported to the labor force, 2006-2012)

* Annual and periodic data

Table 6:
Consumption main variables' table

Variables available for departmental and regional territorial units*
Fleet of private cars under 15 years (on january the 1st of each year) by energy source and administrative power class (2006-2014):
Gasoline <6 horsepower
Gasoline [6;7] horsepower
Gasoline ≥8 horsepower
Gasoline LPG <6 horsepower
Gasoline LPG ≥6 horsepower
Diesel <6 horsepower
Diesel ≥6 horsepower
Electric cars
Gasoline + super ethanol <6 horsepower
Gasoline + super ethanol [6;7] horsepower
Gasoline + super ethanol ≥8 horsepower
Electricity + gasoline <6 horsepower
Electricity + gasoline ≥6 horsepower
Electricity + diesel
Undetermined energy source and/or adminitrative power class
Fleet of private cars under 15 years (on january the 1st of each year) by administrative power class (highly detailed, 2006-2014):
[1;4] horsepower
5 horsepower
6 horsepower
7 horsepower
8 horsepower
9 horsepower
[10;11] horsepower
≥12 horsepower
Undetermined horsepower
Total number of new cars registered in the year by CO2 emissions (2006-2014):
≤100 g
[101;120] g
[121;130] g
[131;160] g
[161;165] g
[166;200] g
[201;250] g
>250 g
Undedermined carbon emissions
Total number of new cars registered in the year (2006-2014)
Total number of diesel new cars registered in the year (2006-2014)

* Annual and periodic data

Table 7:
Geographical main variables' table (by size of the territorial unit)

Communal variables
Municipality's geographic code (INSEE)
Municipality's name
Department's number
Department's name
Indicator of border department
Region's name
Telephone indicator

Departmental variables
Number of cities in category 1 (Paris, >2 million inhabitants)
Number of cities in category 2 (from 200 000 to 1 999 999 inhabitants)
Number of cities in category 3 (from 100 000 to 199 999 inhabitants)
Number of cities in category 4 (<100 000 inhabitants)
Department's number
Department's name
Indicator of border department
Region's name
Telephone indicator

Regional variables
Number of cities in category 1 (Paris, >2 million inhabitants)
Number of cities in category 2 (from 200 000 to 1 999 999 inhabitants)
Number of cities in category 3 (from 100 000 to 199 999 inhabitants)
Number of cities in category 4 (<100 000 inhabitants)
Region's name
Indicator of border region
Telephone indicator

Figure 6: Cluster segmentation for $K = 3$ and $K = 5$

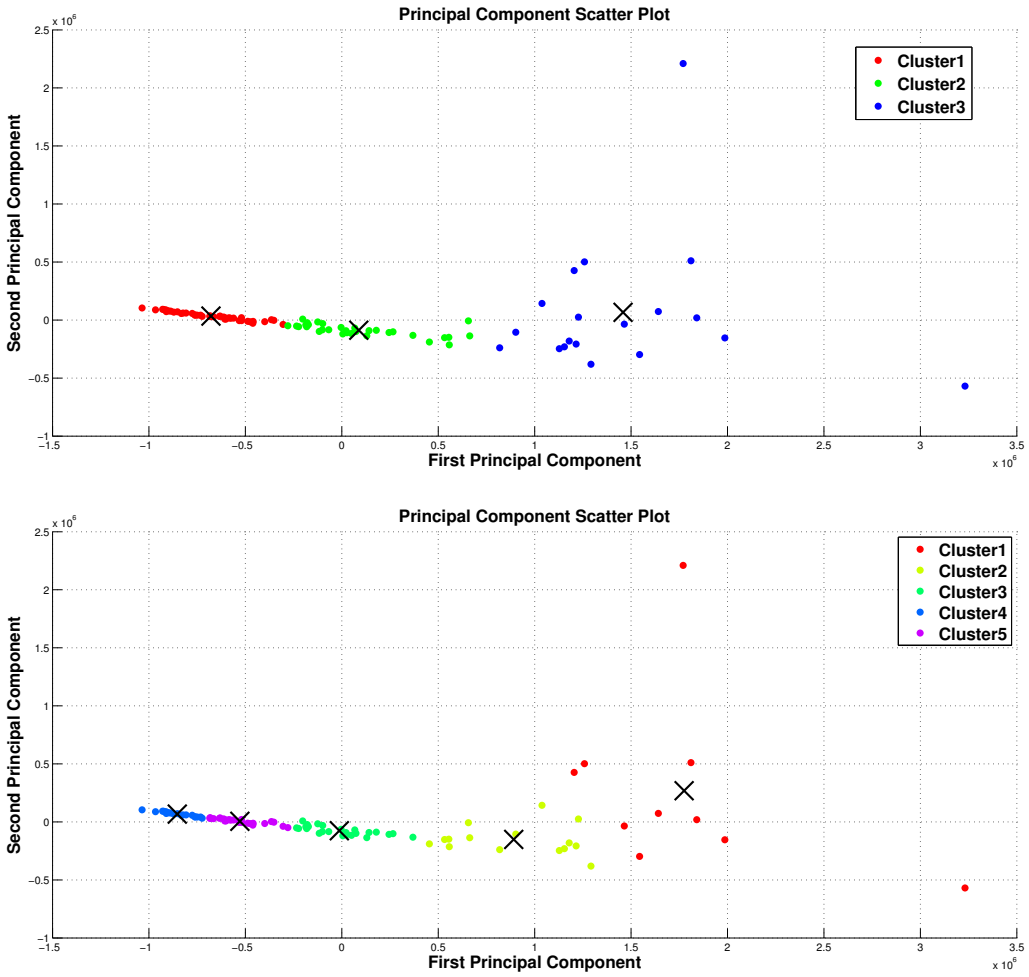


Figure 7: Cluster segmentation for $K = 8$ for each cluster (ordered Cluster 1 to 4)

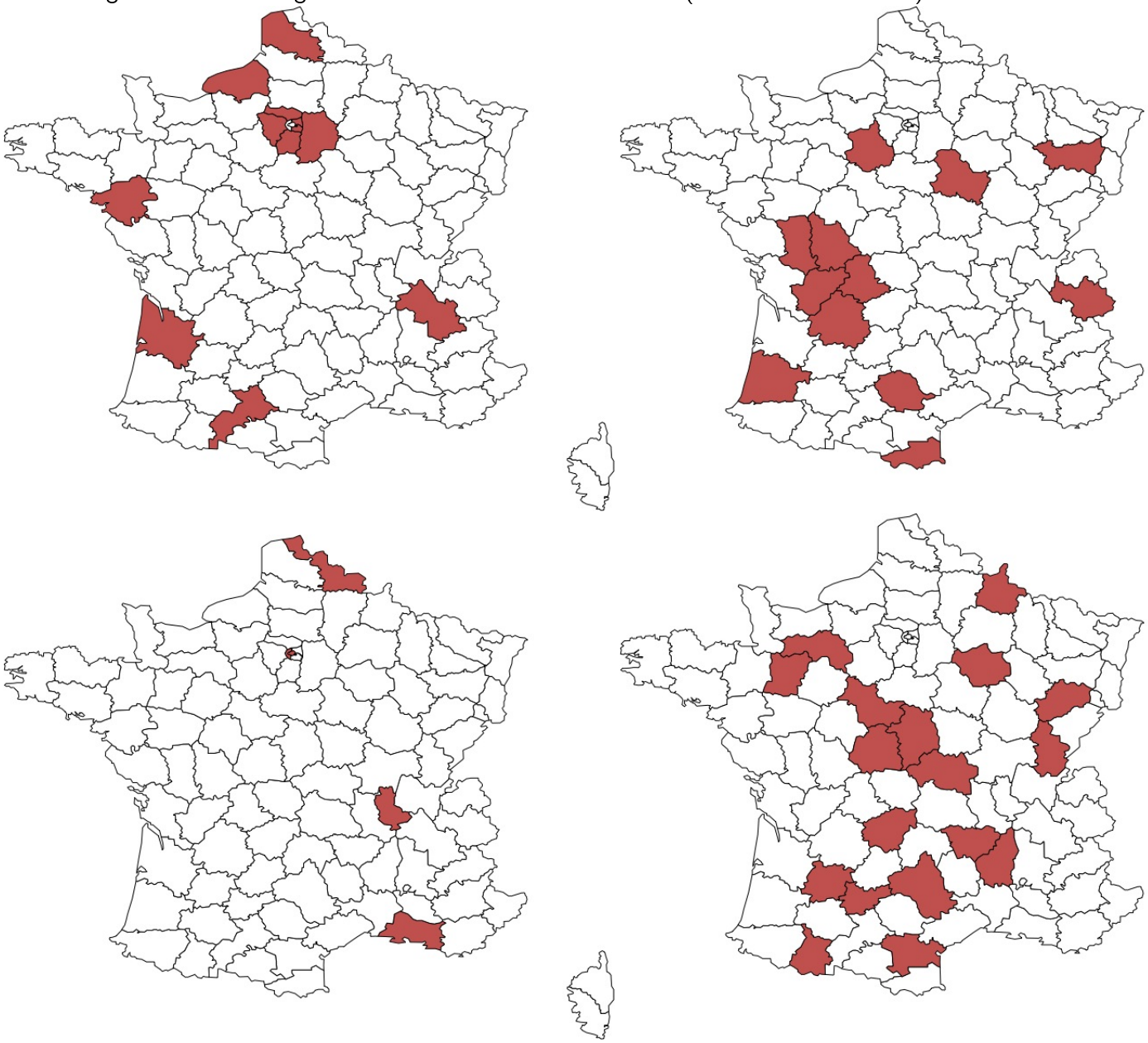
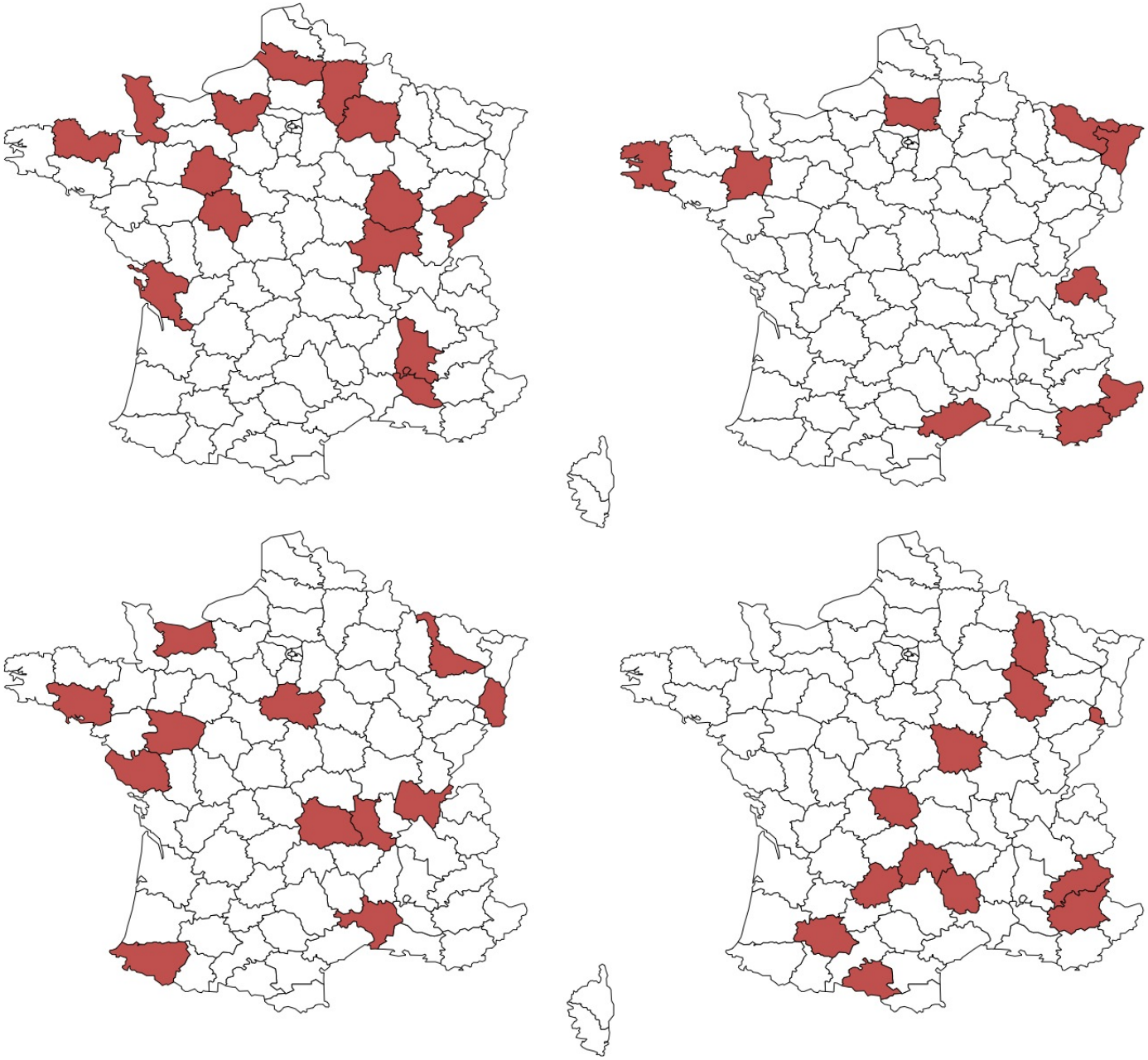



Figure 8: Cluster segmentation for $K = 8$ for each cluster (ordered Cluster 5 to 8)