Car-fuel poverty: determinants and policy implications for France

Ariane Bousquet, Maria-Eugenia Sanin

Car-fuel poverty: determinants and policy implications for France

Ariane Bousquet[a,b], Maria-Eugenia Sanin[a]

[a] *Université Paris-Saclay, Univ Evry, EPEE, 91025, Evry-Courcouronnes, France*
[b] *Direction de la Recherche, Renault*

ARTICLE INFO

Keywords:
Transport poverty
Fuel poverty
Car-fuel poverty
Inclusion
France
Europe

ABSTRACT

In the face of inflation following the Ukrainian crisis, several European governments implemented a generalized gasoline subsidy. In contrast, the reduction of fossil-fuel consumption is crucial to mitigate the current energy and climate crises. Fuel consumption for transport increases with income, making rich households the main beneficiaries of generalized subsidies. In this context, a thorough investigation of the nature of vulnerability to rising gasoline prices is needed to formulate targeted policies. Herein, we contribute in this line for the case of France. Firstly, we develop three metrics of car-fuel poverty. Secondly, we use multivariate statistic analysis to identify car-fuel-poor household profiles. Then, we estimate the socioeconomic determinants of such vulnerability. We find that, aside from income, household composition, region, access to public transport, and house ownership significantly impact the probability of being car-fuel-poor. Then, using our results, we evaluate the impact of recent subsidies implemented in France and suggest alternative targeted policies. We find that the policies that have been implemented are regressive and incur in inclusion and exclusion errors. Instead, a targeted subsidy fully compensating the car-fuel-poor would have been efficient implying, at the same time, important government savings.

1. Introduction

Social and environmental justice are at the heart of the current energy crisis of soaring prices caused by the post-Covid demand recovery together with the Ukrainian conflict. This could represent an opportunity to decrease the use of fossil fuels for transport, both by decreasing the use of cars and by switching to less polluting vehicles (intensive and extensive margins). Yet, environmental policies aiming at reducing the consumption of fossil fuels may reinforce existing inequalities and create new ones (Belaïd (2022a)) while generalized subsidies that protect the population against this type of inflation may deepen inequalities. These issues have appeared in the economic literature in the 1970s under the umbrella concept of *energy poverty*. Since then, European countries have created institutions, like the European Observatory for Energy Poverty (EPOV), to measure and monitor energy poverty. Yet, the focus of this concept is on the use of domestic fuels, disregarding fuel consumption for transport. Instead, the issue of poverty in transport has mainly been studied through the lens of inequalities in access to mobility solutions and the resulting social exclusion (Lucas, Mattioli, Verlinghieri and Guzman (2016); Lucas (2012)) with little or no attention given to fuel poverty related to the use of the car, or what we will call from now on car-fuel poverty.[1]

In this line, we make several contributions. First, we bridge the gap between the literature on energy poverty and transport poverty by adapting energy affordability indicators - which are extensible used and accepted - to transport. We then apply these metrics to a representative sample of the French population using the most recent National Household Travel Survey (*Enquete Mobilité des Personnes* 2018-2019), which provides information on household socio-economic characteristics, equipment, and travel. We find that 3.2% of households have both high fuel expenditures and a residual income after fuel consumption for transport that is under the national poverty line, what we will consider as Low Income High Transport Cost (LIHTC). Similarly, we find that 31% have fuel expenditures for transport that are less than half the population´s median (M/2). Finally, we find that 20% of households have a budget share for car-fuel that is dis-proportionally high, above twice the national median (2M).

A second contribution is the identification of car-fuel profiles that should be targeted by policy using statistical clustering methods. These methods, never applied to transport poverty research so far, allow us to identify three typical profiles of households in situations of car-fuel poverty for each of the three indicators used (LIHTC, M/2,

ariane.bousquet@universite-paris-saclay.fr (A. Bousquet); eugenia.sanin@univ-evry.fr (M. Sanin)

[1] The only exception is Matioli et al. (2018) that defines a specific metrics for the UK as we will discuss in the next section.

2M). The plurality of these profiles demonstrates the importance of choosing a relevant indicator before making policy recommendations.

A third contribution is the econometric estimation of the determinant factors affecting the probability of becoming car-fuel poor. Our results show that the most significant determinants are: (i) for LIHTC, the number of cars, specific socio-professional categories of the household´s head, being a tenant in isolated rural areas and with poor access to public transport, (ii) for 2M, being multi-motorised, specific socio-professional categories, being a tenant in either commuting or isolated rural areas with little or poor access to public transports, and (iii) for M/2, owning only one car, living in urban centers with access to public transport, specific socio-professional categories and belonging to the poorest incomes deciles.

In 2022, many European countries have implemented generalized gasoline subsidies to compensate households for their purchasing-power loss. A last contribution is to use our results to suggest alternative ways to target subsidies. In this regard we find that present schemes have failed to sufficiently compensate vulnerable population for their purchase power loss while including non-vulnerable households under the umbrella of the subsidy. We also show that targeted policies could have lead to major savings while compensating completely the car-fuel poor.

The rest of this paper is as follows. In section 2, we review the literature on energy and transport poverty in order to place our contribution in the existing literature. In section 3, we describe the methods and data used in this study. We then present results in terms of car-fuel profiles and their determinants in section 4. In section 4.3, we discuss the implications in terms of public policies, suggesting alternative subsidies and targeting methods, and in section 5 we present some concluding remarks.

2. Literature Review

The issue of car-fuel poverty may be considered at the crossroads between energy and transport poverty. In this section we will first review the definitions of fuel poverty that we will adapt to the car-fuel context as well as its application in France. We then will review the few studies that have been done in the transport literature where most have used very different definitions and, with the exception of two, have not studied determinants.

Energy poverty can be defined as the ‘inability to obtain sufficient energy for essential services such as cooking, heating, cooling and household lighting’ (Belaïd (2018)). It is a generic concept that at least encompasses two aspects: affordability and access to energy services (or to infrastructure). Herein, our main focus is on the affordability aspect, and more specifically on the use of car fuels. The closest concept in the energy poverty literature is *fuel poverty*, which is defined as the inability to afford adequate energy services at home.

The first measure of fuel poverty appeared in Boardman (1991), in which she defined fuel-poor households as those with a budget share for domestic fuel greater than twice the population median (2M). Applying this measure to a 1988 survey in the United Kingdom, she established that the median budget share was 5%, so her definition was generalized as the *10% indicator*. Her definition infused both academic literature (Legendre and Ricci (2015); Romero, Linares and López (2018)) and public policy, first in the UK, but also in the rest of the EU (Thema and Vondung (2020); Oxley (2023)). It was the official measure of poverty in the UK from 2001 to 2012.

Other expenditure-based indicators are used at the EU level, such as the “half the median” (M/2) indicator that captures the under-consumption of fuel inside the home relative to the national median. In 2012, Hill’s report (Hills (2012)) proposed alternative measures: the *After fuel cost* indicator, which identifies people falling below a poverty line after paying for their domestic energy expenditures, and the *Low-Income High Cost* (LIHC) indicator, which defines as energy poor households that have both (1) required fuel costs above the national median and (2) "were they to spend that amount, they would be left with a residual income below the official poverty line". Widely used and discussed in academic literature (Romero et al. (2018); Belaïd (2018, 2022b); Legendre and Ricci (2015); Belaïd and Flambard (2023)), the LIHC indicator also became the official measure of fuel poverty in the UK from 2013 to 2021 (Oxley (2023)). It was in turn replaced by the Low-Income Low Energy Efficiency (LILEE) in 2021, which targets households with low incomes living in energy-inefficient housing.

In France, the monitoring of fuel poverty began after the *Grenelle Environment Forum* in 2007, with measures inspired by the 10% indicator. Since 2016, several indicators have been used: (i) the 10% indicator restricted to the first three income deciles; (ii) versions of the LIHC indicator; and, (iii) a declared qualitative indicator of discomfort and cold in the home. To tackle fuel poverty, France has set up the annual energy voucher scheme, dedicated to the payment of housing energy bills. This measure was generalized in 2018 to the entire national territory. Moreover, to reduce energy bills and accelerate the energy transition, the government has implemented aid for housing renovation,

with mixed results. In Appendix 6.1 we perform a complete review of the fuel poverty literature that is relevant to this study detailing the alternative definitions used.

In line with the energy poverty literature, Lowans, Furszyfer Del Rio, Sovacool, Rooney and Foley, 2021 suggests defining transport poverty as the "inability to adequately meet commuting and mobility needs in a household". In an earlier review, Lucas et al. (2016) considers that individuals are transport-poor if, to satisfy their daily basic activity needs, at least one of the following conditions apply : (i) "There is no transport option available that is suited to the individual's physical condition and capabilities"; (ii) "The existing transport options do not reach destinations where the individual can fulfill his/her daily activity needs, to maintain a reasonable quality of life"; (iii) "The necessary weekly amount spent on transport leaves the household with a residual income below the official poverty line"; (iv) "The individual needs to spend an excessive amount of time traveling, leading to time poverty or social isolation"; (v) "The prevailing travel conditions are dangerous, unsafe or unhealthy for the individual." Several aspects emerge from this definition; they are not mutually exclusive but imply different policy responses (Lucas et al. (2016)). More generally, the previous definition can be summarized in 4 categories of problems as in Table 1.

Concept	Description
Affordability problem	lack of financial resources to afford available transport options
Mobility poverty	lack of adequate (often motorised) transport options that lead travel difficulties.
Accessibility poverty	households experience difficulties reaching essential activities and services (for e.g. employment, healthcare, school or food shops) "*at a reasonable time, ease, and cost.*" Lowans et al. (2021)
Externalities Exposure	exposure to transport externalities, such as pollution.

Table 1: Transport poverty aspects, based on Lucas et al. (2016).

The UK has been a leader in both policy and research on the issue of transport poverty. The first wave of research occurred after the 2003 Social Exclusion Unit report that established a link between social exclusion and transport disadvantage. The focus of this research was to measure inequalities in *mobility* (number of trips, distance traveled) and in *accessibility* to essential activities (employment, healthcare, school, food shops, etc.) for different social groups (Lucas (2012)). Households in a situation of intense mobility poverty are usually very disadvantaged in terms of income and location and do not own a car. Relatively less academic attention has been given to car owners with difficulties affording their travel needs in car-dependent societies.

The strand of literature focused on transport *vulnerability*, a concept imported from the climate science literature and used in IPCC reports (Adger (2006)). This definition of vulnerability covers three aspects (Mattioli, Philips, Anable and Chatterton (2019)): (i) exposure to fuel costs (ii) sensitivity (often measured by income), and (iii) resilience or adaptive capacity, measuring the ability to shift to other transport options, should car-fuel prices increase. Vulnerability measures are therefore composite indicators accounting for the four aspects defined earlier and are often used in socio-spatial studies. For example, Sustrans (2012) maps the risk of transport poverty using three criteria: (i) low income, (ii) areas where people live more than one mile away from a bus station and, (iii) reaching essential services takes more than an hour by public transport or walking. Mattioli et al. (2019) defines a composite indicator of exposure (define as the ratio of motor fuel expenditure over income), sensitivity (income) and adaptive capacity (sum of estimated travel time to reach 8 key services).

Affordability metrics in the transport poverty literature are often adapted from the fuel poverty literature and use household survey data. Table 2 presents the main measures of affordability and vulnerability upon which we build on. The concept of *Forced Car Ownership* (FCO) emerged in Australian research to define households that have low income while being highly motorized. In the EU, it has been used by Mattioli (2017b) to characterize a household that "(i) owns at least one car and (ii) reports difficulties to afford at least one of five items (rent, mortgage, household maintenance, energy bills, and food)". Similarly, the Royal Automobile Club Foundation (RAC (2012)) uses the controversial 10% indicator, which results in a very large number of UK households being defined as fuel-poor (21 million). Nicolas, Vanco and Verry (2012) derives the burden of daily mobility as the ratio of daily mobility expenditures over equivalent income. It defines as transport-poor households those that present a ratio between their daily mobility expenditures and equivalent income that is in the 20% that spend the most.

Mattioli, Wadud and Lucas (2018) modifies the LIHC indicator with a combination of the "After Fuel Costs" and 2M indicators. Precisely, for Mattioli et al. (2018) households in Car-Related Economic Stress (CRES) have: (i) equivalent income after housing and running motor vehicle costs below 60% of the median; and (ii) ratio of car spending over income is above twice the median (2M). In line with the energy poverty literature, the paper defines the transport energy poverty gap as the mean difference between the cost ratio of the household and the cost ratio threshold.

Berry, Jouffe, Coulombel and Guivarch (2016) adapts three fuel poverty measures to transport: the 2M indicator, defined as fuel spending over income, the LIHC indicator with (i) fuel costs per active persons instead of fuel costs; (ii) residual equivalent income for the income threshold, and a composite indicator combining mobility practices, conditions of mobility and financial resources.

Methodologically, Mattioli et al. (2018) and Berry et al. (2016) are the closest to our study since they both adapt the LIHC metric of energy poverty to specific considerations regarding their case study. Moreover, differently from Berry et al. (2016), Mattioli et al. (2018) econometrically estimate the determinants of energy poverty for the UK.

As in the fuel poverty literature, different measures lead to significant differences in the extent of transport poverty. Mattioli et al. (2018) finds that 9% of the UK population is at risk of Car Related Economic Stress (CRES). On the other hand, Sustrans (2012) finds 1.5 million UK households at risk of transport poverty. Berry et al. (2016) finds that 10.5% of French households are car-fuel poor in terms of the 2M indicator, but only 2% when she restricts to poor households. She also finds that 3.3% of French households are car-fuel-poor in terms of the LIHC. Finally, using her composite indicator, she finds that 12% are fuel vulnerable, 7.5% are fuel dependent and 7.8% are fuel poor.

Regarding the socioeconomic determinants, a strand of the literature on fuel poverty that investigates this (Belaïd (2018); Legendre and Ricci (2015); Romero et al. (2018); Belaïd (2022b); Belaïd and Flambard (2023)) find that key factors include household size and composition, employment status, size of home, dwelling and tenure type, insulation type, household head's age, energy used for heating and cooking, and location (urban or rural). Having poor roof insulation, having an individual boiler, and cooking with gas instead of electricity is increasing the chances of being fuel-poor (Legendre and Ricci (2015)). In terms of the estimation method, Belaïd (2018); Romero et al. (2018); Belaïd (2022b) use binary logistic models, Belaïd and Flambard (2023) use a trivariate probit regression model, Legendre and Ricci (2015) uses a C log-log regression model and a mixed logit model. Belaïd (2022b, 2018) also uses a two-step clustering procedure combining Multiple Correspondence Analysis (MCA) and Ascendant Hierarchical Clustering to identify different profiles of domestic-fuel poor in France, Egypt, and Jordan.

In the transport poverty literature, only two papers investigate the determinants of car-fuel poverty, all for the UK case, and only two use an econometric approach based on logistic regressions (Mattioli (2017b); Mattioli et al. (2018)). Mattioli (2017b) finds that having children, being inactive, being a middle-aged tenant with low-to-middle income affect significantly the odds of being FCO. Mattioli et al. (2018) shows that determinants of being in Car-Related Economic Stress (CRES) include having children, having a portion of household members unemployed, and living in rural areas or Northern Ireland.

For France, no econometric study has been done so far but Nicolas et al. (2012) quantifies vulnerable households defined as low income (25% lower than local average) that are very mobile (75% more km driven than local average) and presents, through descriptive statistics, few categories of socio-economic variables that are in general correlated with vulnerability. These include household size, living in peri-urban areas, socio-professional categories (farmers, employees, independent), having several active people, and being multi-motorized.

Study	Transport Poverty Metrics	Determinants
RAC (2012)	10%: transport poor households spend more than 10% of income on running motor vehicles	Not studied
Nicolas et al. (2012)	Burden of daily mobility in **France**: ratio of daily mobility expenditures over equivalent income. Vulnerable households: they have low income (25% lower than local average) and are very mobile (75% more km driven than local average).	Not studied

Sustrans (2012)	Composite risk of transport poverty index (**UK**): 1. Areas of low income, where running a car would put high stress on households' budgets. 2. Areas where people live more than one mile from nearest bus station 3. Number of essential services that would take more than 1 hour to access by walking, cycling and public transport	Not studied
Berry et al. (2016)	For **France** - 2M indicator only for households under the poverty line. - LIHC if 1. Fuel spending per (active) person > median 2. If its residual income per consumption unit < poverty line - Composite indicator combining 1. Mobility practices (high fuel spending, extra travel time & car use restriction) 2. Conditions of mobility (poor spatial matching, no alternative & low vehicle performance or no vehicle) 3. Financial resources (income)	Not studied
Mattioli (2017b)	Forced-Car Ownership (FCO) if household 1. owns at least one car 2. finds difficulties to afford other essential needs such as rent, mortgage, domestic energy and food	Binary logit in **Germany & UK**. Significant factors: having children (in the UK, not Germany), low activity rates, being tenants, living in semi-detached houses (UK) or small blocks of apartments (Germany), adults in the middle age groups, with low-to-middle income.
Mattioli et al. (2018)	Households experiencing CRES[2]: 1. equivalent income after housing and running motor vehicle costs is below 60% of the median (as in LIHC) 2. percentage of income spent on running motor vehicle > twice median (as in 2M) Also measures the *transport-energy poverty gap* : mean difference (in %) between the cost burden ratio of LIHC households and the cost threshold	Binary Logit in **UK**. Significant factors include having children, unemployed members and living in rural areas or Northern Ireland.
Mattioli et al. (2019)	Composite indicator of vulnerability : 1. Exposure (ratio between estimated mean expenditure on motor fuel and median income) 2. Sensitivity (median income of area) 3. Adaptive capacity (sum of estimated travel time to eight key services by walking or public transport	Not studied

Table 2: Transport poverty metrics

[2]Car-Related Economic Stress

3. Methodology and Data

In this section, we first describe the databases used and their interrelations. Then we present the methodologies we have chosen to tackle, First, the way in which car-fuel-poor households could be classified into clusters according to their main characteristics. This will allow us to simulate targeted policies. Then, we present the methodology to empirically estimate car-fuel poverty determinants. Finally, we present the different techniques we have used to evaluate the impact of the gasoline and diesel subsidy imposed in France in 2022 as well as the way we have built the alternative policies that we investigate.

3.1. Database and metrics construction

Bridging the gap between the energy and transport poverty literature, as well as following the EU Energy Poverty Observatory directives, We choose three indicators: an adaptation of Hills (2012)'s LIHC, the 2M indicator, and the M/2 indicator. Due to the limitations in our data, we use the average income of each income decile in both the adapted LIHC and 2M indicators.

- **Low-Income High Transport Costs (LIHTC)**: as in fuel poverty inside the home literature as well as in Berry et al. (2016) and Mattioli et al. (2018), LIHTC households are those who are both above a cost threshold and below a poverty line. In our case (i): car-fuel expenditures are higher than the national median, and (ii) the mean income of household net of car-fuel expenditures is lower than 60% of the median income net of car-fuel expenditures.[3]
- **Twice the median indicator (2M)**: that ratio of car-fuel expenditures over income is above twice the national median.
- **Under-consumption: half the median (M/2)**: car-fuel expenditures are below half the national median.

Despite the suggestions of Hills (2012), we do not consider income after housing costs since, in the Transport Survey we use, we do not observe it (same as Berry et al. (2016)). In contrast to the literature on fuel poverty, we choose not to normalise income with household consumption units or m^2 since, differently from domestic fuel consumption, travel is motivated by individual needs so there is no consensual normalization available (Mattioli, Lucas and Marsden (2017)).

We use the 2018-2019 French National Household Travel Survey, which we refer to as EMP (*Enquête Mobilité des Personnes*)[4]. Aside from information on mobility, the database contains socio-demographic information such as income decile, household size and composition, age and gender of households' head and tenure type (either owner or tenant).[5] We also have information on socio-professional category of households' heads; these socio-professional categories are *farmer*, *self-employed*, which includes craftsmen, shopkeepers, company directors and other self-employed occupations, *executives*, *employees*, *intermediate occupations*, which includes different intermediate jobs between executives and employees in all sectors, and finally, *unemployed that never worked and inactive* households. Finally, we observe information on the municipality of residence. We build a *municipality type* index with three categories: *urban* (high-density areas), *commuting area* (middle and low-density areas with at least 15% of households working in the closest urban center), and *isolated rural* (low-density municipalities, isolated from urban centers). Public transport access is proxied by the distance to the nearest local, regional or national train station. In the survey, we do not observe fuel expenditures but kilometers driven. To compute fuel expenditures, we use 2019 fuel prices from (French Ministry of Finance (2019)). The main drawback of our data is that we only observe households' income deciles and

[3] 60% of national median income is the national poverty line

[4] Berry et al. (2016) defines car-consumption units since it focuses on specific routes (such as home-work) for which she has information for every member of the household.

[5] It was conducted face to face by INSEE (Institut National de la Statistique et des Etudes Economiques) interviewers between May 2018 and April 2019, in 6-month waves of 2 months each, targeting 20,000 households in metropolitan France. The success rate of the survey is 75.6% which represents 13,825 households who made 45,169 daily trips in the 6 weeks prior to the survey. These households belong to dwellings drawn in the annual census surveys. The interview begins with the equipment census, hence the number of cars, vans, light-duty vehicles, motorcycles, and bicycles, that the household owns. For each car, we have information on age, fuel type, and fuel consumption (l/100km). Then, one individual is drawn from the household and is surveyed on his or her local and long-distance travels. Distance travelled is often poorly estimated by individuals. Therefore, the database consolidates this information using additional databases (for e.g., Répertoire Statistique des Véhicules Routiers, built from registration data, to estimate annual kilometers driven). Since travels of one household's individual might not be representative of the entire household, we used the consolidated annual distances travelled of each car own by each household to compute their fuel expenditures.

not their actual income. We thus match the previous database with INSEE income distribution data on the mean income per decile from Fabien Delmas (2020).

3.2. Multiple Correspondence Analysis and Hierarchical Ascendant Clustering

We identify different fuel-poor household profiles using a two steps clustering procedure commonly used in the field of data science. First, we use Multiple Correspondence Analysis (MCA) to represent the categorical variables in our database in a bi-dimensional space where the proximity of the points in the space indicates the degree of association between the corresponding categories. More details on MCA can be found in Appendix 6.3.1.

The second step is a Hierarchical Ascendant Clustering (HAC). HAC explores the existence of natural groups and their underlying structure within the car-fuel poor household categories. Results of HAC are presented in a dendrogram[6], in which we can visualize the different population clusters. We choose the number of clusters by applying the *Elbow approach*. This method relies on computing the Within-Cluster-Sum of Squared Errors (WSS) for various numbers of clusters (k) and determining the value of k at which the decrease in WSS begins to diminish (forming an *elbow* as shown on Figure 6). Details on the methods can be found in Appendix 6.3.2. Given the desired level of heterogeneity within and between clusters, this technique allows us to draw 3 different profiles per car-fuel poor definition.

3.3. Econometric method

We then study the odds of being LIHTC, 2M and M/2. As in the energy poverty literature (Belaïd (2018); Legendre and Ricci (2015); Belaïd (2022b); Romero et al. (2018)) and in the two previous papers on transport poverty (Mattioli (2017b); Mattioli et al. (2018)) we estimate three binary logistic regressions as in McFadden (1974b), one for each of the car-fuel poverty definition chosen. Each logistic model can be described by:

$$log(\frac{p}{1-p}) = \beta_0 + \sum_i \beta_i x_i \tag{1}$$

with p the probability of being either LIHTC, 2M or M/2. β_0 is the intercept and β_i are coefficients for predictors x_i. Each dependent variable is binary and can be written Y_n, with

$$Y_n = \begin{cases} 1 & \text{if household n is LIHTC, 2M or M/2 respectively} \\ 0 & \text{otherwise} \end{cases}$$

In line with the energy and transport poverty literature, we consider the following explanatory variables: income decile (when not used to derive the indicator as for the LIHTC case), household composition (number of adults, number of children), motorization (number of cars), socio-professional categories, tenure type, location and distance to public transport. We use variables presented earlier in section 3.1, which we describe in details in Appendix 6.2.

3.4. Subsidy simulation

By analyzing the observed annual mileage of individuals in France, we estimate the impact of 2022´s fossil-fuel inflation on household´s purchasing power[7]. First, we compute the additional fuel expenditures as a consequence of inflation in 2022 across different groups (income decile, car-fuel poverty groups, and within-group clusters). We find the effect of rising fuel price is regressive or has a greater impact on middle-income households, depending on whether the analysis includes all households or only those with motor vehicles. This way, we highlight the importance of defining a consensual indicator of car-fuel poverty.

Next, we examine the impact of the 2022 fuel subsidy program that has been applied in three different phases. First, from April 1^{st} to August 31^{st} 2022, the amount was €18 cents per liter of gasoline. It reached €30 cents per liter from September 1^{st} until November 15^{th}, and it has decreased to €10 cents per liter from November 15th to December 31^{st}. Through survey data, we compute the average subsidy per income group and we evaluate the cost for public finances.

Then, we compare the 2022 generalized subsidy that was in place with alternative subsidies targeting the car-fuel poor according to the alternative metrics and belonging to the different clusters identified.

We also compare these outcomes with the government's 2023 €100 fuel vouchers that has replaced the subsidy program as from January 1^{st} 2023. The government announced 11 million eligible households[8]. The measure was

[6] A dendrogram is a visual depiction of a tree structure. This graphical representation finds wide application in various contexts, particularly in hierarchical clustering, where it portrays the organization of clusters resulting from the analysis

[7] Since we only observe km travelled for 2019, we assume that people did not change their behavior after price increase.

[8] Households eligible to the €100 voucher are those who belong to the poorest five income deciles and use their cars to go to work.

meant to be equivalent to a subsidy of €10 cents per liter for an average individual driving 12,200km[9] with a 6.5l/100km car per year. To get the voucher, individuals must register on a government website and provide their income, vehicle registration number, and a sworn statement that the vehicle is used to go to work. On February 2023, 5.17 million individuals had applied to the program and 3.3 million had received the voucher (French Ministry of Finance (2023)).

4. Results and discussion

Herein we describe the main results of the alternative approaches used to identify and quantify car-fuel-poor households, what determines this situation, and what could be done to compensate for their vulnerability to high gasoline prices.

4.1. Car-fuel poverty extent and profiles

Applying the three definitions to households in metropolitan France, we find that 3.2% are car-fuel poor according to the LIHTC definition, 20% have high exposure, spending a disproportionate share of income on car-fuels (2M) and 31% are under-consuming (M/2) (including people with no car-fuel expenditures). Similarly to Legendre and Ricci (2015) for the case of fuel poverty inside the home and to the descriptive study of Berry et al. (2016), the extent of poverty is very different depending on the definition used. Our result for the LIHTC indicator is close to Berry et al. (2016) that uses the 2008 national French household travel survey and finds that 3.3% of French households are fuel poor under its adaptation of the LIHC indicator. On the other hand, it differs from Mattioli et al. (2018) that find that 9% of the UK population is fuel poor under their adaptation of the LIHC indicator (see Table 2). Moreover, our results show that only 18% of households in the 2M group are also LIHTC, hence fall below the poverty line after paying for car-fuels.

The two step clustering method allows us to identify nine main car-fuel poor types. Detailed statistics on the composition of clusters as well as the biplots and dendrograms for each poverty indicator can be found in Appendix 6.3. For the LIHTC definition, the three clusters identified are represented in Figure 8 of the Appendix. Cluster 1 represents about half of LIHTC households. They are single individuals, most under 45 (most of them women under 25), with either one child or no children, belonging to the poorest 20% and renting a home in a rather dense location (urban or commuting area). Despite good access to public transport, (53% leave close to a train station), the vast majority owns a car and uses it intensively. These individuals are mostly employees, intermediate, unemployed, or inactive. Cluster 2 (18% of LIHTC) contains single individuals, with a majority of men, retired, over 60, owners, living alone in isolated rural or commuting areas, owning one car and belonging to the poorest 20%. In cluster 3 (31% of LIHTC), we have an over-representation of large families (two adults and two or more children) compared to the rest of the LIHTC population. They belong to the 30% poorest households. Typical household heads of cluster 3 are men, self-employed, farmers or employees. They are homeowners living in isolated rural and commuting areas with poor access to public transport. They own two cars.

For 2M households, we obtain three clusters. Households in cluster 1 (45% of 2M) are multi-motorised large families belonging to the fifth income decile or above, with the majority belonging to the richest 30% of total population. There is a large majority of home-owners living in commuting areas. Household heads are men aged between 30 and 60 , with an over-representation of intermediate, self-employed, farmers, employees. This is in line with the literature (Hills (2012); Legendre and Ricci (2015); Mattioli et al. (2017)) that shows that the twice-the-median (or 10%) indicator does not exclude rich households that can cope with an increase in fuel prices. Cluster 2 (16% of 2M) has households with income below national median. They have retired men over 60 as household heads. They live mostly in commuting areas with poor access to public transport. They have either one or two cars. Cluster 3 (39% of 2M) contains low-income single adult households (40% poor) with either one child or no children and one car. There is a majority of employees and an over-representation of inactive but also intermediate and self-employed. They rent their home. Most are either in commuting or isolated rural areas with poor access to public transport. 2M clusters are represented on Figure 1.

Finally, M/2 households can be partitioned into three clusters. Cluster 1 (47% of M/2) is composed of single individuals, mostly women, owners, retired, over 70 years old. They belong to the poorest 40%, own either one or no cars and live in commuting or urban areas with relatively good access to public transport. Cluster 2 represent only 15% of M/2 households. It is composed of households who might be volontarily consuming less, by choice, rather than because of budgetary constraints. In fact, they have income above the median, live in dense urban areas (93%) with very good access to public transport. They are families with children; they are tenants and own either one or no cars.

[9] Which corresponds to the national average

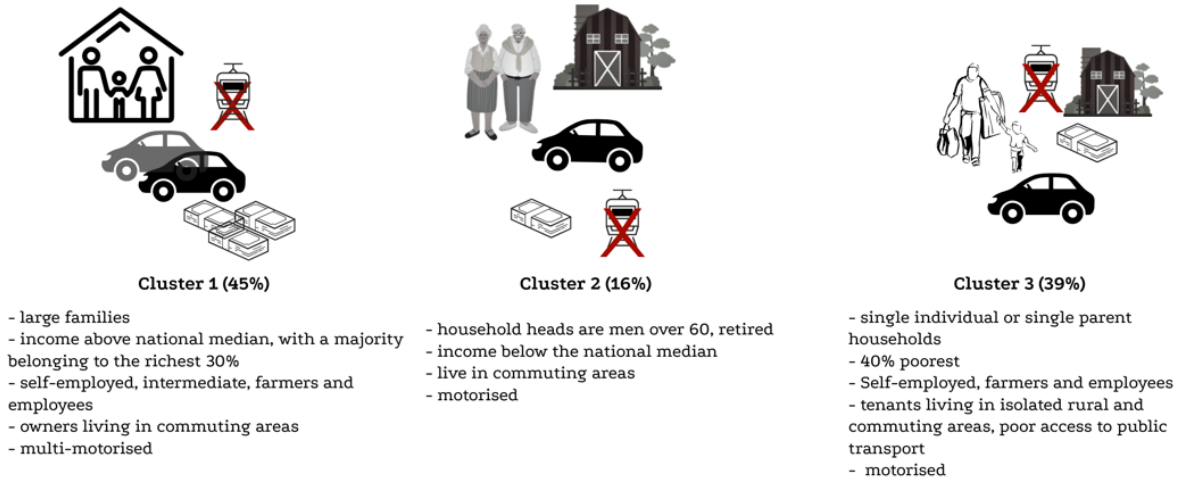

Figure 1: 3 clusters inside the 2M household population

Typical household heads are men, between 30 and 50, working as executives or employees. Cluster 3 (38% of M/2) has a majority of single-parent households with either one child or no children. They have low-income (30% poorest). The large majority does not have a car and are tenants living in urban areas with good access to public transport. They are less than 60 years old and are mostly employees. In section 4.3, we describe specific policies for each cluster according to the alternative definitions, which in turn capture very different vulnerability dimensions.

4.2. Determinants of car-fuel poverty

Now that we have identified different car-fuel poor profiles, we study the determinants of falling into car-fuel poverty. In Table 3, we present logistic regression results for the three alternative definitions. We display both regression coefficients, which are in the log-odds scale, as well as odds ratio (OR). We compare the odds of a social group to a reference category: no children for the number of children, one adult for the number of adults, one car for motorization, pensioners for socio-professional categories, owners for tenure type, urban center for municipality type, below 2 km for distance to train station and the poorest income decile (D1) for income decile membership.

	LIHTC		**2M**		**M/2**	
	Coef.	OR	Coef.	OR	Coef.	OR
1 child	-0.52^{**}	0.60^{**}	-0.09	0.91	-0.09	0.91
	(0.16)	(0.16)	(0.07)	(0.07)	(0.10)	(0.10)
2 children	-0.57^{**}	0.57^{**}	-0.12	0.89	-0.03	0.97
	(0.18)	(0.18)	(0.07)	(0.07)	(0.11)	(0.11)
3+ children	-0.33	0.72	0.02	1.02	-0.01	0.99
	(0.23)	(0.23)	(0.09)	(0.09)	(0.13)	(0.13)
2 adults	-1.74^{***}	0.18^{***}				
	(0.14)	(0.14)				
3+ adults	-2.02^{**}	0.13^{**}				
	(0.73)	(0.73)				
2 cars	0.53^{***}	1.71^{***}	0.59^{***}	1.80^{***}	-2.08^{***}	0.12^{***}
	(0.14)	(0.14)	(0.06)	(0.06)	(0.09)	(0.09)
3+ cars	0.68^{**}	1.97^{**}	1.33^{***}	3.77^{***}	-3.22^{***}	0.04^{***}
	(0.23)	(0.23)	(0.08)	(0.08)	(0.32)	(0.32)
Farmer	1.71^{***}	5.50^{***}	1.30^{***}	3.67^{***}	-0.95^{***}	0.39^{***}
	(0.26)	(0.26)	(0.16)	(0.16)	(0.28)	(0.28)

	Coef.	OR	Coef.	OR	Coef.	OR
Self-employed	0.99***	2.69***	0.98***	2.67***	−0.90***	0.41***
	(0.22)	(0.22)	(0.10)	(0.10)	(0.15)	(0.15)
Executives	−0.68*	0.51*	−0.18	0.84	−0.80***	0.45***
	(0.31)	(0.31)	(0.10)	(0.10)	(0.12)	(0.12)
Intermediate	0.09	1.10	0.67***	1.95***	−1.23***	0.29***
	(0.19)	(0.19)	(0.08)	(0.08)	(0.10)	(0.10)
Employees	0.79***	2.20***	0.97***	2.65***	−1.02***	0.36***
	(0.14)	(0.14)	(0.07)	(0.07)	(0.08)	(0.08)
Unemployed and inactive	1.32***	3.74***	1.15***	3.15***	−0.26	0.77
	(0.21)	(0.21)	(0.14)	(0.14)	(0.14)	(0.14)
Tenant	0.94***	2.55***	0.65***	1.91***	−0.01	0.99
	(0.12)	(0.12)	(0.06)	(0.06)	(0.07)	(0.07)
Commuting area	0.19	1.21	0.45***	1.57***	−0.40***	0.67***
	(0.13)	(0.13)	(0.06)	(0.06)	(0.07)	(0.07)
Isolated rural	0.72***	2.05***	0.72***	2.06***	−0.51***	0.60***
	(0.18)	(0.18)	(0.09)	(0.09)	(0.13)	(0.13)
2-5km	0.33*	1.39*	0.22**	1.24**	−0.27***	0.77***
	(0.14)	(0.14)	(0.07)	(0.07)	(0.07)	(0.07)
5-10km	0.30	1.35	0.39***	1.48***	−0.27**	0.76**
	(0.17)	(0.17)	(0.08)	(0.08)	(0.10)	(0.10)
> 10km	0.73***	2.08***	0.71***	2.03***	−0.33***	0.72***
	(0.16)	(0.16)	(0.07)	(0.07)	(0.09)	(0.09)
D2					−0.17	0.84
					(0.12)	(0.12)
D3					−0.26*	0.77*
					(0.12)	(0.12)
D4					−0.37**	0.69**
					(0.12)	(0.12)
D5					−0.65***	0.52***
					(0.12)	(0.12)
D6					−0.61***	0.54***
					(0.13)	(0.13)
D7					−0.95***	0.39***
					(0.14)	(0.14)
D8					−0.91***	0.40***
					(0.14)	(0.14)
D9					−1.28***	0.28***
					(0.16)	(0.16)
D10					−1.23***	0.29***
					(0.17)	(0.17)
AIC	3205.05	3205.05	11281.44	11281.44	8128.19	8128.19
BIC	3366.81	3366.81	11428.50	11428.50	8341.42	8341.42
Log Likelihood	−1580.53	−1580.53	−5620.72	−5620.72	−4035.10	−4035.10
Deviance	3161.05	3161.05	11241.44	11241.44	8070.19	8070.19
Num. obs.	11530	11530	11530	11530	11530	11530
McFadden Pseudo R-squared	0.15		0.11		0.24	
Nagelkerke	0.18		0.17		0.33	

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3: Regression results.

Having children has a significant and negative impact at the 5% level for LIHTC households but not for 2M and M/2 households. In fact, Table 3 shows that the odds of being LIHTC are 1.7 to 1.8 times greater for households that do not have children, compared to households with at least one child. This is at odds with Mattioli et al. (2018) that finds a positive relationship between the number of children and the probability to be in Car Related Economic Stress (CRES). Our results also show that the odds of being car-fuel poor under the LIHTC for a household with one adult are 6 times the odds of having respectively 2, 3 and more adults. In the energy poverty litterature household composition has no consensual effect: Belaïd (2018), finds that larger families have more chances to be fuel-poor while Legendre and Ricci (2015) find that single-person households are more likely to be vulnerable.

The likelihood of being LIHTC and 2M increases with the number of cars while the likelihood of underconsuming (M/2) decreases with the number of cars. In fact, households with two cars have 1.71 times the odds of being LIHTC than households owning only one car (respectively 1.8 times for 2M). On the other hand, households owning one car have 8.3 times the odds of underconsuming compared to households with two cars. We cannot compare our results with Mattioli et al. (2018) since they do not use the number of cars as a predictor.

Households headed by farmers, self-employed, employees, unemployed or inactive tend to be more car-fuel poor in terms of LIHTC and 2M definitions compared to pensioners, which is our reference category. In facts, the odds of being vulnerable for farmers, self-employed, employees, unemployed and other inactive are respectively 5.5, 2.7, 2.2 and 3.8 times the odds of pensioners for LIHTC households (respectively 3.7, 2.7, 2.7 and 3.2 for 2M). In comparison, Mattioli et al. (2018) do not have detailed information on household socio-professional categories but find that having unemployed members in the households increases the probability of being vulnerable, which is consistent with our findings. Our results are also consistent with the fuel poverty at home literature. Romero et al. (2018); Belaïd (2018, 2022b) find that having household heads with low education or job instability increases the probability of fuel poverty. Finally, the odds of under-consuming are higher for pensioners than for most of the other categories. Executives, self-employed and farmers have the lowest chances to under-consume, as compared to pensioners.

We find that tenants have higher chances than owners to be LIHTC and 2M while tenure type is not significant for M/2 households. This is in line with the energy poverty literature (Belaïd (2018); Legendre and Ricci (2015)) that find that tenants are about twice as likely to experience fuel poverty than owners.

Our results show that the odds of being car-fuel poor for households living in isolated rural areas are 2 times those of households living in dense urban areas. Living in commuting areas has no significant impact on the probability of being LIHTC but significantly increases the probability of being 2M and decreases the probability of being M/2 at the 1% level.

Results suggest that accessibility to public transport has a significant impact on the three outcome variables. We find that living more than 2 km away from a train station increases the probability of being 2M and decreases the probability to under-consume. The odds of being LIHTC when living between 2 and 5 km away from a station are 1.4 times the odds (respectively living 10 km away has 2.1 the odds) of a household living close to a station. Distance to train station has a negative effect on under-consumption, increasing in absolute with distance.

Finally, the odds of under-consuming belonging to D1 are 1.3 times the odds of under-consuming belonging to D3 and 3.3 times the odds of rich households (D10).

Logistic regression results in Table 3 shows that most of our independent variables are significant at the 5% or 1% level. Our predictors present a low correlation, with Variance Inflation Factors less than 2. To avoid collinearity with income deciles, we exclude the number of adults from the predictors for the M/2 and 2M model[10]. Since we used income deciles in the LIHTC and 2M calculations, we remove it from the predictors for these outcome variables, to avoid endogeneity issues.

To assess the predictive capacity of the econometric estimations, we use the traditional R-squared indicator[11] McFadden (1974a) and the Nagelkerke's R-squared that is similar to the previous but adjusted to the sample size N[12].

The McFadden pseudo R-squared is above 0.2 for the M/2 model which suggests a good predictive capacity. It is close but less than 0.2 for the LIHTC and 2M models (respectively 0.15 and 0.11), which is likely caused by the omission of income to avoid endogeneity. Low pseudo R-squared are common in this type of models, as in Mattioli et al. (2018) that obtain a McFadden R-squared of 0.062. The same analysis applies to Nagelkerke's pseudo R-squared, which is considered a more complete measure of predictive capacity than McFadden's.

[10]The number of adults is correlated with income deciles since income is not equivalised as in the fuel consumption in the home literature.

[11]McFadden Pseudo R-squared is $R^2 = 1 - \frac{\text{Log-likelihood full model}}{\text{Log-likelihood null model}}$

[12]The Nagelkerke's R-squared is defined as: $R^2 = \frac{1-(\text{Log-likelihood null model}/\text{Log-likelihood full model})^{2/N}}{1-(\text{Log-likelihood null model})^{2/N}}$

In terms of goodness of fit, the Hosmer and Lemeshow test (Hosmer Jr, Lemeshow and Sturdivant (2013)) compares the observed and predicted probabilities of the outcome variable for different (here, 10) groups in the sample. It tests if there is a significant difference between them. A non-significant test validates the model. For our three models, the Hosmer and Lemeshow tests result in non-significant differences, which suggest a good fit of the models. Detailed tables are available in Appendix 6.4.

4.3. Subsidy simulations

Understanding what it means to be car-fuel poor and which are its key determinants has several implications for energy and climate policies.

In Figure 2 (left) we show the % income loss due to the gasoline and diesel price spikes in 2022 if prices had not been subsidized. We observe that households in D1 that own a car lost: (i) more than 1% more than households in D2; (ii) double than the middle class; and, (iii) four times than the loss by D10.

Car-fuel poverty group	Average Loss (€/household)
LIHTC	581
2M	745
2M-C2	552
2M-C3	510
low-income workers with a car	352

Table 4: Average loss due to inflation for each targeted group (€/households).

In Table 4 we have calculated the average loss provoqued by the inflation (without subsidy) in euros per vulnerable group identified by our estimation and clustering as well as for the targeting group chosen by the government for the €100 cheque implemented in 2023. We observe that the yearly loss due to inflation are all above €500 for the vulnerable groups identified.

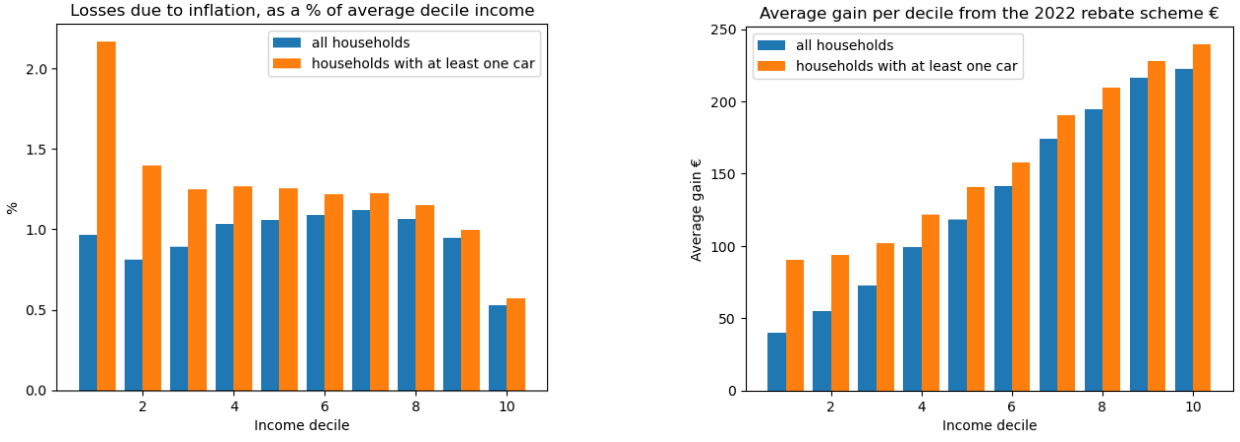

Figure 2: Additional budget share due to 2022 inflation (left) and subsidy perceived in 2022 (right) for different income groups

In Figure 2 (right), we compute the average subsidy per household belonging to each income group resulting from the 2022 generalized policy. We find that the households in D10 have gain in average 2.5 more than the poorest decile if all households are considered, and 1.6 times if we only account for households that own at least one car[13]. We observe that the policy has been extremely regressive.

[13]This result is in line with the work of the Conseil d'Analyse Economique that is part of the Ministry of Economics that investigated the effects of the first phase of the rebate (c€18/l for 4 months) using transaction data from *Crédit Mutuel* bank and finds that the measure benefited the 10^{th} decile households twice as much as households in the 1^{st} decile.

Then, in Table 5 we show our calculations for the total cost of the subsidy over the 9-month period finding that it has cost €3.9 billions[14]. We also show that, if the government had chosen to target only the most vulnerable (LIHTC) with their 2022 subsidy, they could have saved about €3.7 billion since the measure would have cost only 234 million euros. On the other hand, if the rebate had been targeted towards 2M households, it would have cost the state €1.6 billion. Let us now consider alternative policies to perfectly compensate for the loss due to inflation in different vulnerability groups. We find in the second part of the Table that if they had completely compensated the LIHTC, they could have saved €3.3 billion. We also find that fully compensating the 2M households would have led to additional expenses compared to the 2022 scheme, since this category, if not complemented with income restrictions, includes households that are not poor. Therefore, distinguishing among the 2M households between non-poor households that can cope with an increase in fuel prices and households that cannot is key. This is how clustering analysis can help. In table 5, we suggest compensating fully LIHTC households without distinguishing between clusters since all LIHTC experience a severe form of vulnerability. On the other hand, we compute the cost of targeting specifically the vulnerable households within the 2M group (clusters 2 and 3).

Scheme	**Target**	**Cost** (M€)	**Households**
2022 subsidy	No target	3,900	All
	LIHTC	234	1M
	all 2M	1,600	5.2M
	2M-C2	174	0.8M
	2M-C3	419	2.1M
Compensation	LIHTC	592	1M
	all 2M	4,200	5.2M
	2M-C2	440	0.8M
	2M-C3	1,100	2.1M
€100 voucher	Low-income workers with a car	640	6.4M individuals

Table 5: Actual and simulated policy schemes with their targets and costs. We compute the number of eligible households for each case.

We also compute the number of eligible households for the 2023 voucher of €100 per year as well as the potential cost of this policy. We find that, since it only applies to active individuals within households belonging to the five poorest income deciles and owning at least one car, we identify 6.4M households elegible, which yields a cost of €640 millions[15]. Although this measure is better targeted than the 2022 generalized subsidy, we have shown in section 4 that 16% of 2M and 18% of LIHTC are retired, and that unemployed and inactive heads are also heavily represented among car-fuel poor households. Moreover, the policy does not target specifically households with high fuel expenditures. Finally, we observe that 4 are much higher than 100 (or 200 for couples) euros a year meaning that the subsidy´s compensation is low compared to estimated losses for the most vulnerable groups.

All in all, our results show that both policies applied in France, the 2022 generalized subsidy and the €100 voucher that replaced it, have not been well designed. Both schemes have failed to sufficiently compensate vulnerable households for their loss in purchase power. Moreover, the inclusion error in the 2022 subsidy is very important, spending taxpayers money to subsidize rich households. The exclusion error in the €100 voucher is the main problem of this scheme, leaving out important portions of vulnerable population. The results herein as well as the methodology used suggest alternative ways to correctly target and compensate population vulnerable to car-fuel poverty.

5. Conclusion

In this paper, we bridge the gap between the energy and transport poverty literature by adapting metrics extensively used to define fuel poverty inside the home to transport. We choose to focus on three metrics: (i) Hills (2012)'s Low-Income High Cost adapted to consider the fact that transport is an individual matter; (ii) an indicator that identifies

[14]The government announced a budget of €7.6 billion for the policy but here we only consider private cars. Our results for private cars' fuel consumption are in line with national accounts for individual cars.

[15]Note that we only account for metropolitan territory as does the national households travel survey

households that spend a disproportionate share of their income on car fuels (2M); and (iii) a measure of under-consumption (M/2) identifying both households that are restricting their fuel consumption for affordability reasons and households that benefit from good access to public transportation.

We study the extent of transport poverty in the French population and identify three profiles of car-fuel poor households for each metric, which can help design targeted policies that are as we simulate herein. We find that overall, a small percentage of the French population is car-fuel poor (3.2%, 1 million) while a larger part is very fuel dependent but not necessarily poor (20% for which 3 million are in cluster 2 and 3 being car-fuel poor). Then, we investigate key socio-economic factors that increase the probability of being car-fuel poor. Our findings suggest that households at risk of poverty are single, usually low-income, farmers, unemployed or inactive heads, tenants living in isolated rural or commuting areas with very low access to public transport.

Finally, we use our results to suggest alternative targeting schemes for fossil-fuel subsidies to compensate households for the energy price spikes observed after the Ukrainian crisis. We show that targeted policies could have led to significant savings, while compensating completely the car-fuel poor. This result is at odds with the proposal presented by the French government in September 2023 to amend competition law to allow fuel distributors to sell at a loss.

Our study has focused on affordability. Yet, transport poverty cannot be reduced to a single dimension. To better account for the multidimensional aspects of transport poverty, some papers have developed composite indicators (Berry et al. (2016); Mattioli (2017a)). A new strand of the literature also investigates energy and transport poverty jointly, to account for the double vulnerability that certain households face (Lowans et al. (2021); Simcock, Jenkins, Lacey-Barnacle, Martiskainen, Mattioli and Hopkins (2021). This could be a topic for future research.

Disclaimer

The authors affirm that they do not possess any financial conflicts of interest or personal associations that might have seemed to impact the findings presented in this manuscript. The authors bear full responsibility for the results and conclusions of this study.

Acknowledgments

Authors acknowledge the work of two anonymous referees.

References

Adger, W.N., 2006. Vulnerability. Global Environmental Change 16, 268–281. URL: https://www.sciencedirect.com/science/article/pii/S0959378006000422, doi:https://doi.org/10.1016/j.gloenvcha.2006.02.006.

Belaïd, F., 2018. Exposure and risk to fuel poverty in france: Examining the extent of the fuel precariousness and its salient determinants. Energy Policy 114, 189–200. URL: https://www.sciencedirect.com/science/article/pii/S0301421517308200, doi:https://doi.org/10.1016/j.enpol.2017.12.005.

Belaïd, F., 2022a. Implications of poorly designed climate policy on energy poverty: Global reflections on the current surge in energy prices. Energy Research Social Science 92, 102790. URL: https://www.sciencedirect.com/science/article/pii/S2214629622002936, doi:https://doi.org/10.1016/j.erss.2022.102790.

Belaïd, F., 2022b. Mapping and understanding the drivers of fuel poverty in emerging economies: The case of egypt and jordan. Energy Policy 162, 112775. URL: https://www.sciencedirect.com/science/article/pii/S0301421521006418, doi:https://doi.org/10.1016/j.enpol.2021.112775.

Belaïd, F., Flambard, V., 2023. Impacts of income poverty and high housing costs on fuel poverty in egypt: An empirical modeling approach. Energy Policy 175, 113450. URL: https://www.sciencedirect.com/science/article/pii/S0301421523000356, doi:https://doi.org/10.1016/j.enpol.2023.113450.

Berry, A., Jouffe, Y., Coulombel, N., Guivarch, C., 2016. Investigating fuel poverty in the transport sector: Toward a composite indicator of vulnerability. Energy Research & Social Science 18, 7–20. URL: https://linkinghub.elsevier.com/retrieve/pii/S2214629616300123, doi:10.1016/j.erss.2016.02.001.

Boardman, B., 1991. Fuel poverty: from cold homes to affordable warmth. Pinter Pub Limited.

Fabien Delmas, J.G., 2020. En 2018, les inégalités de niveau de vie augmentent. Insee Première n° 1813 .

French Ministry of Finance, 2019. Plateforme ouverte des données publiques françaises. data.gouv.fr.

French Ministry of Finance, 2023. Communiqué de presse du ministère de l'économie et des finances. URL: https://www.prefectures-regions.gouv.fr/ile-de-france/content/download/103418/656109/file/2023%2002%2013%20CP%20Prolongement%20indemnit%C3%A9%20carburant.pdf.

Hills, J., 2012. Final report of the Hills Independent Fuel Poverty Review: Getting the Measure of Fuel Poverty. CASE Reports casereport72. Centre for Analysis of Social Exclusion, LSE. URL: https://ideas.repec.org/p/cep/sticar/casereport72.html.

Hosmer Jr, D.W., Lemeshow, S., Sturdivant, R.X., 2013. Applied logistic regression. volume 398. John Wiley & Sons.

INSEE, 2020. Zonage d'Etude. URL: https://www.insee.fr/fr/information/2114631.

Legendre, B., Ricci, O., 2015. Measuring fuel poverty in france: Which households are the most fuel vulnerable? Energy Economics 49, 620–628. URL: https://www.sciencedirect.com/science/article/pii/S0140988315000390, doi:https://doi.org/10.1016/j.eneco.2015.01.022.

Lesueur, T., 2020. La précarité énergétique en 2021 : une hausse limitée par le chèque énergie URL: https://www.ecologie.gouv.fr/sites/default/files/thema_essentiel_25_precarite_energetique_2021_mars2023.pdf.

Lowans, C., Furszyfer Del Rio, D., Sovacool, B.K., Rooney, D., Foley, A.M., 2021. What is the state of the art in energy and transport poverty metrics? A critical and comprehensive review. Energy Economics 101, 105360. URL: https://linkinghub.elsevier.com/retrieve/pii/S0140988321002668, doi:10.1016/j.eneco.2021.105360.

Lucas, K., 2012. Transport and social exclusion: Where are we now? Transport Policy 20, 105–113. URL: https://www.sciencedirect.com/science/article/pii/S0967070X12000145, doi:https://doi.org/10.1016/j.tranpol.2012.01.013. uRBAN TRANSPORT INITIATIVES.

Lucas, K., Mattioli, G., Verlinghieri, E., Guzman, A., 2016. Transport poverty and its adverse social consequences. Proceedings of the Institution of Civil Engineers - Transport 169, 353–365. URL: https://www.icevirtuallibrary.com/doi/10.1680/jtran.15.00073, doi:10.1680/jtran.15.00073.

Mattioli, D.G., 2017a. Developing an index of vulnerability to motor fuel price increases in England , 12.

Mattioli, G., 2017b. 'Forced Car Ownership' in the UK and Germany: Socio-Spatial Patterns and Potential Economic Stress Impacts. Social Inclusion 5, 147–160. URL: https://www.cogitatiopress.com/socialinclusion/article/view/1081, doi:10.17645/si.v5i4.1081.

Mattioli, G., Lucas, K., Marsden, G., 2017. Transport poverty and fuel poverty in the uk: From analogy to comparison. Transport Policy 59, 93–105. doi:10.1016/j.tranpol.2017.07.007.

Mattioli, G., Philips, I., Anable, J., Chatterton, T., 2019. Vulnerability to motor fuel price increases: Socio-spatial patterns in England. Journal of Transport Geography 78, 98–114. URL: https://linkinghub.elsevier.com/retrieve/pii/S0966692318308767, doi:10.1016/j.jtrangeo.2019.05.009.

Mattioli, G., Wadud, Z., Lucas, K., 2018. Vulnerability to fuel price increases in the uk: A household level analysis. Transportation Research Part A Policy and Practice 113. doi:10.1016/j.tra.2018.04.002.

McFadden, D., 1974a. Conditional logit analysis of qualitative choice behavior , 105–142.

McFadden, D., 1974b. The measurement of urban travel demand. Journal of Public Economics 3, 303–328. URL: https://www.sciencedirect.com/science/article/pii/0047272774900036, doi:https://doi.org/10.1016/0047-2727(74)90003-6.

Moore, R., 2012. Definitions of fuel poverty: Implications for policy. Energy Policy 49, 19–26. URL: https://www.sciencedirect.com/science/article/pii/S0301421512000833, doi:https://doi.org/10.1016/j.enpol.2012.01.057. special Section: Fuel Poverty Comes of Age: Commemorating 21 Years of Research and Policy.

Nicolas, J.P., Vanco, F., Verry, D., 2012. Mobilité quotidienne et vulnérabilité des ménages:. Revue d'Économie Régionale & Urbaine février, 19–44. URL: https://www.cairn.info/revue-d-economie-regionale-et-urbaine-2012-1-page-19.htm?ref=doi, doi:10.3917/reru.121.0019.

OECD, 2016. Terms of reference oecd project on the distribution of household incomes URL: https://www.oecd.org/statistics/data-collection/Income%20distribution_guidelines.pdf.

Oxley, S., 2023. Fuel poverty methodology handbook (low income low energy efficiency) URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1138926/fuel-poverty-methodology-handbook-lilee-2023.pdf.

RAC, 2012. URL: https://www.racfoundation.org/media-centre/transport-poverty.

Romero, J.C., Linares, P., López, X., 2018. The policy implications of energy poverty indicators. Energy Policy 115, 98–108. URL: https://www.sciencedirect.com/science/article/pii/S0301421517308789, doi:https://doi.org/10.1016/j.enpol.2017.12.054.

Simcock, N., Jenkins, K.E., Lacey-Barnacle, M., Martiskainen, M., Mattioli, G., Hopkins, D., 2021. Identifying double energy vulnerability: A systematic and narrative review of groups at-risk of energy and transport poverty in the global north. Energy Research Social Science 82, 102351. URL: https://www.sciencedirect.com/science/article/pii/S2214629621004424, doi:https://doi.org/10.1016/j.erss.2021.102351.

Sustrans, 2012. Locked out, transport poverty in england. URL: https://www.sustrans.org.uk/media/3706/transport-poverty-england-2012.pdf.

Thema, J., Vondung, F., 2020. (2020) epov indicator dashboard: Methodology guidebook. wuppertal institut für klima, umwelt, energie gmbh guidebook. URL: https://energy-poverty.ec.europa.eu/system/files/2021-09/epov_methodology_guidebook_1.pdf.

6. Appendix

The appendix contains additional elements on the energy poverty literature, metrics and determinants in section 6.1, a detailed description of the predictors used in the multivariate analysis and the logistic regressions in section 6.2, an in depth description of the two-step clustering method that we use in this paper as well as the results for all car-fuel poor groups and the descriptive tables for the clusters' population, in section 6.3, and finally, robustness tests for the three logistic regression models in section 6.4.

6.1. Literature Review on Fuel poverty definitions and metrics

In developed economies, concerns about fuel poverty arose following the oil crises of the 1970s. Since then, defining and measuring fuel poverty has been at the heart of a rich and growing literature (Boardman (1991); Moore (2012); Belaïd (2018); Romero et al. (2018)). In this section, we propose a detailed review of the fuel poverty literature. We summarize domestic fuel poverty measures and determinants in Table 6. As mentioned in section 2, the first measure that both infused UK policies and the academic literature was that of Brenda Boardman (Boardman (1991)). Using a 1988 consumer expenditure survey, she identifies the expenditure in fuel inside the home and considers households as fuel-poor if their fuel budget share is above twice the population´s median (2M). In her sample, the population´s median budget share was 5% so her definition was generalized as the *10% indicator*.

In the studies that afterwards have used this definition, fuel expenditures can be actual household expenditures or required fuel expenditures, which corrects for self-rationing. Similarly, income can be gross or equalised in consumption units, full income or income net of housing costs, or approximated by total expenditures. These methodological choices may be driven by data availability and have significant impacts on the composition of fuel-poor households (Moore (2012)).

Boardman (1991)´s definition became the UK official measure of energy poverty from 2001 to 2012 and is still used at the European level by the Energy poverty observatory (Thema and Vondung (2020)), either in absolute terms (the 10% indicator) or in relative terms (the 2M indicator). Other expenditure based indicators are used at the EU level, such as the "half the median" (M/2) indicator that captures under-consumption of fuel inside the home relative to the national median.

Over the years, Boardman's 10% indicator has faced criticism. Hills (2012) and Romero et al. (2018) argued that the indicator is too sensitive to price increases, that the 10% threshold is arbitrary and that the indicator does not exclude rich households. In fact, spending more than 10% or income on fuels inside the home does not necessarily make a household at risk of falling into poverty. Hills (2012) states that the "definition can encompass households that clearly are not poor", which has been confirmed (for e.g. Legendre and Ricci (2015)). Hills (2012) then suggested two alternative indicators: the *after fuel cost poverty* indicator, which targets households that have residual income after fuel cost inside the home below the official poverty line (60% of national median income) and the Low-Income High Cost (LIHC) indicator, a dual indicator that identify households that are above a fuel cost threshold and below an income threshold. In Hills' LIHC metric, fuel poor households have (1) required fuel costs that were above the national median and (2) "were they to spend that amount, they would be left with a residual income below the official poverty line".[16]. With the LIHC indicator comes the measure of the depth of fuel poverty through the 'energy poverty gap', i.e. the additional income required to get out of poverty. Other indicators exist; for example Moore (2012) suggests using a budget standard approach, with fuel poor households having fuel costs inside the home above income net of housing and minimum living costs.

In 2013, the UK chose the LIHC to replace the 10% indicator in their revised energy poverty strategy (Oxley (2023)). In the academic world, the LIHC indicator has been widely used, criticized and improved over the years. For e.g., Romero et al. (2018) considers the poverty threshold as 60% of median income net of average fuel costs. Belaïd (2018) changes units of energy costs from € to €/m^2 to better account for low income households living in small dwellings. Finally, Belaïd (2022a) considers the Low-Income Low Energy Efficiency (LILEE) indicator as an alternative to the LIHC. This indicator was first introduced in the UK in 2021 with the need to build an absolute measure to track progress and policy efficiency (Oxley (2023)). It is meant to capture all low income households living in inefficient homes. Few of the papers just mentioned have also used econometric methods to study the main socio-demographics determinants for households to fall into the alternative categories of fuel poverty.

[16] In his policy paper, he also makes several recommendations, among which he suggests to use income after housing costs normalised by consumption units in line with OECD recommendations (OECD (2016))

Study	Energy poverty metrics	Determinants
Boardman (1991)	Ratio = $\frac{\text{Required Fuel expenditures}}{\text{Income}}$ > 10% which corresponds to twice the median (2M) in the sample	Not studied
Hills (2012)	Low Income High Costs (LIHC) if 1. "required fuel costs that were above the median level" 2. "were they to spend that amount, they would be left with a residual income below the official poverty line" Residual income is considered after housing costs.	Not studied
Moore (2012)	Budget standard approach, Minimum Income Standard (MIS) fuel poor if fuel costs > Net household income – housing costs – minimum living costs	Not studied
Romero et al. (2018)	• 10% indicator • Moore (2012)'s indicator: fuel costs > household income - housing costs - minimum living costs • Operational adaptation of Hills (2012)'s LIHC with : household income - household expenditure on energy < 60% [median household income - mean expenditure on energy]	Binary Logit Regression results: the most vulnerable to energy poverty are low-income households, with children, household heads with job instability
Belaïd (2018)	Hills (2012)'s LIHC with energy expenditures in €/m^2	Multiple Correspondence Analysis (MCA) and Ascending Hierarchical Classification (AHC) to identify 4 fuel poor profiles. "(i) foreign family, employed, in shared building group, (ii) single person, retired, in small size flat group, (iii) family in individual house with gas and individual central heating system group and (iv) owner of high size rural house group". Also use logit regression to identify critical factors impacting the odds of being fuel poor.
Belaïd (2022b)	same as Belaïd (2018)	Clustering and regression methods similar to Belaïd (2018). Fuel poor household types in Jordan and Egypt are: (i) older households with higher incomes relative to the rest of the fuel-poor households, homeowners living in rural areas (ii) married homemakers, living in apartments (iii) lower incomes relative to other fuel poors, homeowners living in appartments

Belaïd and Flambard (2023)	Extension of LIHC: defines three categories of fuel poor to disentangle the effects of housing and fuel consumption: 1. low income, high housing costs and high fuel costs 2. low income, high housing costs and low fuel costs 3. low income low housing costs and high fuel costs	Trivariate probit regression to investigate critical factors of fuel poverty in Egypt: fuel poor households have large families, live in detached houses and have a low educated household's head.
Belaïd (2022a)	Low-Income Low Energy Efficiency (LILEE) 1. "have an FPEER[17] equal or lower than D" 2. "were they to spend that amount, they would be left with a residual income below the national standard poverty line." Income threshold set at 60% national median equivalized residual income after all housing-related expenditures per consumption unit and energy bill	Not studied
Legendre and Ricci (2015)	• 10% indicator • After fuel cost: Income - housing costs - fuel costs < 60% median. Focus on *vulnerable* households that are not below the poverty line before fuel costs. • Hills (2012)'s LIHC	C log-log and mixed effect logit model to investigate wich factors influence the odds of being fuel vulnerable using the French 2006 National Housing Survey: the probability of being vulnerable is higher for retired households, living alone, renting their home, with poor roof insulation, using an individual boiler for heating and cooking with gas.
EU Energy Poverty Observatory (Thema and Vondung (2020))	Expenditure-based indicators • M/2: Low absolute energy expenditure. Energy expenditures below half the national median. • 2M: Share of energy expenditure over income above twice the national median	Not studied

[17] Fuel Poverty Energy Efficiency Rating

French National observatory on Energy Poverty Lesueur (2020)	• 10% indicator restricted to the poorest 30% (TEE indicator)[18] • Hills (2012)'s LIHC • LIHC adaptation with energy expenses in relation to the size of the dwelling (m^2). • Qualitative declarative metrics of discomfort and cold	No studied

Table 6: Fuel poverty metrics and determinants

The choice of the indicator has significant effects on the extent of fuel poverty inside the home. Legendre and Ricci (2015) apply three indicators to the 2010 French household survey (the 10% indicator, the after fuel cost indicator and the LIHC indicator) and find that 17% spend more than 10% of income on energy, 21% are under the poverty line after housing and fuel expenditures at home and 9.2% are fuel-poor in terms of the LIHC indicator.

6.2. Description of predictors

In the following table, we describe in details the variables and their categories, used in this paper.

Variable	Description
Income decile	We observe income decile membership of households from 1 (poorest) to 10 (richest)
Household size & composition	We observe the number of children in a household with 4 levels: no children, one child, two children and three and more children[19]. We also take into account the number of adults, but only include it as an explanatory variable in the LIHTC[20].
Tenure type	We consider two categories: homeowners and tenants.
Socio-professional categories	We observe 7 categories: Farmers, Self-employed (craftsman, company director, merchant, and others), executives, intermediate professions, employees, unemployed who never worked and inactive (including students), and pensioners.
Categories of municipalities	We use the 2020 typology of urban catchment areas[21]. We consider 3 categories: urban (city centers of high and medium density), commuting areas (medium and low density), and isolated rural (low and very low density) areas.
Distance to closest train station	This variable is a proxy for access to public transport. It has 4 levels: below 2km, between 2 and 5km, between 5 and 10km, and above 10km.
Number of cars	Number of cars owned by the household. We consider four categories: no cars, 1 car, 2 cars, and 3 and more cars.
Age class	We observe the age of the household's reference individual. We group them into age classes. Categories are: below 25, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, and above 70.

[18] Taux d'Effort Energétique in French

[19] We assume that households with children may have additional travels for child-related activities, resulting in higher fuel consumption

[20] This is because, in this model, we cannot include income decile membership as an explanatory variable due to the use of mean income to calculate the indicators. Therefore, the number of adults serves as a proxy for income. However, to avoid collinearity, we exclude this variable in the other two models.

[21] It defines urban areas as city centers and commuting areas. A municipality belongs to a commuting area if more than 15% of its population works in the city center. There are principal and secondary city centers. Municipalities that do not belong to these commuting areas are considered isolated (INSEE (2020))

Gender of household's reference individual	Gender is a binary variable with two categories: man and woman.

Table 7: Description of variables

6.3. Multi-correspondence Analysis and Hierarchical clustering

In this section, we detail the method and results of the two-step clustering process that we perform in this paper.

6.3.1. Multi-correspondence Analysis (MCA)

Multi-correspondence Analysis is a data analysis method extending the principal component analysis (PCA) to categorical variables. Categorical variables are combined into synthetic continuous variables called principal components. Formally, if there is K variables and I observations, principal component analysis consider household observations as a cloud of points in a K multidimensional space and the K variables as a cloud of points in a I multidimensional space. If K≤3 (respectively I), it is possible to represent the cloud of points in a two- or three-dimension graph. However, if K≥4 (respectively I), visualization of the data is not possible. The idea behind principal component analysis is to look for the two-dimension subspace that best summarizes the data.

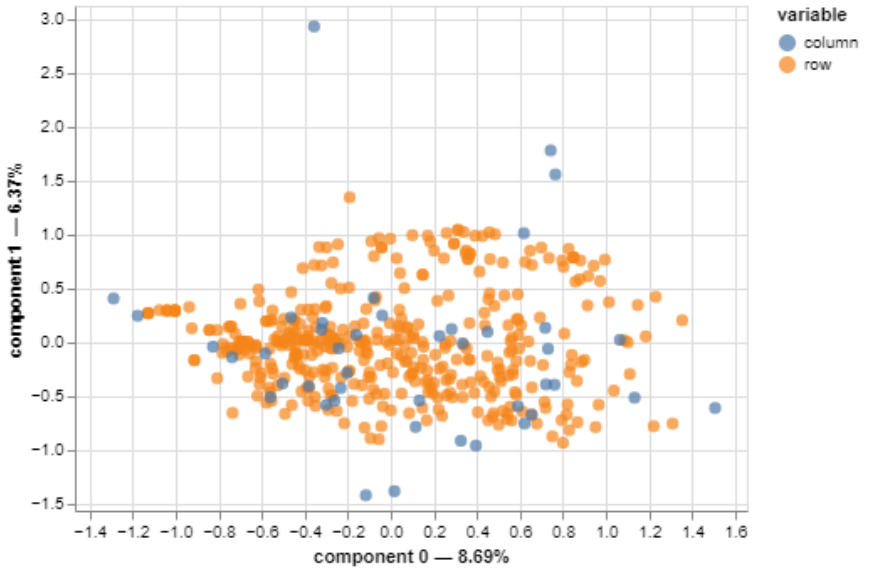

Figure 3: MCA biplot for LIHTC households. Rows are households (observations) and columns are variable categories

The proximity of the points in the space indicates the degree of correlation between the corresponding categories. The closer the points, the stronger the association. For example, for the Low-income High Transport Cost (LIHTC) group, we find that isolated rural and farmers are very close on the first dimension, while inactive and below 25 are also very close on this dimension but at the opposite side. We indicate the percentage of variance captured by each principal component on the horizontal and vertical axes in Figure 3. Note that for the LIHTC category, the cumulative percentage of variance captured by the first two principal components is low. This is because we choose to consider all variables as active variables. In an alternative setting, we separate a reduced set of three active variables that we believe to best capture the variability within the LIHTC population (Tenure type, Municipality type and Distance to train station). The remaining variables are called supplementary variables and help interpret the results but do not constitute the main axes. We can plot all variables in the updated two-dimension subspace figure 4 and we see that the two updated axes capture about 50% of data variability. For the remainder of this paper, we keep all the variables as active variables.

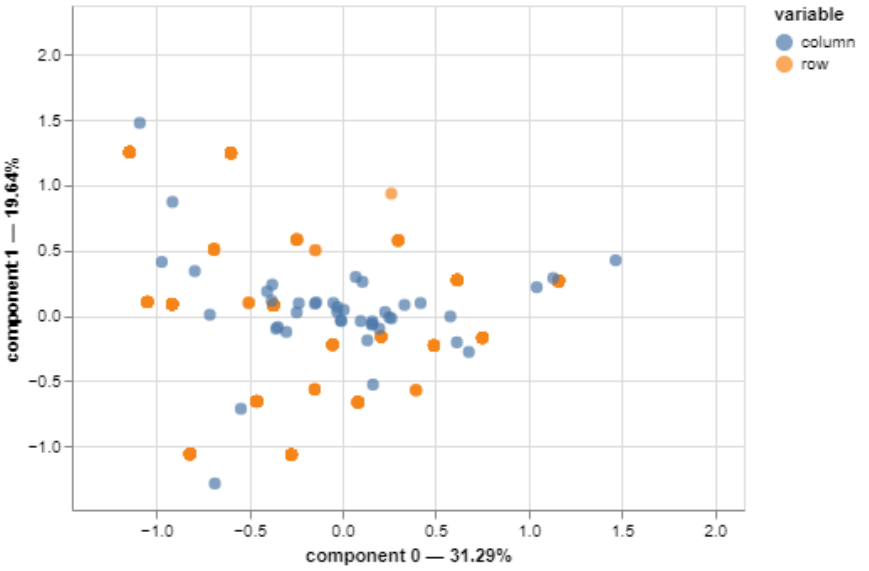

Figure 4: MCA biplot for LIHTC households. Rows are households (observations) and columns are variable categories. Our active variables are tenure type, municipality type, and distance to train station

6.3.2. Hierarchical clustering

We use hierarchical clustering to explore the existence of *natural* groups of households in the fuel poverty categories and their underlying structure.

Formally, considering K variables, Q clusters of individuals, and I_q individuals in each cluster, the total inertia (multidimensional variance) is:

$$\underbrace{\sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{i=1}^{I_q}(x_{iqk}-\bar{x}_k)^2}_{\text{total inertia}} = \underbrace{\sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{i=1}^{I_q}(x_{iqk}-\bar{x}_{qk})^2}_{\text{within inertia}} + \underbrace{\sum_{k=1}^{K}\sum_{q=1}^{Q}\sum_{i=1}^{I_q}(\bar{x}_{qk}-\bar{x}_k)^2}_{\text{between inertia}} \tag{2}$$

With x_{iqk} the value of factor k for observation i in cluster q, $\bar{x}_{qk}$ the mean value of k in cluster q and $\bar{x}_k$ the mean of factor k accross all clusters. We choose *Ward's criterion* to aggregate observations into clusters. It consists in the following algorithm. We start with n clusters of size 1, one cluster per observation (here, household). At the beginning, the within-class inertia is null because there is a unique observation in each cluster, and the *total inertia* is equal to *between inertia*. In the first step, n-1 clusters are formed, with one cluster of size 2 and n-1 clusters of size 1. The cluster of size 2 contains the two observations for which the merger yields the smallest increase in within-cluster inertia. Then, at each step, the algorithm merges clusters A and B if they are the closest according to Ward's measure of similarity:

$$\Delta_{ward}(A,B) = \frac{I_A I_B}{I_A + I_B} d^2(\mu_A, \mu_B) \tag{3}$$

With μ_i and I_i the barycentre and size of cluster i, respectively. d is the Euclidean distance between the barycentres. The algorithm stops when all observations are grouped into a single cluster.

The result of hierarchical clustering is presented in the dendrogram Figure 5. The length of the branches represents the distance between the clusters. Choosing the number of clusters means deciding where to *cut the tree*. Different criteria and methods can be used to determine the appropriate number of clusters, such as visualizing the dendrogram, setting a maximum distance for observations of the same class, and using the *elbow method*. For the LIHTC population (Figure 5), the dendrogram presents three clear clusters and around 12 sub-clusters. Setting a maximum distance of 6 leads to an optimal number of clusters of 3. Note that the maximum distance criterion is subjective in the sense that a

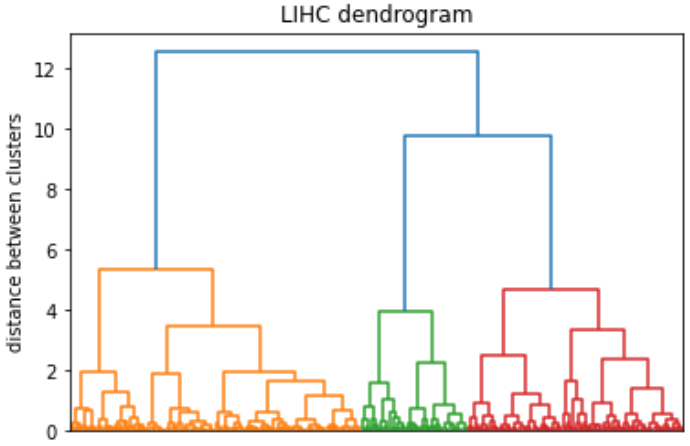

Figure 5: Dendrogram for the low-income high transport cost population

lower maximum distance value will result in more clusters, while a higher value will yield fewer clusters. Finally, the *elbow method* considers the Within-Cluster Sum of Squares (WCSS) as a function of the number of clusters. The result is presented in Figure 6. Visual inspection of the elbow plot allows for the identification of the number of clusters for which the WCSS takes off. In Figure 6, we observe that going from 3 to 2 clusters yields an increase in WCSS that is about twice the increase that we get if we go from 4 to 3 clusters, which confirms that 3 clusters is a relatively good number of clusters for our analysis of the Low-Income High Transport Cost population.

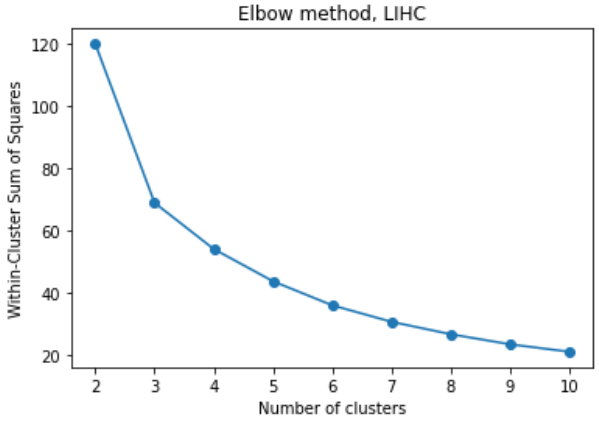

Figure 6: Elbow plot for low-income high transport cost households

6.3.3. Characterising the partition

We want to make sure that our categorical variables characterize the clusters. For the resulting partition of the data, we compute the chi-square statistics and the corresponding p-value between the class variables and the categorical variables. This way, we can assess the importance of the association and determine if it is statistically significant. Results for the low-income high-cost population are presented in table 8. Values are sorted by increasing p-values. Table 8 shows that all variables are significant to characterize the clusters and that the most significant variables are occupation status (pensioner, farmer, Self-employed, worker, etc) and age class (below 25, 25-29, ..., above 70). On the other hand, the gender of the household head has a rather low impact on the LIHTC partition.

Variable	Chi-square statistic	P-value
Socio-prof. cat.	350	2e-67
age	346	1e-62
Tenure	131	3e-29
Number of car	128	1e-26
Category of municipality	125	4e-26
dist train	130	1e-25
income decile	65	5e-11
Number of adults	44	6e-09
Number of children	44	9e-08
Gender	12	3e-03

Table 8
Chi-square test for the Low-Income High Transport Costs cluster variables

Then, we describe the composition of the 3 clusters and observe the categories within our social groups are over-represented in a specific cluster, compared to the rest of the low-income high-cost population, as well as the total population. We describe the clusters population in section 4.

Category	Cluster 1 (51%)	Cluster 2 (18%)	Cluster 3 (31%)	LIHTC	All
decile 1	44	28	28	36	11
decile 2	46	64	33	45	10
decile 3	10	6	37	18	11
decile 4	1	1	2	1	9
decile 5	0	1	0	0	10
1 child	14	1	16	12	16
2 children	9	0	18	10	14
3+ children	7	0	6	5	6
no children	70	99	60	72	64
1 adult	87	80	51	74	47
2 adults	13	20	49	25	52
3+ adults	0	0	0	0	1
1 car	89	80	32	70	45
2 cars	10	18	50	24	28
3+ cars	0	2	18	6	7
Employees and workers	57	5	40	42	25
Executives	2	1	2	2	11
Farmers	1	0	21	7	2
Self-employed	7	1	20	10	7
Intermediate	17	0	8	11	14
pensioners	0	91	8	19	37
Unemployed and inactive	16	1	1	9	4
Owner	15	59	75	42	59
Tenant	85	41	25	58	41
Commuting area	33	49	56	43	40
Isolated rural	2	23	31	15	7
Urban	65	28	13	42	54
2-5km	26	21	19	23	22
5-10km	11	20	22	16	15
Above 10km	10	35	55	29	16
Below 2km	53	25	4	32	47

Continued on next page

Category	Cluster 1 (51%)	Cluster 2 (18%)	Cluster 3 (31%)	LIHTC	All
25-29	10	0	8	7	6
30-34	14	0	7	9	7
35-39	10	0	7	7	9
40-44	18	0	8	12	9
45-49	8	0	21	11	10
50-54	9	2	15	10	9
55-59	6	3	19	9	10
65-69	1	39	11	11	10
Above 70	0	56	2	11	25
below 25	24	0	3	13	4
man	41	58	60	50	58
woman	59	42	40	50	42

Table 9: Distribution of each cluster, all LIHTC and the entire population within the variable categories. Results are expressed in percentage (%)

Table 9 presents the distribution of each cluster within each variable category as well as the distribution of all low-income high cost households (LIHTC) and the entire population as a comparison.

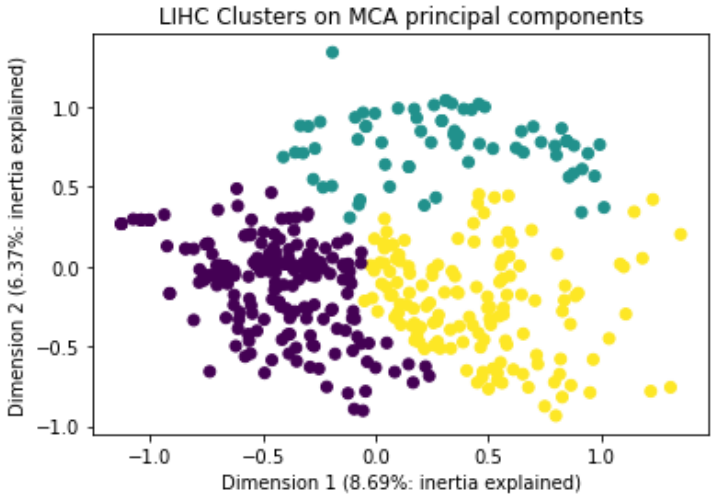

Figure 7: Projection of the 3 LIHTC clusters on the two principal components

Visualization of the clusters in the MCA biplot: In Figure 7 we see clearly the three clusters: cluster 1 on the bottom left (single-person household living in urban areas with access to public transport, employees, workers and inactive), cluster 2 in the middle (large families in rural areas) and cluster 3 on the bottom right (pensioners living in rural areas).

6.3.4. Twice the median - 2M households

Hierarchical clustering of 2M households leads to the identification of 3 main household clusters presented in Figure 9. The Elbow plot in Figure 10 confirms that three is an appropriate number of clusters.

The distribution of clusters within the variable categories are the following:

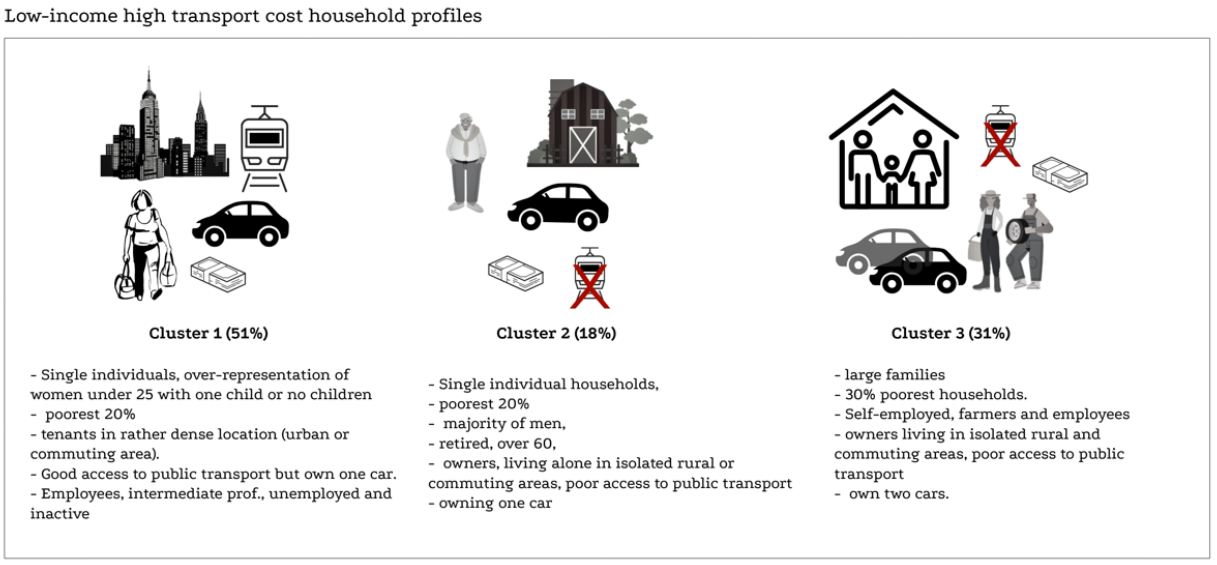

Figure 8: 3 clusters inside the LIHTC category

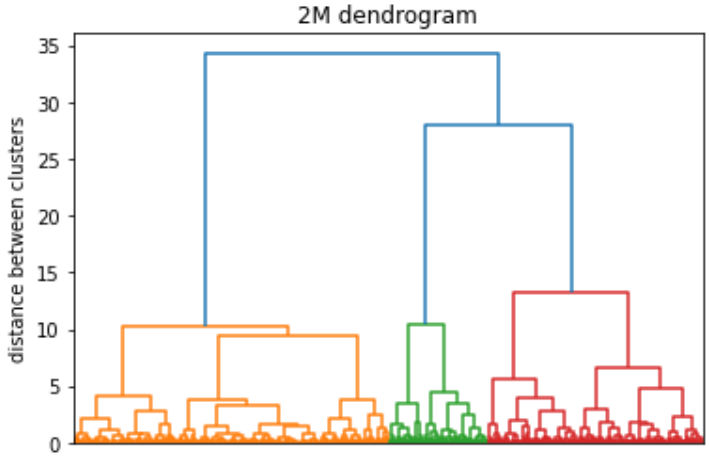

Figure 9: Dendrogram for *twice the median* households

Category	Cluster 1 (45%)	Cluster 2 (16%)	Cluster 3 (39%)	2M	All
decile 1	1	20	27	14	11
decile 2	2	18	18	11	10
decile 3	3	15	20	12	11
decile 4	5	14	17	11	9
decile 5	12	13	10	11	10
decile 6	16	11	5	11	10
decile 7	25	5	2	13	11
decile 8	20	3	1	10	10
decile 9	15	1	0	7	10

Continued on next page

Category	Cluster 1 (45%)	Cluster 2 (16%)	Cluster 3 (39%)	2M	All
decile 10	2	0	0	1	10
1 child	28	6	19	21	16
2 children	32	1	10	19	14
3+ children	16	0	6	10	6
no children	24	92	65	51	64
1 adult	6	50	78	41	47
2 adults	92	49	22	58	52
3+ adults	2	1	0	1	1
1 car	3	54	80	41	45
2 cars	63	37	19	42	28
3+ cars	34	9	1	17	7
Employees and workers	40	2	53	39	25
Executives	9	0	6	6	11
Farmers	7	4	1	4	2
Self-employed	18	1	9	11	7
Intermediate	24	0	21	19	14
pensioners	1	91	1	16	37
Unemployed and inactive	1	1	8	4	4
Unknown	1	0	0	0	0
Owner	78	82	28	59	59
Tenant	22	18	72	41	41
Commuting area	65	62	39	54	40
Isolated rural	15	15	6	11	7
Urban	21	22	55	34	54
2-5km	22	21	25	23	22
5-10km	30	23	12	22	15
Above 10km	32	33	17	26	16
Below 2km	17	23	46	29	47
25-29	7	0	13	8	6
30-34	10	0	10	9	7
35-39	14	0	12	11	9
40-44	16	0	13	12	9
45-49	19	0	11	13	10
50-54	14	0	13	12	9
55-59	14	3	11	11	10
65-69	5	42	3	10	10
Above 70	0	54	0	9	25
below 25	1	0	15	6	4
man	75	67	47	63	58
woman	25	33	53	37	42

Table 10: Distribution of each cluster, all 2M households and the entire population within the variable categories. Results are expressed in percentage (%)

6.3.5. Half the median - M/2 households

For M/2 households, hence households that spend less than half the median of the population on car fuels, the dendrogram is presented in Figure 12.

We could choose to cut the tree with a distance of 30 and select only two clusters (see Figure 12). However, the Elbow plot (Figure 13) shows that selecting two clusters instead of three substantially increases within-cluster inertia. Therefore, we keep the third cluster even if it represents only a small part of the M/2 households (15%). We think that

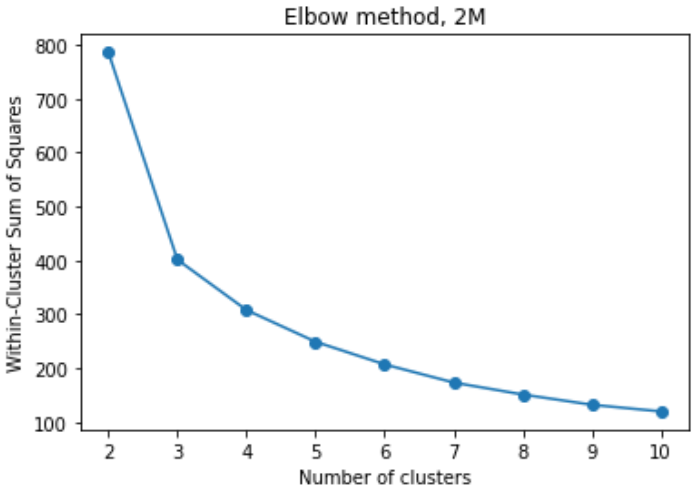

Figure 10: Elbow method for 2M

Variable	Chi-square statistic	P-value
Socio-prof. cat.	2.23E+03	0.00E+00
age	2.15E+03	0.00E+00
Number of car	1.38E+03	1.61E-296
income decile	1.35E+03	1.15E-275
Number of adults	1.15E+03	9.21E-249
Number of children	7.09E+02	6.37E-150
Tenure	6.46E+02	4.37E-141
Category of municipality	3.05E+02	8.61E-65
dist train	2.71E+02	6.54E-54
Gender	1.63E+02	3.23E-36

Table 11
Chi-square test between the class variable and the categorical variables for 2M clusters.

it is relevant as it allows to separate low-income households from rich urban households, as described in the results in section 4.

Category	Cluster 1 (47%)	Cluster 2 (15%)	Cluster 3 (38%)	M/2	All
decile 1	17	2	37	22	11
decile 2	19	3	21	18	10
decile 3	19	4	16	16	11
decile 4	14	7	9	11	9
decile 5	10	10	7	9	10
decile 6	8	12	5	7	10
decile 7	5	13	3	6	11
decile 8	4	16	1	5	10
decile 9	2	13	0	3	10
decile 10	1	19	0	4	10
1 child	3	25	15	11	16
2 children	1	26	7	7	14
3+ children	0	17	3	4	6
no children	96	32	76	78	64

Continued on next page

Category	Cluster 1 (47%)	Cluster 2 (15%)	Cluster 3 (38%)	M/2	All
1 adult	75	21	90	72	47
2 adults	25	75	9	27	52
3+ adults	0	4	1	1	1
1 car	55	47	19	40	45
2 cars	4	4	0	3	28
3+ cars	0	1	0	0	7
no cars	40	49	80	57	20
Employees and workers	5	35	41	23	25
Executives	1	35	6	8	11
Farmers	1	0	0	1	2
Self-employed	2	10	3	4	7
Intermediate	1	17	13	8	14
pensioners	86	2	17	47	37
Unemployed and inactive	4	0	19	9	4
Unknown	0	1	0	0	0
Owner	68	35	11	41	59
Tenant	32	65	89	59	41
Commuting area	42	7	10	24	40
Isolated rural	8	0	1	4	7
Urban	50	93	89	72	54
2-5km	22	14	16	18	22
5-10km	15	1	3	8	15
Above 10km	20	2	3	11	16
Below 2km	43	83	77	62	47
25-29	1	10	9	6	6
30-34	1	17	8	6	7
35-39	1	15	10	6	9
40-44	1	17	8	6	9
45-49	2	17	8	6	10
50-54	2	7	10	6	9
55-59	4	10	11	8	10
65-69	11	4	7	8	10
Above 70	78	1	13	42	25
below 25	0	1	16	6	4
man	42	73	36	44	58
woman	58	27	64	56	42

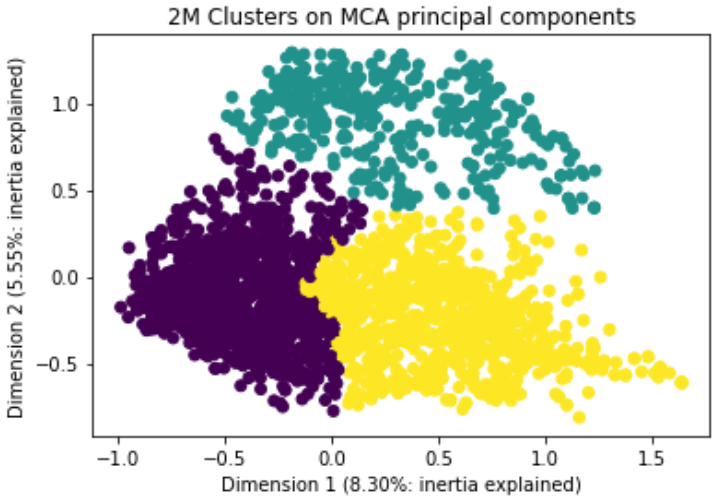

Figure 11: Projection of 2M clusters on the MCA principal components

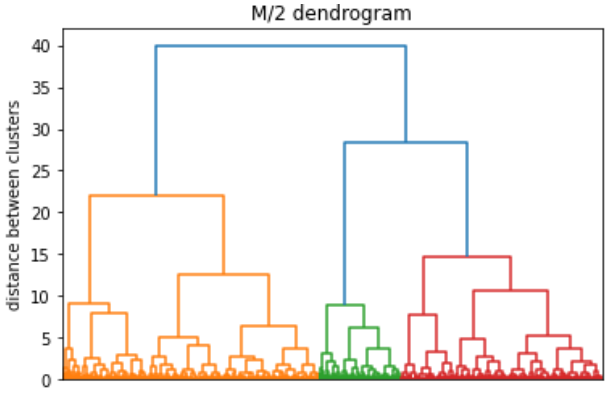

Figure 12: Dendrogram for M/2 households

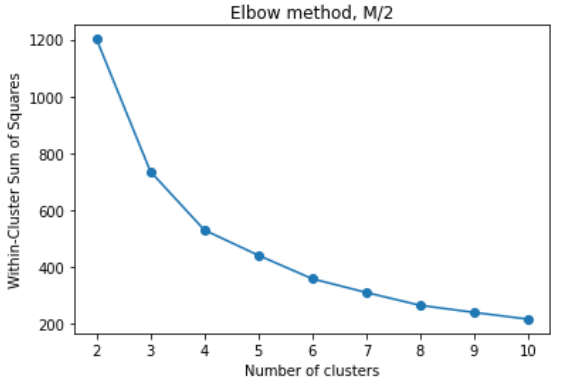

Figure 13: Elbow plot for M/2 households

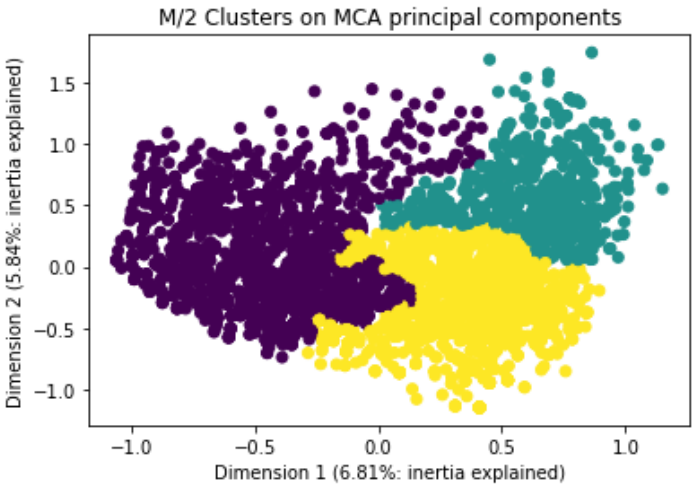

Figure 14: Projection of M/2 cluster on principal components

Variable	Chi-square statistic	P-value
Socio-prof. cat	2.66×10^{3}	0.00×10^{0}
age	2.37×10^{3}	0.00×10^{0}
Number of children	1.31×10^{3}	3.34×10^{-279}
decile_rev	1.33×10^{3}	8.37×10^{-272}
Tenure	1.24×10^{3}	6.84×10^{-270}
Number of adults	9.94×10^{2}	5.71×10^{-214}
Category of municipality	8.44×10^{2}	2.14×10^{-181}
dist_train	7.02×10^{2}	2.95×10^{-146}
Number of car	6.32×10^{2}	2.65×10^{-133}
SEXEPR	2.54×10^{2}	7.53×10^{-56}

Table 13
Chi-square test between the class variable and the categorical variables for 2M clusters.

6.4. Robustness of the three regression models

In this section, we present three robustness tests for the logistic regressions. First, we test for multicollinearity between our variables. Then we assess the predictive capacity of our model with McFadden and Nagelkerke pseudo-R-squared measures. Finally we test for goodness of fit with the Hosmer and Lemeshow test.

6.4.1. Variance Inflation Factor (VIF)

We assess the degree of multicollinearity between our predictors for the three model. We reduce the number of explanatory variables in each model in order to obtain VIF below two for all predictors of each of the three regression models. VIF results are presented in Tables 14, 15 and 16.

Term	VIF	VIF 95% CI	Increased SE	Tolerance	Tolerance 95% CI
Number of children	1.26	[1.23, 1.29]	1.12	0.79	[0.78, 0.81]
Number of adults	1.66	[1.62, 1.70]	1.29	0.60	[0.59, 0.62]
Number of car	1.85	[1.80, 1.90]	1.36	0.54	[0.53, 0.56]
Socio-prof. cat	1.36	[1.33, 1.40]	1.17	0.73	[0.72, 0.75]
Tenure	1.30	[1.27, 1.33]	1.14	0.77	[0.75, 0.79]
Category of municipality	1.63	[1.59, 1.68]	1.28	0.61	[0.60, 0.63]
dist train	1.60	[1.56, 1.64]	1.26	0.63	[0.61, 0.64]

Table 14
Variance Inflation Factors (VIF) for the LIHTC model

Term	VIF	VIF 95% CI	Increased SE	Tolerance
Number of children	1.35	[1.32, 1.38]	1.16	0.74
Number of car	1.30	[1.27, 1.33]	1.14	0.77
Socio-prof. cat	1.42	[1.39, 1.45]	1.19	0.70
Tenure	1.31	[1.28, 1.34]	1.15	0.76
Category of municipality	1.53	[1.49, 1.57]	1.24	0.65
dist train	1.48	[1.45, 1.52]	1.22	0.68

Table 15
Variance Inflation Factors (VIF) for the 2M model

Table 16
Variance Inflation Factors (VIF) for the M/2 model

Term	VIF	VIF 95% CI	Increased SE	Tolerance	Tolerance 95% CI
decile rev	1.62	[1.58, 1.66]	1.27	0.62	[0.60, 0.63]
Number of children	1.37	[1.34, 1.40]	1.17	0.73	[0.71, 0.75]
Number of car	1.12	[1.10, 1.15]	1.06	0.89	[0.87, 0.91]
Socio-prof. cat	1.84	[1.80, 1.89]	1.36	0.54	[0.53, 0.56]
Tenure	1.36	[1.33, 1.39]	1.16	0.74	[0.72, 0.75]
Category of municipality	1.56	[1.53, 1.60]	1.25	0.64	[0.62, 0.66]
dist train	1.55	[1.51, 1.59]	1.24	0.65	[0.63, 0.66]

Pseudo R-squared	LIHTC	2M	M/2
McFadden	0.15	0.11	0.24
Nagelkerke	0.18	0.17	0.33

Table 17
Pseudo R-squared for the three logit models

6.4.2. Predictive capacity

Table 17 presents the pseudo R-squared of the three logistic regression models computed as in McFadden (1974a).

Group	Total	Observed (def=1)	Expected (def=1)	Observed (def=0)	Expected (def=0)
1	1168	3	3.53	1165	1164.47
2	1199	2	7.49	1197	1191.51
3	1097	7	9.78	1090	1087.22
4	1148	16	13.59	1132	1134.41
5	1158	26	18.35	1132	1139.65
6	1149	25	24.99	1124	1124.01
7	1158	27	34.38	1131	1123.62
8	1279	66	59.14	1213	1219.86
9	1030	77	78.54	953	951.46
10	1144	191	190.21	953	953.79

Table 18
Partition for the Hosmer & Lemeshow Test, LIHTC model

Chi-Square	DF	Pr > ChiSq
11.1035	8	0.1959

Table 19
Goodness of Fit Test, LIHTC model

6.4.3. Goodness of fit

To test the goodness of fit of our three logistic model, we perform the Hosmer and Lemeshow test- (Hosmer Jr et al. (2013)). In simple terms, the test partitions the sample into groups, based on the predictive probabilities. For each group, the test compares the observed probabilities with the expected probabilities. The null hypothesis is that the differences between these probabilities are not significant. Having large p-values means that we cannot reject the null hypothesis, which is an argument in favour of good fit of our model. Results for the three models are presented in Tables 18, 20 and 22 show that in each case we can reject the null hypothesis.

Group	Total	Observed (def=1)	Expected (def=1)	Observed (def=0)	Expected (def=0)
1	1188	55	69.49	1133	1118.51
2	1148	92	106.04	1056	1041.96
3	1127	129	131.70	998	995.30
4	1150	191	175.89	959	974.11
5	1166	233	216.42	933	949.58
6	1141	264	253.00	877	888.00
7	1167	311	309.94	856	857.06
8	1148	386	373.42	762	774.58
9	1147	444	459.67	703	687.33
10	1148	624	633.43	524	514.57

Table 20
Partition for the Hosmer & Lemeshow Test, 2M model

Chi-Square	DF	Pr > ChiSq
10.8629	8	0.2096

Table 21
Goodness of Fit Test, 2M model

Group	Total	Observed (def = 1)	Expected (def = 1)	Observed (def = 0)	Expected (def = 0)
1	1154	8	7.54	1146	1146.46
2	1169	17	16.70	1152	1152.30
3	1136	15	23.50	1121	1112.50
4	1153	38	38.08	1115	1114.92
5	1153	81	74.29	1072	1078.71
6	1153	178	168.45	975	984.55
7	1169	271	257.30	898	911.70
8	1139	342	346.56	797	792.44
9	1151	429	457.50	722	693.50
10	1153	630	619.09	523	533.91

Table 22
Partition for the Hosmer Lemeshow Test, M/2 model

Chi-Square	DF	Pr > ChiSq
8.8373	8	0.3562

Table 23
Goodness of Fit Test, M/2 model