Energy Poverty:

Measurement Strategies and Solutions

Master Thesis defended by:

Matteo Farinoni

Universitat Politècnica de Catalunya

KIC InnoEnergy, MSc Energy for Smart Cities

Directors:

Andreas Sumper

CITCEA – Universitat Politècnica de Catalunya

Adriana Anguera Jordà

everis

July 2016

ETSEIB

an NTT DATA Company

Abstract

This work wants to propose measurement methodologies and solutions for tackling the energy poverty and affordability issue in developed countries, focusing on the European Union and in particular on Spain and Catalonia.

The research is carried out as a support tool for policy makers and public authorities, providing an objective and scientific evaluation of a problem which is currently at the centre of both the political and economic debate. Two are the aims of this project.

First aim is to analyse and test, on a real database, all the indicators used throughout Europe so far. This will lead to the choice of a suitable indicator that could be applied to Spain for assessing and estimating the energy poverty extension and impact over Spanish society.

Second aim, based on previous step, is to model the phenomenon in an innovative manner using machine learning instruments. This will allow to understand what are the variables that increase the risk for a single households of facing an energy vulnerability situation. As a core added value, the analysis will not take into account information that are commonly owned by private utility companies.

In the final part of the project the results obtained from the trained model are applied and tuned to a specific study case: the city of Barcelona. An energy vulnerability ranking will order all city neighbourhoods according to their probability of hosting families in energy deprivation conditions. Moreover, it will be possible to evaluate the drivers of the problem case by case.

This outcomes can set the base for the implementation of effective policies following a specific and demonstrated framework and order of action, optimizing and controlling the use of public financial resources.

ETSEIB

everis

an NTT DATA Company

Table of Contents

1 Introduction and Objectives

The *energy poverty* issue has gained particular importance in the last ten years due to the significant increase of energy products prices. The economic and financial crisis, started in 2008, has heavily contributed to weaken European citizens' ability of affording *adequate comfort levels* in their dwellings. Energy efficiency also plays a major and central role in the energy poverty issue, particularly in relation with the poor conditions of Southern European countries' housing stocks.

This work, as final contribution to the study track MSc in Energy for Smart Cities, wants to assess and tackle the energy poverty issue at urban level proposing new solutions and approaches for supporting public authorities' decision making process. The analysis will focus on Spain, and in particular on the city of Barcelona. The latter is facing a particularly grief situation, where the number of forced disconnection by utilities companies is significantly increasing, contributing to the establishment of the current social crisis widely discussed in public debates and newspapers.

The intrinsic lack of statistically available data increases problem's complexity, making particularly difficult to define which households should or should not be considered energy poor. The problem is strictly related to the concept of Smart City, since we are seeking methodologies and tools for gathering and optimizing citizens' information and data with the scope of enhancing their life quality, social equality in cities and *energy affordability*. In particular the latter has been defined by the International Energy Agency (IEA) as a fundamental pillar for the future of energy systems besides *security of supply* and *sustainability*.

This study addresses this issue, providing some instruments for identifying firstly a coherent and precise energy poverty definition and secondly an algorithm for modelling and studying the problems drivers in depth based on the freely available data.

In particular, the main aim is to provide an *active energy poor households research methodology* to report which city's neighbourhood are facing stringent energy vulnerability conditions, diagnosing, case by case, which are the causes of the problem (e.g. unstable job condition or low energy efficiency standards).

As a major added value, the study has the objective of solving the intrinsic information asymmetry existing between public authorities and private utilities.

The previous step will allow policy makers and public authorities to have a better understanding of the phenomenon and a better control and optimization of the financial resources made available to solve it.

A systematic and well-planned energy poverty solution strategy would result in lower health burdens for the country and in higher policies targeting efficiency, increasing the total *surplus* for society.

In the initial part (Chapter 3) we will cover what is the background scenario of *energy poverty* in the European context, analysing which are the aspect to be considered, underlining the central role of a coherent *energy poverty* definition and a precise assessment strategy that is still lacking in the majority of Member States.

Chapter 4 will analyse a series of proposed indicators, providing an evaluation and decisional framework, and it will conclude by choosing a specific and appropriate measurement strategy.

In Chapter 6, three machine learning instruments will be used, with the main aim of recognizing meaningful and significant patterns of the problem using a national database made available by the Spanish Statistical Institute (INE). A modelling technique will prove to have best performances and it will be chosen for quantifying the importance and partial contributions of the variables considered in the database without taking into account the energy expense variable. In order to reduce the number of variables' considered, the work will select a subgroup that allows to achieve acceptable *accuracy* and model's *predictability power*. Afterwards, the methodology will be replicated for the Spanish urban context.

Chapter 7 will apply the modelling results, to the specific case of Barcelona, using and combining the freely available Open Data of the city. The analysis will end with the identification of the city's most vulnerable neighbourhoods and *energy poverty* drivers and causes on a case by case basis.

2 Methodology

In this preliminary Chapter we will spend a few words defining the approach and methodology followed in this work, describing both the instruments and data sources considered. As it will be further explained in next Chapters, statistics is at the core of the energy poverty discussion, so it has been crucial to retrieve and study different databases both at European and Spanish level.

The study will consider the *European Survey on Income and Living Conditions* (*EU SILC*) offered by EUROSTAT for studying the situation across the European Union.

In the following part the work will focus on Spain and Catalonia and the main reference would be the *Instituto Nacional de Estadística (INE)*, considering in particular two dataset publications:

- The *Encuesta de Presupuestos Familiares (EPF).*
- The *Encuesta de Condiciones de Vida (ECV).*

The former focuses on the analysis of the annual expenses of a selected population sample of 22,146 Spanish households divided among nineteen regions (i.e. *Comunidades Autónomas*). It is widely used to assess the economic and consumption patterns and yearly performance of the entire population. The latter has the scope of evaluating social issues, population necessities, and the impact of social and economic policies. The two data collections provide anonymous results for a particular population sample that can be extended to the entire Spanish population through particular weights and elevation factors. The latter represents the number of real families that are represented by each single sampled households.

The data will be processed with the R statistical software which offers a specific package (`MicroDatosEs`) for dealing with the aforementioned Spanish dataset. It is available from 2011 to 2014.

3 Energy Poverty Theory

3.1 General Concept

According to the Energy Union Package “A Framework Strategy for a Resilient Energy Union with a Forward-Looking Climate Change Policy” [1] [2] the main target is providing European citizens with *secure*, *sustainable*, and *affordable* energy. This work is going to analyse the status of the third element (i.e. *affordable energy)* at European level, mainly focusing on Spain and Catalonia.

The European Survey on Income and Living Condition (EU SILC) [3], redacted in May 2015, estimated that 54 million European citizens were unable to keep an *adequate temperature* in their dwellings in 2012. The Building Performance Institute Europe [4] concluded that, in 2014, between 50 to 125 million people in the European Union were unable to afford proper indoor thermal comfort.

It is possible to define such a situation, subject of this work, with the expressions: *energy vulnerability or energy poverty*. In the British case the term *fuel poverty* is used, too .

It refers to the inability of an household of paying or affording *adequate energy services* (electricity, natural gas…), which are considered as vital goods without proper substitutes, that allow sufficient integration in society and healthy life conditions [5]. Energy poverty can be related to all kind of domestic energy consumption, although, particular attention is paid on heating and refrigeration energy demand, since directly linked to most serious and dangerous health effects [6], as it will be demonstrated in Paragraph 3.3. Due to this proved relationship between energy poverty and serious health consequences is necessary to distinguish it from the general concept of monetary poverty and to treat has a separated social, technical and economic topic.

This is just a qualitative description of energy poverty, in fact, at European level, there is not, at the moment, a *formal and official definition* of the problem. Only the United Kingdom offers a legal identification and definition of energy poverty (see Annex E).

In the European Union context, energy vulnerability considerations were integrated with Directives 2009/72/EC and 2009/73/EC, concerning common rules for the internal Electricity and Natural Gas markets. Among other points, the Directives required Member States to adopt a definition of *vulnerable customers* [7]. Up to now (June 2016) the majority of European countries still have not fulfilled this Communitarian requirement and Spain belongs to this latter category, too (see Annex E).

INSIGHT-E (2015) [8] defines an evaluation framework to compare the situation across different Member States dividing the policies implementation process in three steps. Those are:

1. *Targeting* is the strategy that policy makers or public authorities adopt to choose the requirements that a family must fulfil to be officially considered in vulnerability conditions. Politically, it is the most complex step because several trade-offs should be made to define which category is really facing energy poverty conditions. The more elaborated and specific the targeting phase is, the higher would be the implementation cost. A *measurement tool* and strategy has to be determined to complete this step based on the targeting criterion adopted.

2. Once the problem is defined, population's statistical data must be collected to start the *Identification* phase. The statistical aspect plays a crucial role in detecting which are the households that meet the conditions determined in the first step.

3. The *Implementation* phase takes advantage of the previous two, providing aids or support to specifically identified families, according to a rigorously determined and objective definition of energy poverty.

It is not possible to efficiently tackle the energy vulnerability issue without considering the just described three steps ladder. At the moment, throughout European Member States, it appears that policy makers, finding the first two points particularly difficult to face, are starting by implementing policies without solid bases in targeting and identifying vulnerable population groups [5].

This work addresses only the first two points (i.e. *Targeting* and *Identification*) with the purpose of providing innovative tools and solutions for estimating, measuring and tackling the issue, using real population data.

We will now spend a few words discussing what are the energy poverty drivers and causes.

3.2 Energy Poverty Drivers

Energy poverty is, without any doubt, part of the monetary poverty issue from which it is *partially* not separable. It is crucial to use the word “partially”. Indeed, it is not possible to demonstrate or prove a *biunique relationship* existing between energy poverty and “general” poverty. A family can face energy vulnerability conditions without being officially recognized as living in “general” poverty condition, due, for instance, to major expenses on energy (e.g. caused by low efficiency standards or high dwelling’s floor area). It is therefore not possible to assess energy poverty just relying on “general” poverty indicators (i.e. based on household’s income), as it is a more *complex* and *heterogeneous phenomenon* [8].

To completely understand this consideration it is important and useful to explain what are the three recognised causes and drivers for energy poverty reflected in Figure 1.

These are: *energy price*, *energy efficiency*, and *household’ income*.

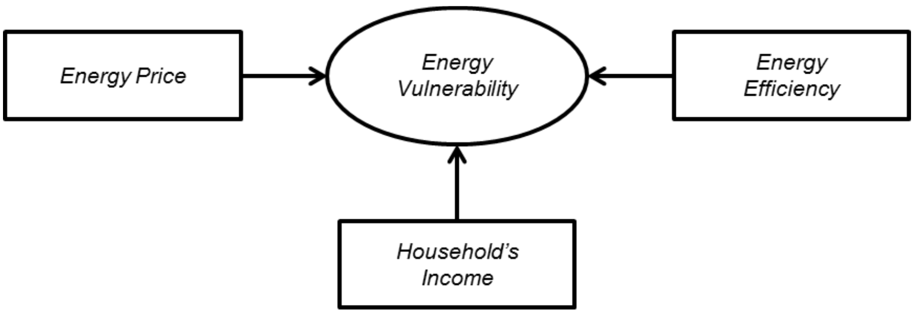


Figure 1 - Energy Poverty Drivers [Own Elaboration]

The main difficulties linked to energy poverty’s drivers analysis are caused by the specific nature of the problem. The latter is, in fact, *private* (being confined to the domestic domain), *temporally and spatially dynamic* (energy price fluctuations due to geopolitical or financial markets oscillations), and *culturally sensitive* (expectations of energy services are subjective and socially constructed). All these factors combined increase even more the difficulties in evaluating and assessing the phenomenon, as it combines factors that are not constant in time and affected by significant subjective components. Moreover, the private aspect makes the *identification* phase particularly difficult, mainly concerning household's income and energy expenses data collection.

In the following part we will spend some words for each of the three drivers treating, in particular, the Spanish situation.

Energy Retail Price

It refers to the chronological and geographical difference existing between energy products *retail price* values. Due to the *domestic* connotation of energy poverty, it is important to specify that the analysis focuses only on this sector, evaluating, as two principal energy sources, Electricity and Natural Gas. This hypothesis is justified in Figure 2, where the domestic final energy from Electricity and Natural Gas amounts to the 67.2% of the total.

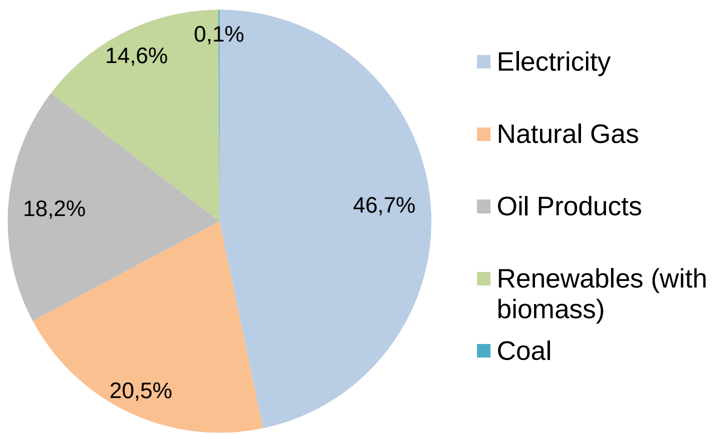


Figure 2 - Spanish residential sector final energy consumption share for different energy source [Own Elaboration based on IDEA data (2011)] [9]

In order to briefly describe what has been the trend for the last ten years, we refer to EUROSTAT data [10], reporting the retail prices of Electricity and Natural Gas for Spain, related to second semesters of each of the sampled years (between 2006 and 2015).

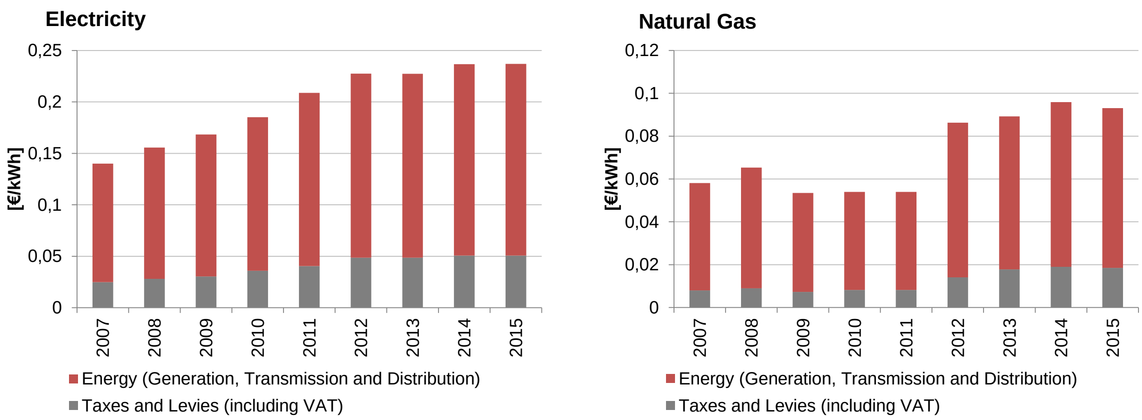
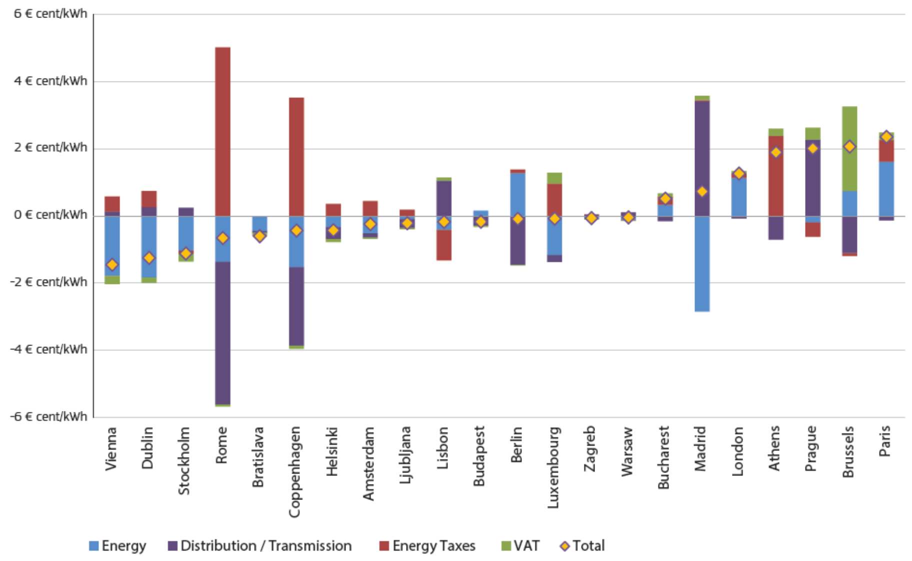


Figure 3 - Electricity and Natural Gas retail price evolution for the domestic sector between 2007 and 2015. [Own Elaboration based on EUROSTAT Energy Survey data (2007:2015)] [10]

Figure 3 shows how the Electricity price has increased by 70% from 2007 to 2015, where the Taxes and Levies component has increased, over the same period by 104% and the energy product component by 61%. The same can be said for Natural Gas with a total increment of 60%, where the Taxes and Levies variation has been of the 131.3%.

In order to put this data into perspective, assuming a recent high energy efficiency standards fridge (i.e. with a yearly consumption of 355 kWh), the difference for a family paying its annual running costs between 2007 and 2015 is almost €40. On the other hand, if one considers a fridge from the period 2000-2010 with average annual energy needs for 600 kWh the difference in running energy prices reaches €58.2.

The European Commission Quarterly Report on European Electricity Market (2015) [11] adds a further consideration to highlight the burden of the Taxes and Levies component on Spaniards' energy bills. Figure 4 represents, for European capital cities, the variations of electricity price components from 2014 to 2015. The reader can notice that for Madrid, the energy component decreased by almost 3 cents per kWh. On the other hand, the Distribution and Transmission component has increased at even higher pace bringing to a total increase of the domestic electricity retail price of almost 1 cent € per kWh.



Source: Vaasaett No data were alvailable for Nicosia, Riga, Sofia ,Tallinn, Vilnius and Valetta in June 2014

Figure 4 – Change in Electricity Price Components across selected European capital cities, between June 2014 and June 2015 [11]

The debt that customers encounter with distribution system operators (DSOs) is raising. Their cumulative debt in 2012 was €24,000 million and in 2014 was €30,000 million [12]. Further increases in tariffs were decided by utilities mainly to cover this huge debt, increasing for example the Distribution and Transmission component. Needless to say that such decisions were responsible for an even higher number of vulnerable customers following in stringent financial conditions and to forced disconnections.

It is possible to state, according to this chronological price evaluation, that, combined with household's *energy demand inelasticity*, the burden for Spanish families has increased considerably between 2007 and 2015, making the energy price component a primary driver for the increase of energy poor in Spain and for bringing the situation to the actual social emergency status, addressed by a considerable number of Spanish local and regional authorities (see Annex A).

In Annex D, the reader can find both Electricity and Natural Gas retail price (for the domestic sector) for all European Member States.

Energy Efficiency

The problem of energy efficiency has been widely treated in many European studies and projects with the objectives of fighting Climate Change and reducing energy consumption throughout the Union. It is strictly related to the energy affordability issue, too. Needless to say, in fact, that the total expenditure on energy is directly proportional to the energy needed to keep adequate and liveable conditions all year long.

European studies [13] [14] suggest that the Spanish building stock is particularly old and characterized by extremely low energy efficiency features. Many dwellings were built before that the *Norma Básica de la Edificación* (1979), the first regulation concerning thermal insulation, was introduced. In 2006 and in 2013 the *Codigo Técnico de la Edificación* added further updated building requirements in line with European Commission Directives.

Poor energy efficiency performances can affect people belonging to different social and income deciles, even though for the richest it is possible to improve dwelling's condition in shorter time and without major impact on personal financial balances. The problem can be particularly difficult to solve if poorest families are involved, for this reason, policy makers and public authorities must react and identify what are the families that really need financial intervention to improve their conditions and leave the energy vulnerability status. Once more, an efficient and precise targeting instrument is needed to identify which are the family living in risky situations where energy efficiency improvements might be needed.

Household's Income

Needless to specify that this third energy poverty driver has gained particular importance due to the economic crisis that has hit the whole European Union since 2008. The situation for Spain appears particularly grief with significant consequences on the number and conditions of unemployed and retired citizens. The combination of rising energy prices and lower families' income has significantly increased "general" poverty and energy poverty concerns. The European Commission study, "Alleviating Fuel Poverty in the European Union Investing in Home Renovation and Inclusive Solution" (2014) [4] demonstrates that energy costs are rising much faster than households' income increasing the number of families switching from vulnerability situations to real energy poverty ones. Therefore, energy subsidises and direct financial supports cannot provide a sustainable and *long-term solution* for solving the structural energy poverty causes. It is not in the scope of this work to further analyse the macro economic situation throughout European Union.

3.3 Energy Poverty Health Effects

At this point, the study analyses the consequences of energy poverty over citizens health conditions.

The World Health Organization (WHO) published, in 1987, a milestone study about domestic thermal comfort concluding that the *adequate temperature* range, for people in good health conditions, is between 18 and 24 °C. This non trivial conclusion has been taken as a reference for scientific and regulatory studies on energy poverty in the United Kingdom. In the latter case, it has been concluded that the sensible temperature range has a lower upper limit, being 21 °C for the main dwelling's room and 18 °C for the remaining spaces.

Those standards, although being official and applied in reality, have not been updated for the last twenty-five years [15] and fail to include the health effects caused by *relative humidity index* that have been proved significant in determining domestic comfort levels.

According to Healy (2003) [16] the most serious and direct effect of energy poverty is *winter mortality*. It has been demonstrated that, although winter deaths are not to be exclusively linked to energy poverty, a relevant share has to be attributed to the lack of energy efficiency standards (WHO, 2011). In particular Roberts (2008) [17] has proved that the latter can cause breathing, circulatory, and mental diseases. On the other hand, the ACA (2016) [14] states that excessively high temperature can result in obesity and metabolism issues.

In Southern European countries, it is necessary to refer to the deaths associated with extremely high *summer temperatures*, too. To put this into perspective, Robine (2008) [18] estimates that 70,000 fatalities occurred in 2003, due to the anomalous heat wave of that summer. This proves that, in Mediterranean countries, excess summer mortality should be considered, too.

Excess winter mortality has been reported in medical journals for about 150 years, and most countries suffer from 5% to 30% excess winter mortality. Healy's results conclude that climatic variables such as mean winter environmental temperature and mean winter precipitation are not directly associated with higher levels of relative excess mortality in Europe. The "*paradox of excess winter mortality*" indicates, indeed, that higher mortality rates are generally found in less sever and milder winter climates. This empirical result shows that it is not possible to directly correlate winter deaths with outdoor cold exposures.

Housing standards have been considered as a potential causative factor behind this paradox. This is especially the case in Portugal, Spain, and Ireland. Healy's study demonstrates also a strong correlation between excess winter mortality and overall levels of relative humidity and dwelling's temperature.

Spain is in a particularly harsh position, having a Coefficient of Seasonal Variation in Mortality (CVSM) of 0.21 against a Communitarian (EU-14) mean of 0.16. This coefficient calculates the share of yearly deaths occurring during winter months (December, January, February and March) over total yearly deaths. For Barcelona specific case EPEE (2013) [13] calculates a CVSM of 0.19. Although the latter is not only caused by energy poverty, the WHO suggests that the phenomenon must be considered responsible for 10% to 40% of winter deaths, with a central value of 30%.

ACA (2016) [14] applied this conclusion to the Spanish case, determining that energy poverty caused, in 2013, 7,000 deaths mostly among pensioners. To put this datum into perspective, we can add that in Spain, in the same year, the number of fatalities caused by car accidents was around 2,000. In Catalonia the number of winter deaths, according to the same reference, was 1,200 (second highest in Spain after Andalucía).

The reader should notice that the information given is based on European studies and that are not directly replicable for Spain. Nonetheless, this is a necessary hypothesis since, up to now, there are not specific studies that evaluate the impact of energy poverty over heath conditions at Spanish level.

The effect of energy poverty on health effects has to be linked to public expenditure, too. In the United Kingdom dwellings with low energy efficiency standards cost to the National Health Service (NHS) £760 million per year [14]. One can infer that energy efficiency investments result in major savings in terms of public expense on health services. Moreover, due to the contamination reduction, the gain in population health condition is further improved, reducing, even more, the public spending.

3.4 Energy Poverty Solutions

The purpose here is to describe the policies that have been put in place so far by European politicians and regulators. It is important to specify that, in general, such policies have been characterized by a strong social connotation, based on the fact that the majority of energy poor households are also facing "general" poverty conditions. This will be further demonstrated for the Spanish case in Chapter 5. Due to the inherent lack of *targeting* and *identification* strategy, policy makers are trying to solve the problem merely addressing "general" poverty without considering that there is a significant share of population which is not in monetary poverty conditions, but that is still affected by energy poverty. In all the latter cases, the family income may be above "general" poverty threshold, but, after utilities' bills payment, follows in true poverty conditions. All those families cannot be identified by governments without using specific energy poverty measurement strategies. We are now going to present some "*best practices"* that have been used in different Member States for the last eight years. They can be divided in two phases: *short-term* and *long-term* solutions [8, 13, 19].

As already discussed, the United Kingdom is a pioneer in the energy poverty field having implemented both short-term and long-term solution policies.

The *Winter Fuel Payment* (WFP) is a short-term emergency policy: the government directly pays the bills of the household in vulnerability situations (i.e. with a family member older than 65). The *Cold Weather Payment* (CWP) is on the same line, providing an occasional payment if the temperature decreases below 0 °C. It is addressed to all the households already assisted and supported by social services. The latter two policies are intended to mitigate the *household's income* energy poverty driver.

The *Warm Home Discount* (WHD) targets the *energy price* driver. The latter is, in the British case, highly differentiated from "general" poverty and the eligibility for WHD is supported by an official energy poverty indicator. The policies was implemented in 2011, and require utilities companies to offer discounted and reduced tariffs to reported vulnerable customers. However, such companies can choose their own eligibility criteria which are mostly centred on family members' age and low income groups.

The previous two policies groups are essential to decrease the effects and burdens of energy poverty on identified vulnerable families. However, they do not solve the structural problems that cause an excessive expenditure on energy products. This refers to the third energy poverty driver: energy efficiency.

The *Warm Front Scheme* (WFS) has as objective the improvement of energy efficiency standards of vulnerable households' properties. The government offers financial and technical support for changing heating systems and insulation materials. The policy can achieve high targeting ability and effectiveness due to the British building stocks statistical surveys.

The *Green Deal* (GD) was also introduced for increasing the energy efficiency standards of British homes. It sets minimal energy efficiency label to the renting sector and it introduces a new financial instruments: the Green Deal Finance. The latter allows vulnerable families to afford energy efficiency improvements without all in once payments, but spreading the cost along future energy bills in a controlled manner for a period of 25 years. Moreover, it ensures that the savings would be higher that the costs.

In the French case, energy poverty policies are mostly centred on the reduction of energy prices (TPN and TSS), and on energy efficiency improvements (*Habiter Mieux*). The latter includes a particularly innovative element. In fact, households with major energy expenses are taught and helped not only by installing new efficient equipment, but also with specific courses to change their lifestyle habits.

The German case is particularly relevant as, mostly due to the massive changings in the national energy mix towards renewable sources (*Energiewende)*, the retail price of energy products have increased significantly. Romero (2014) [20] estimates that each German household is contributing to renewables' subsidies for almost €185 per year. The energy poverty issue is not directly addressed as a more general social aids policies has been in place for the last ten years (*Sozialgesetzbuch II*).

Italy introduced in 2005 a *bonus* (*Buono Sociale*) for reducing the impact of utilities (Electricity and Natural Gas) bills over households' financial situation. The program is financed by increasing the fix component of other (non-vulnerable) customers energy bills. This element can be particularly risky as it can result in an overall increase in the number of energy poor households. The policies consider both an income and a contracted power eligibility criterion. The discount, in the case of Electricity, varies between €71 and €153 per year, while, for Natural Gas, varies between €70 and €264 per year (Faiella, 2014) [19]. Faiella also proves that the Precision (i.e. number of real energy poor families among all the households receiving the aid) of such policies is around 20% [19].

The situation in Spain is similar to the Italian case with the implementation (in 2009) of the, so called, *bono social* only for Electricity and not for Natural Gas. The biggest limitation of this policy is its poor targeting ability: a significant share of the households which are eligible for the aid is not experiencing vulnerable conditions. It is estimated (see Annex A) that only the 20% of families that are facing vulnerable situations are also eligible for the *bono social*. We can infer that the current accuracy of the model is extremely poor. Moreover, this policy is only *reactive* the citizens must ask for help and he is not actively detected and helped by the government.

The current Eligibility Criteria (see Figure 5) are:

- A contracted power lower than 3 kW.
- A consumer older than 60 with minimum retirement fees.
- Large families.
- A family with all members in unemployment conditions.

Romero (2014) [20] suggests the inclusion of an income threshold among the *bono social* eligibility criteria and that it should be only financed by public funds and not by private utilities companies.

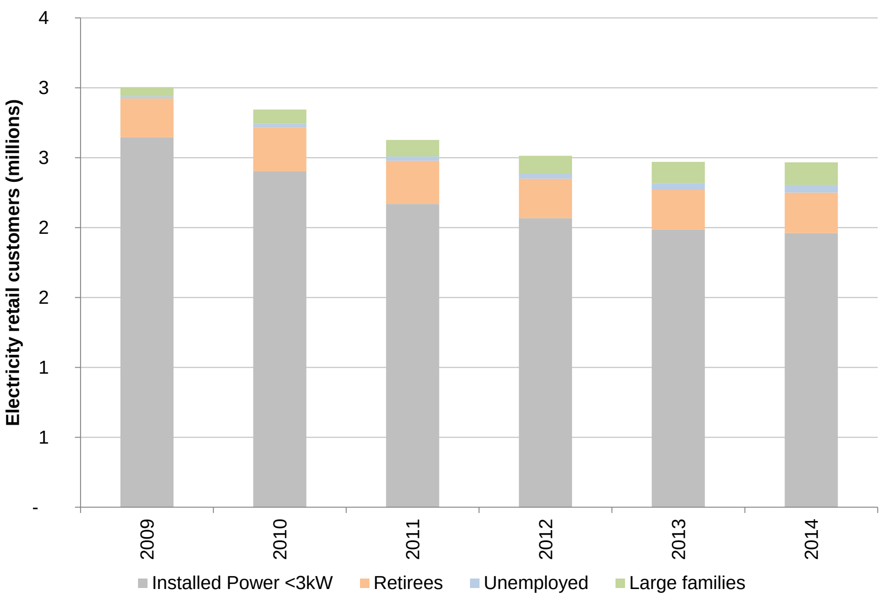


Figure 5 – The assigned Bono Social in the period 2009-2014 divided along Eligibility Criteria categories. [Own Elaboration based on INE data]

Country	Energy Price Policies	Household's Income Policies	Energy Efficiency Policies
United Kingdom	*WHD*	*WFP; CWP*	*WFS; GD*
France	*TPN; TSS*	-	*Habiter Mieux*
Italy	*Buono Sociale* (Electricity and Natural Gas)	-	-
Germany	-	SGB II	-
Spain	*Bono Social* (Electricity)	-	-

Table 1 - Energy Poverty policies implemented in five selected European Member States Own Elaboration]

Table 1 gathers the policies described in Paragraph 3.4, highlighting how the majority of the analysed Member States has focused its attention on policies that mitigate energy products' prices. On the other hand, few countries (United Kingdom and France) have implemented socially driven energy efficiency improvements programs, even though they are the only providing long-term and permanent solutions.

It is obvious that other factors can bring significant and structural improvements to energy poverty, mostly to the household's income component. All those factor can be linked with the macro-economic trends throughout European countries with a specific focus over GDP and unemployment levels. However, in this work, we will consider energy efficiency as the main long-term solution for two main reasons. First, the reduction of energy consumption have a central role in the current and future European strategy towards higher sustainability standards, secondly, the financial benefits (in terms of cost reduction) derived by energy efficiency improvements are pivotal to ameliorate private companies' financial and operative conditions bringing also significant benefits to unemployment and society wealth conditions. Thus, thanks to serious commitments towards higher energy efficiency standards, also the *household's income* energy poverty driver can be addressed.

3.4.1 Energy Affordability vs. Climate Change Policies

The energy poverty solution analysis is also functional for evaluating the contradictory relationship existing between Climate Change and energy affordability policies. In many cases energy sustainability have been claimed to be responsible of the significant increase of energy prices in Europe, for instance, in relation to renewable sources subsidies and carbon prices (i.e. European Trading System). Vorsatz and Herrero (2012) [21] evaluate the existing trade-off or divergent element. The most important is the potential increase in energy poverty levels as a result of strong Climate Change actions increasing energy prices through carbon pricing. If the internalisation of external costs is not totally compensated by enhancement in energy efficiency, the burden of mitigation will mostly affect those worse-off population groups. On the other end, the most significant synergy is offered by improved energy efficiency in buildings. Guaranteeing higher energy efficiency standards is the only option for aligning strong energy poverty alleviation and Climate Change mitigation goals. Short-term support measures implemented as *energy price allowances* or *social tariffs* do not provide a structural solution to the problem, but they divert significant financial resources away from energy efficiency improvements of the residential housing stock.

In conclusion, after the analysis of the policies and the results experienced so far in Europe, it is clear that a serious commitment to fighting energy poverty increasing energy efficiency standards can also bring disruptive improvements to Climate Change reduction.

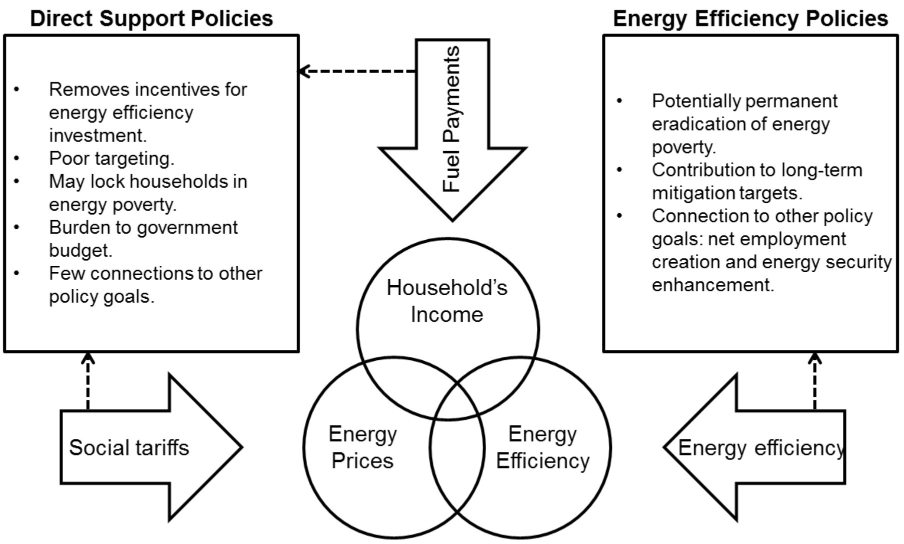


Figure 6 - Synergies and Trade-offs between Climate Change and Energy Poverty policies [21]

4 Measurement Strategies

Chapter 2 has highlighted how important the measurement phase is in the energy poverty solution process. In this work such measurement tools will be called *indicators*.

Choosing suitable indicators (*Targeting*) is the only way to firstly quantify the problem (*Identification*), and, in a second step, to take measures to limit and decrease the phenomenon's impacts over society (*Implementation*).

It is important to specify that exists a wide variety of indicators, and that no one of them has to be taken as a "silver bullet" or as a perfect measurement tool.

The purpose is therefore to analyse the indicators used throughout Europe, identifying, for each one of them, advantages and disadvantages. At the end of this part, it will be possible to choose an eligibility criteria, sorting between social, economic, and energy efficiency focused indicators. This results might be used by policy maker to select a meaningful and fair indicator according to their political visions and strategies.

The majority of them measures the impact of the yearly expense on energy products over households' financial and economic conditions. This is the only approach to understand if there are significant (i.e. beyond a certain threshold) imbalances between utility customer's energy bill and his/her annual income or total expenditures over the same period.

The first problem is that such an approach is based on *real energy expenditures*. Ideally, we should instead compute, case by case, the physical energy demand (e.g. m^3 of Natural Gas needed or kWh electric required per year) that guarantees adequate comfort conditions all year long, according to Paragraph 3.3. In order to follow this optimal estimation approach, one should know dwelling's technical features and the energy equipment used by each family. At national level, the only way to deal with such a complex situation is to statistically collect data about building stock's conditions. These can be, for instance, thermal heat exchange coefficients of the materials used (e.g. walls, windows…), dwellings' surface areas, and heating system's typologies [22].

No Member State possesses such statistics but the United Kingdom [8]. In this country, energy poverty is assessed taking into account a certain expenditure (statistically determined) on energy products necessary for guaranteeing adequate comfort conditions throughout the year. In all other European Countries, the only way to assess energy poverty is to rely on available data, so to say *real energy consumptions*. The latter, in conclusion, will distort the analysis since they do not only focus on customers' needs, but also on their personal preferences. This has to be taken as a necessary and unavoidable first hypothesis.

At this point it is important to provide the reader with an indicators classification framework.

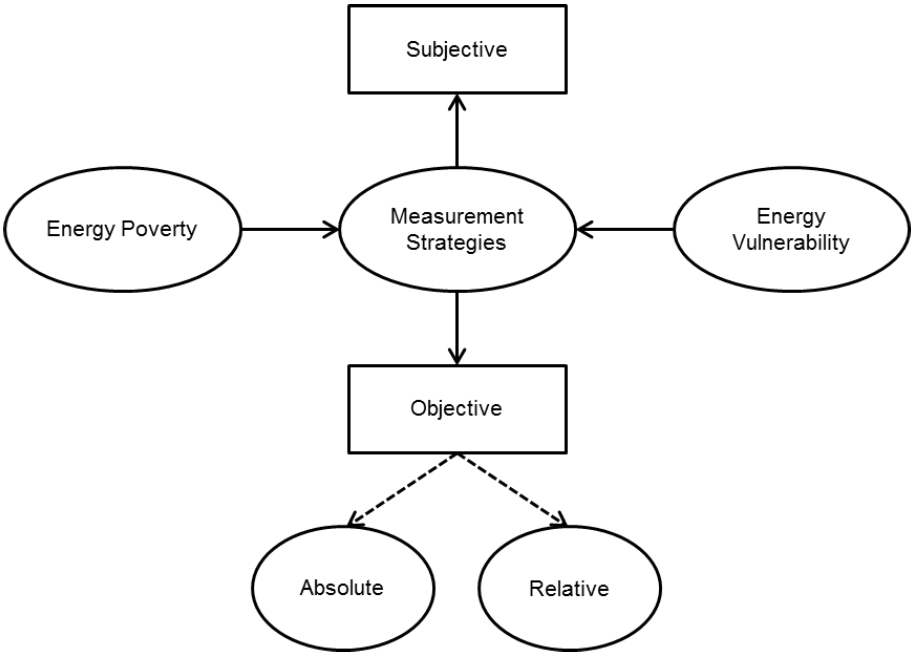


Figure 7 – Energy Poverty Indicators classification strategy [Own Elaboration]

As observed in Figure 7, the measurement strategies can be used to assess two different phenomena that are most of the times confused: energy vulnerability and poverty. Moreover, it is possible to distinguish between subjective and objective measurement strategies, on absolute or relative terms.

4.1.1 Poverty or Vulnerability

Figure 7 shows that the first step to take towards the choice of a suitable measurement strategy is to understand the crucial difference existing between *energy poverty* and *energy vulnerability*. In the study of energy poverty, the vocabulary plays a very important role, even though sometimes the terminology may be misused.

Energy poverty indicates a condition where an household does not have the physical and infrastructural possibility to have access to basic energy services [2]. On the other hand, *energy vulnerability* indicates a condition where the use of energy services brings a family to an unstable economic condition (i.e. "general" poverty). This means that the payment of energy bills subtracts a significant share of one family's financial resources. However, in reality, this terminological distinction, although recognised, is not used and the expression energy poverty has become synonym of energy vulnerability in developed countries.

This study treats the energy affordability issue in developed countries, thus the expression "energy poverty" will be considered as a synonym of vulnerability. Nonetheless, we will demonstrate that also in Spain there are cases of true energy poverty (i.e. lack of access to all energy products, mainly to Natural Gas). The latter element increases considerably the domestic electricity demand inelasticity.

4.1.2 Subjective or Objective

As observed in Figure 7, once defined the object of the study, there are two main measurement strategies: *subjective* and *objective* [19]. The former relates to customers preferences in terms of thermal comfort and housing conditions. The latter, on the contrary, considers only quantitative variables, independently of personal inclinations or habits.

If objective, a strategy can be absolute or relative based.

Absolute indicators use factors that are not influenced by other customers performances or behaviours. They can consist, for instance, of the essential conditions that allow a customer to reach minimum welfare and health levels.

Relative indicators compare the situation of a specific customer with the performances of the whole studied population. In this case, the indicator would be constant with varying energy prices, since such changes would be experienced by the total population and not just by the single household. Relative indicators are thus not sensible to macro variations of the energy price component.

4.1.3 Unit of analysis

Another interesting discussion is whether to consider as analysis unit, the household or the single citizen. In general, it is common practice to use the family as a reference. Considering individual citizens can distort the picture of the problem. Preliminary, in fact, energy vulnerability is more common in few members households [19].We will therefore, from now on, refer to families as unit of reference, since, doing the opposite, would underestimate the impact of energy poverty over certain particularly vulnerable groups. This is shown in Figure 8 where two energy poverty indicators were chosen from the ones that will be analysed later on in this Chapter.

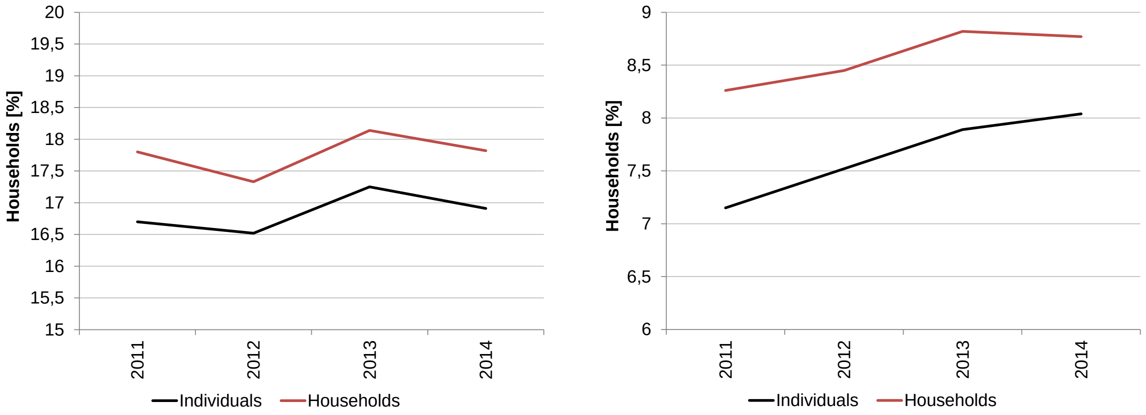


Figure 8 - DM and LIHC energy poverty indicator considering as unit of measurement either Individuals and Household [Own Elaboration based on INE EPF data (2011:2014)]

4.2 Objective Indicators

In the following Paragraph, the analysis will focus on objective indicators either at absolute and relative levels. This category groups all the indicators that consider an *excessive expense on energy services* as the best way for identifying energy poor households. From an engineering perspective this is the most formal and systematic way to tackle the problem.

The expenditure on energy can be normalized by yearly customer's *total expense* or by yearly *total net income*. In general, [19] the choice between the previous two should depend on data availability and quality. Moreover, income and wealth surveys, are affected by misreporting, and it is sometimes controversial choosing between gross or net income.

The following equation defines how the ratio between energy expenses and yearly household's equivalent expenditure is calculated:

$$\rho_i = \sum_{i=1}^{n} \frac{\sum_1^k Expenditure\ on\ energy\ product\ k_i}{Total\ yearly\ expense_i} \cdot 100\ [\%] \quad (4.1)$$

The energy products considered (k) that will be considered in this work are: Electricity, Natural Gas, Liquid Fuel, and Solid Fuels.

A similar version, considering the yearly total household's annual income, instead of total annual equivalent expense, is:

$$\omega_i = \sum_{i=1}^{n} \frac{\sum_1^k Expenditure\ on\ energy\ product\ k_i}{Total\ yearly\ income_i} \cdot 100\ [\%] \quad (4.2)$$

Objective indicators are based on the concept of domestic *energy demand inelasticity*. This means that the demand for energy services is independent of price fluctuations. For this important reason the energy expenditure along all income quintiles will be fairly constant and the share of energy expenses over household's yearly income will be higher for lower income groups. In Figure 9 it is possible to see how the expense on energy has increased between 2006 and 2014, even if both the Electricity and Natural Gas prices have increased (see Paragraph 3.2).

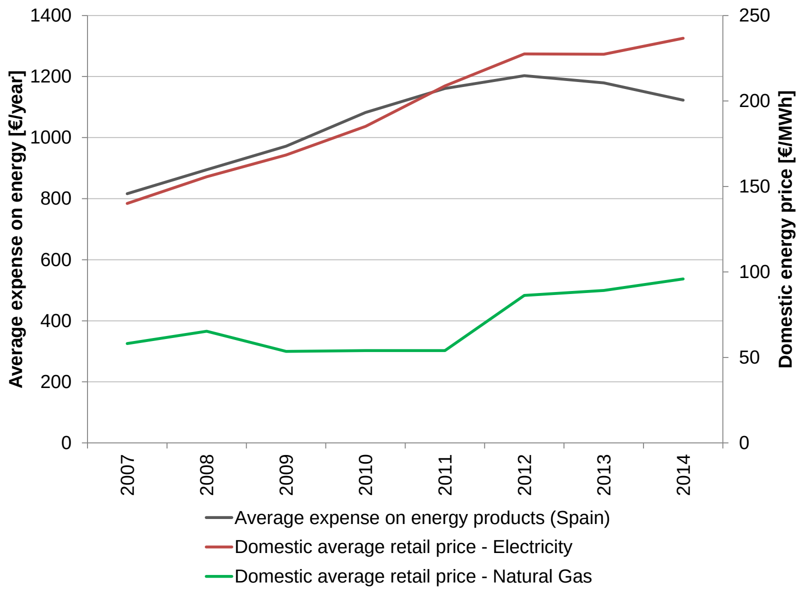


Figure 9 – Household's average expense on energy services in Spain and domestic Electricity and Natural Gas retail prices between 2006 and 2014 [Own Elaboration based on INE EPF data (2006 : 2014)]

Figure 10 shows how the ratio (ω_i) is significantly higher for lower income quintiles. While households in the last decile spent just around 3% of their yearly total expense, poorest families (first tenth) spent more than 11%. It is also possible to appreciate how the ratio ρ_i increased in the period considered. This is mainly due to two different causes: the economic crisis, and the rise of energy prices. The former increases the numerator of Equation 4.1, while the latter decreases the denominator (i.e. impact of economic recession). Both terms combined have increased the value of ρ_i.

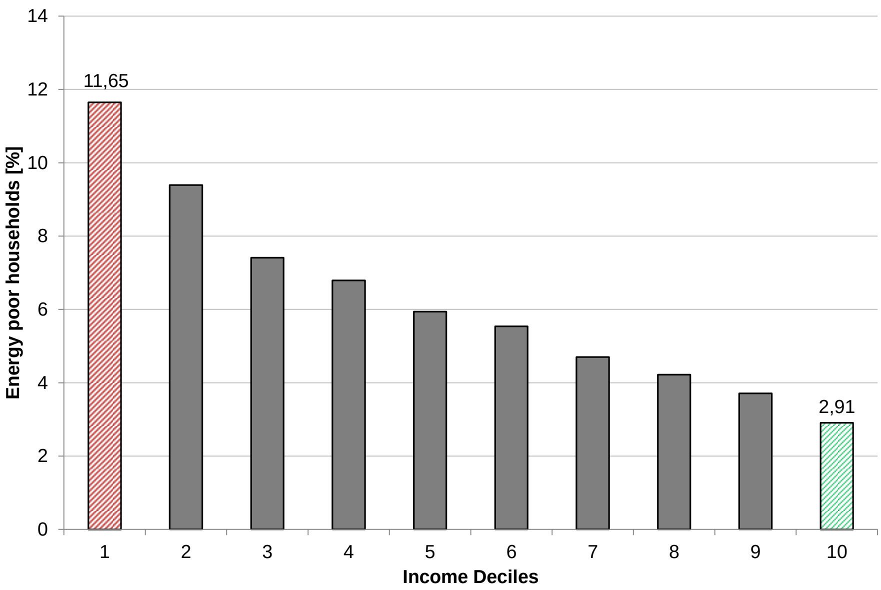


Figure 10 - Energy expense share (ω_i) per income deciles
[Own Elaboration based on INE EPF data (2006,2014)]

In this work, the annual energy expense will be normalized by the annual net household's income. There are two reasons for that. First of all, we want this work to be fully comparable with other studies at European level which are mostly using this normalization strategy. Secondly, using the annual net income will allow us to evaluate the relationship existing between energy poverty and "general" poverty. In fact, according to EUROSTAT, in Europe, the "general" poverty phenomenon must be assessed using income data and not total annual expenditure.

At this point, a threshold can be defined to distinguish energy vulnerable from non-vulnerable households. Thus, the necessary and sufficient condition to be "vulnerable" would be having a ρ_i or ω_i above an *agreed threshold.*

The identification of the latter is not an easy task, as many aspects (social, economic, and technical) must be taken into account and really quantified. The best way, to get to the point is to present some "best practices" that have been proposed in Europe for the last twenty years, to identify which objective indicators can be suitable for our purpose.

4.2.1 Ten Percent Rule (TPR)

This is the oldest indicator for quantifying energy vulnerability and affordability and it was proposed by Boardman in 1991 [15]. It was used in the United Kingdom from 1991 until 2013. Family i should be considered vulnerable if the ω_i ratio is above 0.1 (10%).

$$\gamma_1 = \frac{1}{n}\sum_{i=1}^{n} z_i I(\omega_i \geq 0.1); \tag{4.3}$$

The term z_i takes into account the household's sample weight.

A version similar to the previous can be:

$$\gamma_2 = \frac{1}{n}\sum_{i=1}^{n} z_i I(\rho_i \geq 0.1); \tag{4.4}$$

The indicators are *objective* and *absolute*, in fact, both the ratios are completely independent of other households' expenditure or income conditions. This is mainly due to the fact that the 10% threshold is not obtained through a population analysis but it is simply taken as granted, from the British experience. Originally, the 10% value was taken as it was, in 1989, the double of ρ_i median level of the 30% poorest British households. The empirical nature of the TPR indicator's threshold limits considerably its use and reliability.

4.2.2 Double Median Expense (DM)

It is an objective indicator whose formulation is completely in line with the previous one. Nonetheless, in this case the threshold is obtained by studying the energy expenses patterns of the considered sample. Needless to say that such an approach is highly more flexible and replicable, since the threshold is adapted and modulated upon whole population's features.

The DM indicator is thus *objective* and *relative*, since the logic condition for vulnerability is determined by the performance of the entire population and not just by the single household's:

$$\gamma_3 = \frac{1}{n}\sum_{i=1}^{n} z_i I[\omega_i \geq 2 * (Md_1)]; \tag{4.5}$$

In the equation above, the term Md_1 indicates the median of the vector gathering all the ω_i ratio of the considered sample.

Nevertheless, there is an important limitation in using such an indicator at national or, even more, at European level. This is due to the fact that a unique threshold fails at including the differences existing among climatic zones, household typologies, or dwelling features.

4.2.3 Absolute Measure of energy poverty

Another possibility for measuring energy poverty is defining adequate levels of energy consumption that allow an average family to reach wealthy and socially inclusive living conditions. In other words, the level of energy services necessary to avoid a family experiencing either health problems or social exclusion is taken into account. Simplifying, we consider in the following equation only the expenses on Natural Gas and Electricity, respectively s_i^g and s_i^e.

$$\theta_1 = \frac{1}{n}\sum_{i=1}^{n} z_i I\left[\left(s_i^e + s_i^g\right) < \left(S_i^e + S_i^g\right)\right]; \qquad (4.6)$$

S_i^e and S_i^g are the basic expenditures on Electricity and Natural Gas that allow to reach adequate living standards. The latter should change according to household's typology (i.e. number of members, age, job situation…). Faiella (2014) [19] demonstrates that this approach heavily overestimates the energy poverty impacts, due to the difficulty of choosing and tuning the threshold values S_i^e and S_i^g to different situations. This is caused by the difficulty in modelling the consumption patterns of an entire population. In Italy for instance, the thresholds were calculated considering households with autonomous Natural Gas heating systems and high energy expense shares, without taking into account that many dwellings did not have any kind of heating systems at all. The minimum standard, was, in the Italian case, too high causing the overestimation problem described above.

4.2.4 Low Income, High Cost (LIHC)

This is the newest indicator proposed by Hills (2011) [23] [15] and adopted by the United Kingdom in 2013. Hills considered that all previous indicators, by using a single threshold, were also considering, as energy poor, families without any kind of financial issues. They were indeed just experiencing an anomalous ratio of energy expenses, without falling into real deprivation situations (i.e. monetary poverty).

To avoid this issue, the LIHC indicator uses *two thresholds*. The two conditions are:

- An energy expense that exceeds the national median level (Md_2).
- A residual income (i.e. total income minus energy expense) that makes the household fall into general poverty conditions, according to EUROSTAT methodology.

It is necessary to clarify the second condition: EUROSTAT states that, at European level, an household should be considered at "general" poverty risk if the total annual income is below the 60% of the national median level (Md_T).

Thus:

$$\pi_1 = \frac{1}{n} \sum_{i=1}^{n} z_i \{ I[s_i^{tot} > Md_2] \cdot I[(y_i^{tot} - s_i^{tot}) < Md_T]\}; \tag{4.7}$$

In Equation 4.7 y_i^{tot} indicates the household's total net annual income. The LIHC indicator is a *twin indicator* consisting of:

- The number of households that have both low income and high fuel costs (shown by the shaded area in Figure 11); and
- The depth of fuel poverty amongst these fuel poor households. This measures the energy poverty gap which represents the difference between the median national energy cost and the actual family expense (shown by the red arrow in Figure 11).

The depth allows to distinguish households according to their energy poverty severity degree.

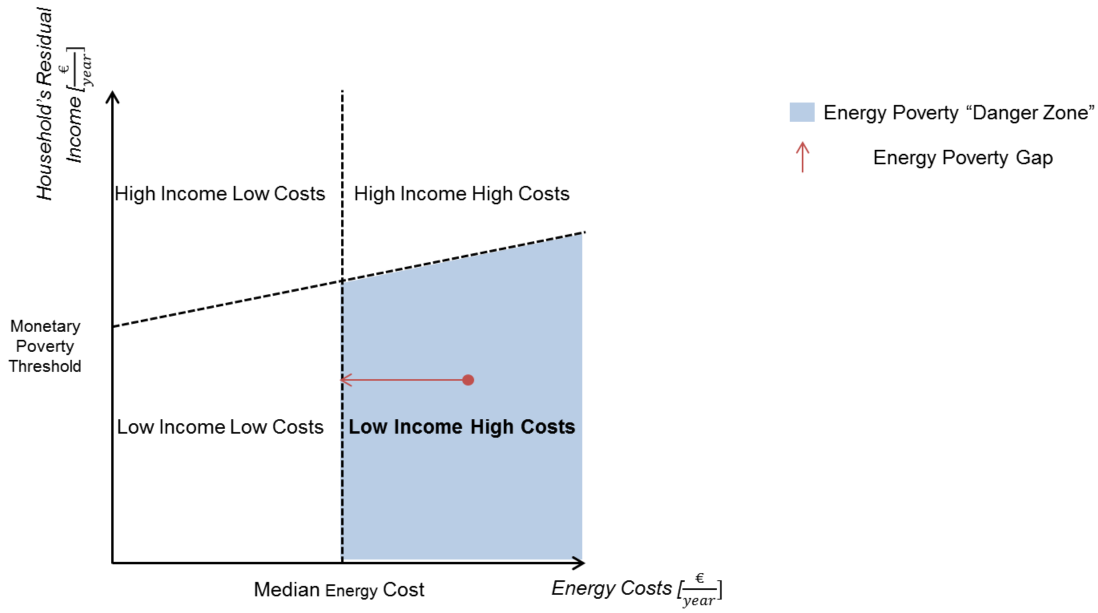


Figure 11 - Energy Poverty Low Income, High Costs indicator. The plot displays the four quadrants of interest, highlighting in light blue the energy poverty "danger zone" [Own Elaboration]

The energy poor quadrant also includes some households who are pushed into fuel poverty by their very high energy requirements, caused, for instance, by the average age of family members or by the number of children. The latter, according to Paragraph 3.3, have higher energy needs to guarantee healthy living conditions. This is reflected in the gradient of the income threshold. The energy poverty depth concept is particularly significant in the British context, where the statistical surveys allow to compute the household's *required energy expenditure*. In England the latter is modelled using data from the English Housing Survey (EHS) which is an annual national survey of citizens' housing and tenures, involving physical inspection of properties by professional surveyors, to determine energy efficiency status and general housing conditions. As said, such statistics are not available in other Member States compromising the applicability of the energy poverty depth.

4.2.5 Minimum Income Standard (MIS)

This indicator is also recent and born in the United Kingdom. It refers to a *minimum income level* that makes a family fully and actively integrated in society. Moore (2012) [24] demonstrates how this can be a precise and efficient instrument to assess energy vulnerability across Europe. However, it is extremely difficult to define what income level should be considered as the adequate minimum. In the United Kingdom, to solve the problem, the government defined a list of products whose purchasing is considered essential, according to household's typology and magnitude (i.e. number of family members). Based on that, a minimum income was determined (MIS) and applied in the following equation:

$$\delta_1 = \frac{1}{n}\sum_{i=1}^{n} z_i I\left[\left(y_i^{tot} - s_i^{tot}\right) > MIS_i\right]; \tag{4.8}$$

The approach of this indicator is similar to the one used for θ_1, in fact both define a minimum standard, being either a physical or a financial quantity, and then compare this with single household's performances.

The MIS standard can be constant or variable. We will see, for the Spanish case, how different standards exist among autonomous regions, and how the choice of a specific one can heavily influence the assessment of energy poverty.

4.3 Households with zero energy bills

All previous indicators, apart from θ_1, do not consider as fuel poor those households that have expenses on energy close to zero. This is a major limitation. The Spanish studies and reports analysed in this work, ACA (2016) and Romero (2014) [14, 20], do not take into consideration this vulnerability aspect limiting their studies to the indicators described so far. This work wants to fill this gap using the considerations and studies proposed by Faiella in 2014 [19].

It has been demonstrated [8] that the lack of heating system, at European level, can be considered an indicator of deprivation, as it is more common within lowest income deciles. We are going to further elaborate this concept for the Spanish case in Chapter 5.

In the following part we will therefore introduce three more indicators, that will add to DM, LIHC, and MIS those families with very low expenses on energy products (i.e. close to zero) or without any kind of heating system.

$$\gamma_4 = \frac{1}{n}\sum_{i=1}^{n} z_i \left\{ \begin{matrix} I[\omega_i \geq 2 * (Md_1)] \\ \cup \\ I[H = 0] \\ \cup \\ I\left[s_i^{tot} < t * (Md_2)\right] \end{matrix} \right\}; \tag{4.9}$$

$$\pi_2 = \frac{1}{n}\sum_{i=1}^{n} z_i \left\{ \begin{matrix} I\left[s_i^{tot} > Md_2\right] \cdot I\left[\left(y_i^{tot} - s_i^{tot}\right) < Md_T\right] \\ \cup \\ I[H = 0] \\ \cup \\ I\left[s_i^{tot} < t * (Md_2)\right] \end{matrix} \right\}; \tag{4.10}$$

$$\delta_2 = \frac{1}{n}\sum_{i=1}^{n} z_i \left\{ \begin{matrix} I\left[\left(y_i^{tot} - s_i^{tot}\right) > MIS_i\right] \\ \cup \\ I[H = 0] \\ \cup \\ I\left[s_i^{tot} < t * (Md_2)\right] \end{matrix} \right\}; \tag{4.11}$$

In previous equations the logical condition $I[H = 0]$ indicates the lack of heating system, and Md_2 is the annual national median expense on energy services. It is important to stress the fact that the latter is different from Md_1, being the national median share of expenditure on energy services (see Equation 4.2).

A further threshold has been introduced, adding a *lower limit*, that can be tuned changing the parameter t between 0 and 1. The latter has to be changed for each specific application to justify the fact that a low expenditure on energy products must be strictly correlated to household's deprivation conditions. Up to now, in fact, it has not been demonstrated which is the minimal energy expenditure that provides *adequate* health and living condition, as it is done in the British case. This limitation is due to the first hypothesis made, linked to the structural lack of data about national building stocks.

An household should therefore be considered energy vulnerable, not only if it has excessive expense, but also if its consumption is *significantly lower* than other households' experiencing similar boundary conditions: weather seasonal patterns, energy prices, comfort requirements among others.

4.3.1 Energy Efficiency Indicator (EEI)

This is an energy vulnerability indicator used for the first time in this work, with no other applications in Europe, so far. It has been elaborated taking into account the data availability of the "*Encuesta de Presupuestos Familiares*" by the Spanish Statistical Institute (INE). The purpose of the indicator is to achieve high insight on the *energy efficiency component* of energy poverty following the LIHC approach and methodology.

In Energy Engineering, the assessment of dwellings' energy efficiency levels is done evaluating energy needs per surface unit (m^2). For this reason it is possible to design an indicator based on the LIHC, normalizing the annual energy expenditure by the dwelling's surface.

As π_2, the *EEI* indicator measures an excessive energy consumption (beyond national median level) and household's financial conditions after utilities' bills payment. The possibility of including the surface component is crucial in identifying the nature of the excessive energy expense reported by indicator π_2. Thus, two are the differences in using *EEI* rather than π_2.

The first is the exclusion of those households with major energy expenses caused by high dwelling's surface, and the second is the inclusion of families with not excessive expenses (i.e. not beyond the national median) in absolute terms, but whose expenditure per m^2 is still considerable, due to poor energy efficiency standards.

The *EEI* is further improved by adding a lower threshold, as for γ_4, π_2, and δ_2.

Thus:

$$EEI = \frac{1}{n}\sum_{i=1}^{n} z_i \begin{Bmatrix} I[k_i > Md_3] \\ \cup \\ I[H = 0] \\ \cup \\ I\left[s_i^{tot} < t * (Md_2)\right] \end{Bmatrix}; \tag{4.12}$$

where k_i is the ratio between household i energy expense and its dwelling's surface. Md_3 is median of the vector containing all k_i.

The indicator focuses on what we will identify to be the major driver for energy poverty in Atlantic and Northern Spanish regions: energy efficiency (see Paragraph 5.2).

4.4 Subjective Indicators

The three following indicators were agreed at European level (EU-SILC) and the data were collected and processed by national statistical institutes (in the Spanish case by INE). The aim was to quantify as precisely as possible, the perception of thermal comfort and the ability of affording adequate energy services across Europe.

A sample of European families, divided among all Member States, were asked to answer three basic questions, taking therefore into account only their subjective and personal opinion and perception of the problem.

They are currently the only indicators available to use and compare the status of energy poverty across the European Union, and therefore, despite their weaknesses, provide an important basis for comparison.

The three SILC subjective indicators are:

- ε_1: assesses the household's ability of keeping an adequate temperature throughout the year.
- ε_2: assesses the household's delay in the energy bill payments (i.e. arrears).
- ε_3: assesses if the household is living in a dwelling with structural deficiencies or in poor energy efficiency conditions.

4.5 Indicators' Characteristics

Up to now, thirteen different energy vulnerability indicators have been identified. The previous analysis represents a good starting point for benchmarking and identifying advantages and disadvantages of each one of them. In order to do that, it is useful to identify which are the properties [19] that make an indicator a suitable and a meaningful measurement tool for supporting policy makers decisions.

The energy poverty and vulnerability issue has particular effects over social and health household's conditions. The key aspect in the analysis is to guarantee that the reported families, really face deprivation conditions. It can happen, in fact, that an indicator reports as vulnerable, households that are not really facing stringent financial and social situations, but just major unbalances in their ω ratio. Of course, it can be interesting to address also the latter problem to identify potential savings and efficiency improvements, but it is important to remember the social core of the energy poverty issue. The best way to do this, is to study energy vulnerability as a consequence or rather a "branch" of general poverty, therefore focusing on lowest income deciles at highest risk of social and monetary exclusion. The latter indicator's property is called *indicator targeting ability*. Figure 12 illustrates this aspect for three indicators' families (DM, LIHC, and MIS). The reader can notice how the DM indicator (black line) reports as energy poor a significant portion of higher income citizens. In such cases, we can infer that there might be significant energy expenses caused by higher living standards or by huge dwellings.

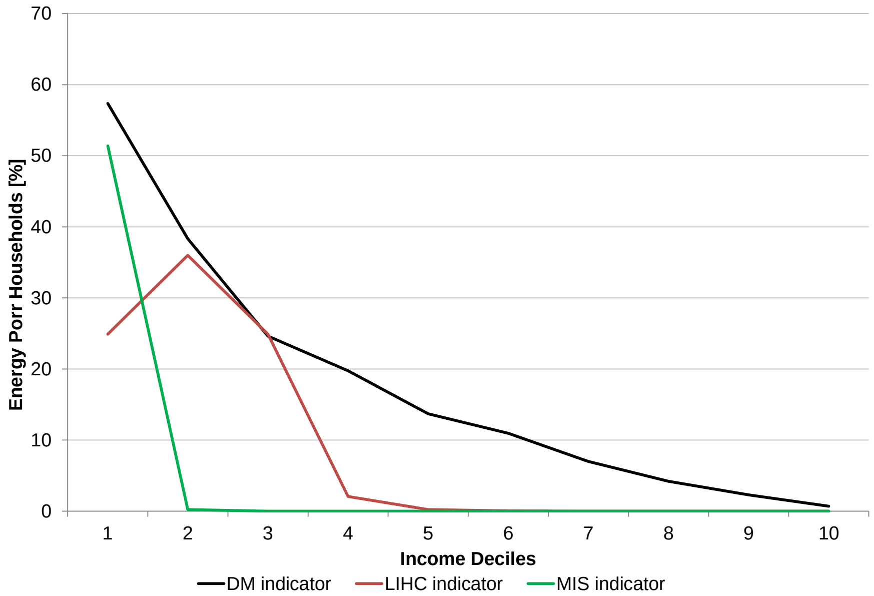


Figure 12 – Example (from the Spanish study case) of DM, LIHC and MIS indicators results along different income deciles [Own Elaboration based on INE EPF data (2014)]

If a policy maker recognizes the need of focusing on social inclusion, facing inequality, it would be advisable to avoid using DM indicators as the black line in the picture is only converging to zero for the last income decile. In Chapter 5 we will consider the indicator's targeting ability in further detail, focusing on the Spanish case.

The energy vulnerability issue involves many and diverse aspects related to energy efficiency, social and financial aspects. It is important for an indicator not to focus only on one of them, but to be flexible and heterogeneous in describing and evaluating the phenomenon. Due to data availability reasons, it is common to take into account mostly economic considerations. However, we have demonstrated that some of the most recent indicators are also including a more energy efficiency focused approach (e.g. LIHC). Having high *heterogeneity* is useful not just to properly quantify the problem, but also to understand which are the first measures to be taken, increasing the chances of solving the problem.

Data availability and statistical concerns are also central in evaluating energy vulnerability. The only way to assess both causes and impacts, is to have an high quality and updated database. For this reason, an indicator has to use data that are fully *available*, *reliable*, and *chronologically updated*. The latter is not significant, at European level, as all the statistical information needed are fully available, however, on a case by case basis, this might not be true in some specific applications.

The last interesting feature is *communicative efficiency*. An indicator should not be too difficult to understand and interpret including the right number of thresholds and logic conditions necessary to fully understand the "physical" laws lying behind each indicator.

Energy Poverty Indicator	Indicator's symbol	Indicator Targeting Ability	Indicator information heterogenity	Data availability and quality	Chronological data availability *	International comparability	Communicative efficiency	Total **
Ten Percent Rule	γ_1	⇩	⇔	⇩	⇧	⇩	⇧	⇩
Double Median	γ_3	⇔	⇔	⇧	⇧	⇩	⇧	⇧⇧
Absolute Measure	θ_1	⇔	⇧	⇩	⇧	⇩	⇧	⇧
Low Income, High Cost	π_1	⇧	⇧	⇧	⇧	⇩	⇔	⇧⇧⇧
Minimum Income Standards	δ_1	⇔	⇔	⇧	⇧	⇩	⇧	⇧⇧
Subjective Indicator	ε_1	⇩	⇩	⇩	⇧	⇧	⇧	⇔
Subjective Indicator	ε_2	⇩	⇩	⇩	⇧	⇧	⇧	⇔
Subjective Indicator	ε_3	⇩	⇩	⇩	⇧	⇧	⇧	⇔
Double Median + ZEBH	γ_4	⇧	⇔	⇧	⇧	⇩	⇧	⇧⇧⇧
Low Income, High Cost + ZEBH	π_2	⇧⇧	⇧	⇧	⇧	⇩	⇔	⇧⇧⇧⇧
Minimum Income Standards + ZEBH	δ_2	⇧	⇔	⇧	⇧	⇩	⇧	⇧⇧⇧
Energy Efficiency Indicator	EEI	⇧⇧	⇧	⇧	⇧	⇩	⇔	⇧⇧⇧⇧

ZEB = zero energy bill households

*Table 2 - Main indicators evaluation table. * Availability of data for a time period at least spanning from 2006 to 2014. **The total is computed as the algebraic sum of the factors considered for each energy vulnerability indicator, considering the following scoring system: ⇧=1; ⇔=0; ⇩=-1 [Own Elaboration based on [19]]*

ETSEIB

everis
an NTT DATA Company

Table 2 shows a *purely qualitative* rating methodology for pointing out each indicator's advantages and disadvantages. The evaluation strategy is based on the work of Faiella (2014) [19] and support the energy poverty indicators' proposal in the Italian case. The table resume, the considerations made so far, gathering the points made in the studies and researches considered [8, 13, 19]. Although just qualitative, this analysis can be useful for policy makers to select indicators which can be suitable and useful for a specific purpose or application. The total column shows the final grades of all the thirteen indicators considered in the study.

First of all, it is possible to notice that the three subjective indicators, have exactly same ratings. Moreover, they are the best according to communicative efficiency. It is not a surprise, as they were chosen as main measurement indicators at European level (EU-SILC). We can consider them as neutral in our study. However, the ε_3 indicator can be important to assess the housing stocks conditions across different European countries, even though not in an objective way. They will be used to position Spain within the European context.

The best indicator, as already anticipated, among the γ family, is γ_4. This is due to its higher flexibility in adapting to each country specific conditions.

π_2 and δ_2, corresponding to LIHC and MIS indicator families, show high features. They perform well according to all factors, apart from communicative efficiency and international comparability. The latter is not unexpected: the thresholds are in fact tuned on specific national conditions, and are not the same throughout Europe. Due to the triple threshold characteristic, the indicator is more difficult to understand, and it is also tougher for households to determine their vulnerability status.

The best scores are achieved by indicators that include a lower threshold, considering families without heating system and with zero energy bills, too. It is important to remember that, until now, in Europe such indicators have not been used yet.

In conclusion, we can consider four indicators particularly suitable for our scopes: γ_4, δ_2, π_2, and *EEI*. For historical reasons, as first energy poverty indicator, we will also take into account the TPR, mainly as a comparison instrument.

Since remembering the properties of those five indicators might be hard, Table 3 summarizes all the considerations made, providing the basic tools to start the energy poverty evaluation phase in Chapter 5.

Indicator	Family	Description
γ_4	DM	Double Median indicator that includes both an upper and a lower threshold. A family is considered vulnerable if its energy expense ratio (ω) is above the median, OR if its expense on energy is lower than a certain percentage (t) of the national median value.
δ_2	MIS	A family is considered vulnerable if its residual income (net income minus energy expenditure) is below a certain agreed level (Minimum Income Standard),OR if its expense on energy is lower than a certain percentage (t) of the national median value.
π_2	LIHC	A family is considered vulnerable if it is energy expense is above the median AND its residual income is below the 60% of the national median ("general" poverty condition), OR if its expense on energy is lower than a certain percentage (t) of the national median value.
EEI	LIHC	A family is considered vulnerable if it is energy expense per unit area is above the median AND its residual income is below the 60% of the national median ("general" poverty condition), OR if its expense on energy is lower than a certain percentage (t) of the national median value.
γ_1	TPR	It is just considered for historical reasons. It evaluates if the family's energy expense ratio (ω) is above 10%.

Table 3 - Summary table reporting the main features of the five chosen indicators [Own Elaboration]

5 Energy Poverty Evaluation

The purpose of this Chapter is to apply the conclusions drawn until now, regarding energy vulnerability indicators, to quantify the phenomenon at European, Spanish and Catalan level. Moreover, a unique suitable indicator will be chosen to evaluate the issue across Spanish regions. The data considered in this Chapter were taken from EUROSTAT [3] and from INE [25] and they corresponded to the most updated statistical series available when this work was written (Spring 2016).

5.1 The European Situation

Subjective indicators, as explained in Chapter 4, are the best according to communicative efficiency and international comparability. The EU-SILC survey by EUROSTAT [3] consists of three qualitative questions, that were answered by sampled European households. Since the latter ones are exactly the same in each Member State, it is possible to draw comprehensive measurements at European level. The following conclusions will be only qualitative but, as already anticipated, ε_3 can provide useful information about Member States housing stocks conditions.

The analysis starts with ε_1. It measures household's ability of keeping an *adequate temperature* in its dwelling. The results are proposed both as percentage of affected population and as absolute number of people facing the problem. In this case, the word *adequate* is completely aleatory and purely dependent on personal preferences and it is not linked to the World Health Organization proposals (see Paragraph 3.3).

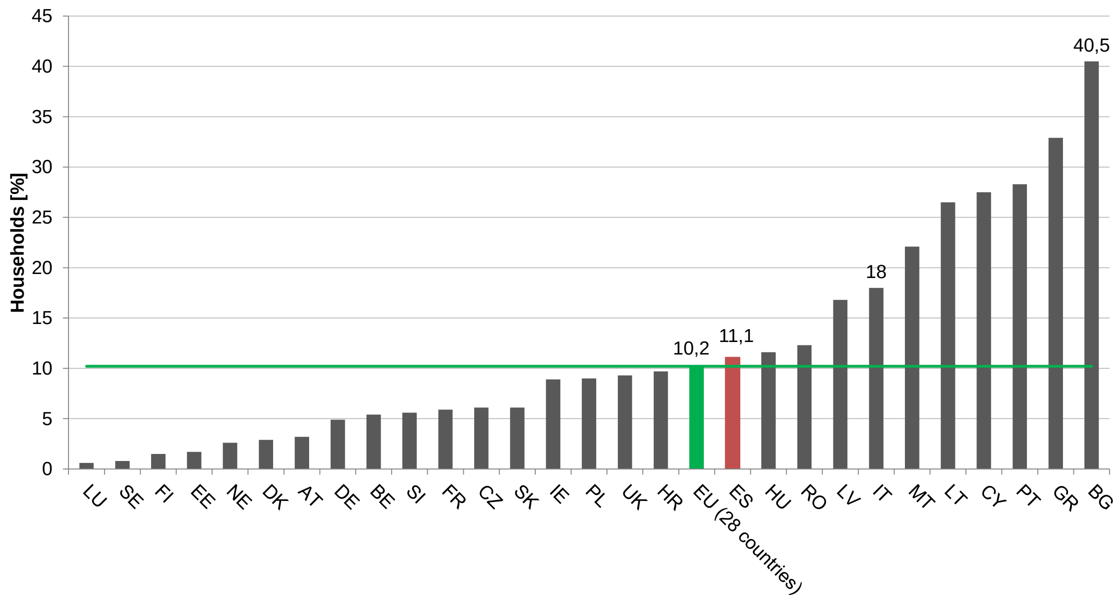


Figure 13 - ε_1 indicator (2014). The line in green represent the European average, while the red column represent the Spanish study case. [Own Elaboration based on EUROSTAT data].

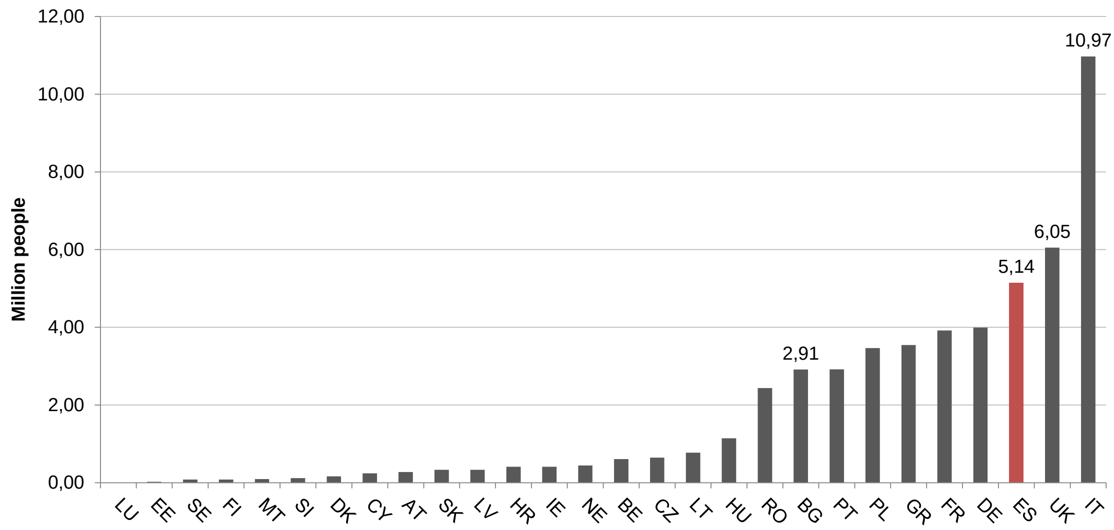


Figure 14 - ε_1 indicator in absolute terms (2014). The red column represents the Spanish study case. [Own Elaboration based on EUROSTAT data].

ε_1 indicator is arguably considered, at European level, the most relevant and truthful measurement tool [8]. For this reason we will spend some words analysing its outcomes, even though Table 2 and our indicator analysis has demonstrated that such conclusions cannot be considered reliable at all.

Figure 13 shows that Bulgaria, Greece, and Portugal present significant levels of inability of keeping adequate and healthy home conditions. Spain presents a level slightly above the

European average (green line). Figure 14, on the other hand, shows absolute values, as millions of people unable to fulfil basic thermal requirements throughout the year. Since the indicator takes into account Member States' population, Italy, United Kingdom, and Spain are now showing the worst performances. Interestingly, Southern European countries have higher values. This is known as the *energy poverty paradox* and it is due to three major causes.

First, generally speaking, housing stocks features are often not adequate, and the lack of insulation makes heating very difficult and expensive, in winter periods. Second, heating systems are often (i.e. 20 to 30% of cases according to INSIGHT-E (2015) [8]) inexistent or have low thermal efficiencies (i.e. lower number of installed central heating systems). Third, financial conditions in Mediterranean countries were more heavily struck by the economic crisis than Central and Northern Europe ones.

ε_2 indicator, in Figure 15, gives a measure of energy affordability across European countries. It shows the share of population with arrears on utility bills. This indicator, according to INSIGHT-E (2015) [8], is highly related to Member States economic situations. Spain is below European average, showing that there are not significant delays in utilities bills payment. This element might appear as a positive proof of good financial stability, but this is absolutely not the case, in fact, due to the low degree of vulnerable customers' protection in Spain the number of forced disconnections has been increasing since 2011, limiting to a minimum the number of arrears on utility bills.

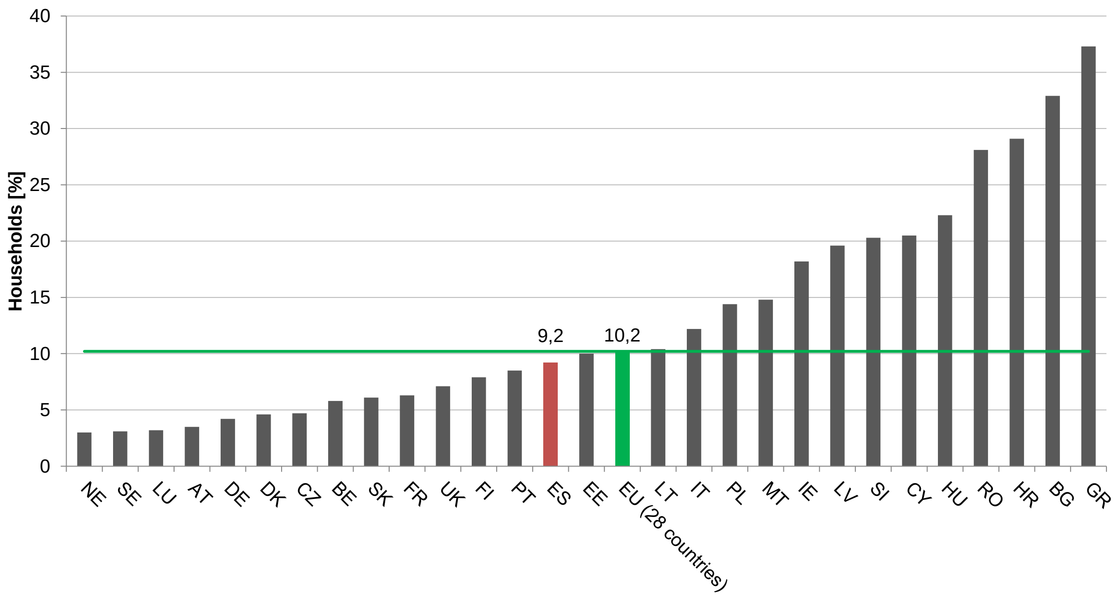


Figure 15 - ε_2 indicator (2014) [Own Elaboration based on EUROSTAT data]. The green line indicates the European average, while the red column represents the Spanish study case

ε_3 indicator gives insight, although just in a qualitative way, over European housing stock conditions. It is not directly possible to draw conclusions on energy efficiency levels, since the indicator considers only the presence of leakages, damping walls and rotten windows.

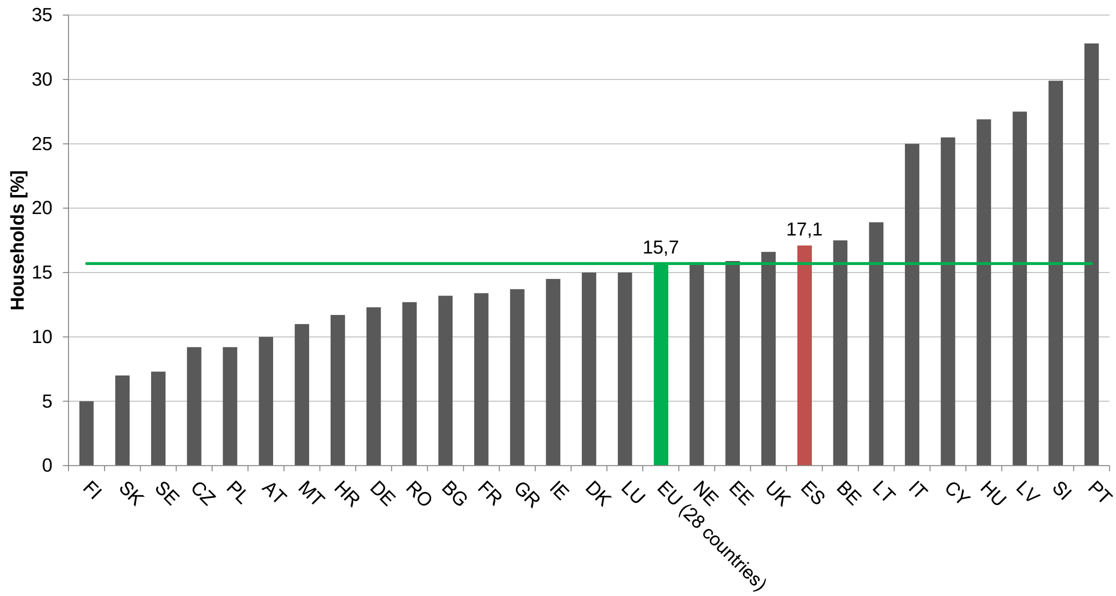


Figure 16 - ε_3 indicator (2014) [Own Elaboration based on EUROSTAT data]. The green line indicates the European average, while the red column represents the Spanish study case

Figure 16 shows that, according to ε_3 indicator, dwellings in Spain have poorer building standards than the European average. This is also in line with what we said analysing indicator ε_1, in fact, Mediterranean countries, like Italy and Spain, are well above Central Europe countries' ε_3 values. This is of course due to milder climates, but as demonstrated in Figure 13, results in lower thermal comfort levels and in major health risks. Considering absolute population, indicator ε_3, estimates that, in 2014, almost 8 million Spanish were living in poor housing conditions (from energy poverty point of view).

European surveys based on subjective indicators, namely ε_1, ε_2, and ε_3, do not consider Spain in a particularly dangerous situation. However, we can conclude that in 2014, 5.14 million Spanish citizens were unable, due to technical and/or economic reasons, to maintain an adequate temperature in their dwellings. Another important feedback, given by ε_3, is that the estimation of energy poverty in Spain cannot be carried out without considering the energy efficiency component. In fact, the 17.1% of the Spanish population is living in dwellings with unhealthy housing conditions, including, for instance mould, drafts, and excessive humidity. It is also possible to demonstrate that the value of ε_1 has been steadily increasing in the last six years, starting, and it is not a coincidence, from 2008. The value of ε_1, for Spain, reached and passed the European average level in 2013.

According to this, it appears evident that the energy poverty issue in Spain has become more serious in the last eight years.

Thomson and Snell (2013) [26] defined four aggregate indicators based on the three proxy ones used by the EU-SILC. In order to do that, different weights (between 0 and 1) were assigned to ε_1, ε_2, and ε_3. Since, we have showed that poor housing stock conditions are, for Spain, above the communitarian average level and appear as main energy poverty drivers, we will show the results for what Thomson and Snell (2013) defined as *Scenario 3*. A weight of 0.5 is assigned to ε_3, and a weight of 0.25 is assigned to both ε_1 and ε_2.

Thus, the Thomson-Snell indicator would be:

$$TS = \sum_{i=1}^{c} 0.25 \cdot (\varepsilon_1 + \varepsilon_2) + 0.5 \cdot \varepsilon_3 \qquad (5.1)$$

Where c represents the number of considered Member States.

The use of this indicator will allow us to define a European ranking where we will notice how Spain is positioned, compared to other Member States.

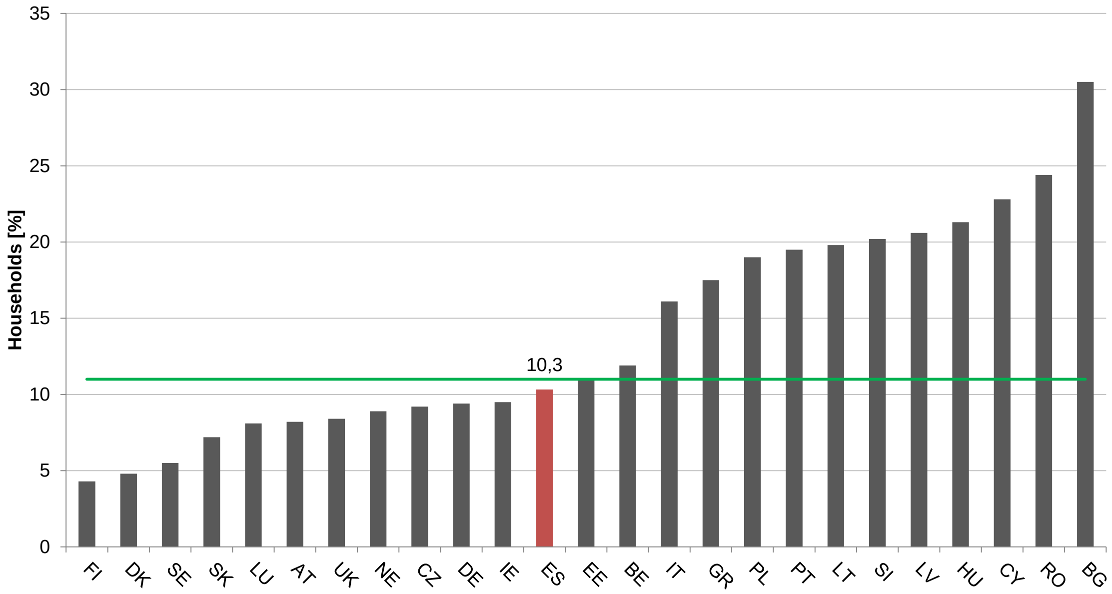


Figure 17 - Thomson-Snell indicator for Scenario 3 (2014) [Own Elaboration based on EUROSTAT data]. The red column represents the Spanish study case, while the green line the European average

The Spanish situation, evaluated by subjective (proxy) indicators, does not appear so serious in comparison with other European countries. According to the *TS* indicator the 10.3% of Spanish population is vulnerable in energy poverty terms.

5.2 The Spanish Situation

The assessment of energy poverty with objective indicators will be based on the statistical experimental evaluation of the disaggregated dataset: *Encuesta de Presupuestos Familiares* by the Spanish Statistical Institute (INE) [25].

The aim of the following part is to provide energy vulnerability measurements for Spain and, in particular, for Catalonia. At the end of the Chapter the reader will be able to estimates which are the Spanish regions with highest vulnerability levels, and to have a more precise understanding of what are the phenomenon's drivers case by case.

The first thing to do is selecting, among all the treated indicators, one that can be precisely and coherently applied to the Spanish case.

The results, according to Chapter 4 conclusions, will be given per households and related to family's annual net income. The analysis will start considering the selected indicators: $\gamma_1, \gamma_4, \pi_2, \delta_2$, and *EEI* (see Table 3).

First of all, it is important to verify whether the analysis should merely focus on energy vulnerability or also on the structural and infrastructural lack of heating system. According to EPF, the 34.3% of Spanish households did not have any kind of heating system in 2014. In the Catalan case, the number decreases to 22.3%. In all these cases we would rather use the term *energy poverty* and not vulnerability, for indicating that the problem is caused by the lack of structural access to certain energy products (mostly Natural Gas). This factor has major impact over domestic *demand inelasticity.*

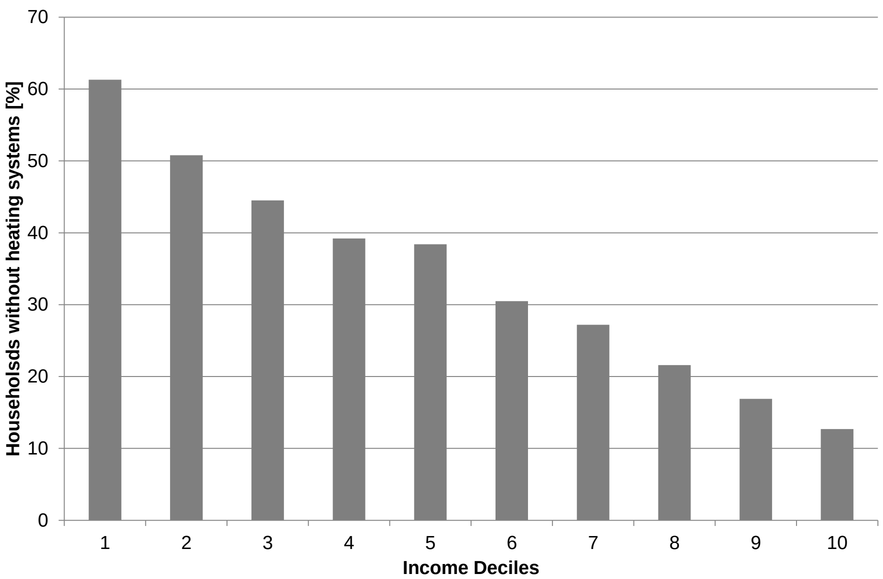


Figure 18 - Distribution per income deciles of households without any kind of heating system (2014) [Own Elaboration based on INE EPF Base 2006 data]

Figure 18 demonstrates that the lack of heating system is more common in lowest income deciles, even if, among all households without heating system, just the 45.5% belongs to the first three income deciles, that indicate monetary poverty (according to EUROSTAT). It is therefore not possible to directly conclude, for the Spanish case, that the lack of heating system indicates a deprivation situation (i.e. “general” poverty). This is particularly significant for some of the defined indicators, as including a specific logic condition related to that.

It must be tested whether what has just been discussed can decrease the *Targeting Ability* of indicators γ_4, π_2, δ_2, and *EEI* (see Table 2). We should in fact verify that the lack of heating system is not reporting too many households from higher income deciles as energy vulnerable. For the sake of brevity only the results for π_2 are presented.

Figure 19 shows how π_2 behaves for different t (see Equation 4.9, 4.10, and 4.11). These represent specific percentage levels of the median energy expense level to adjust the lower indicators’ threshold. The dotted line indicates the “general” poverty limit. The efficiency of an energy poverty indicator measures its ability to cover this “danger zone” area not including families belonging to higher income deciles.

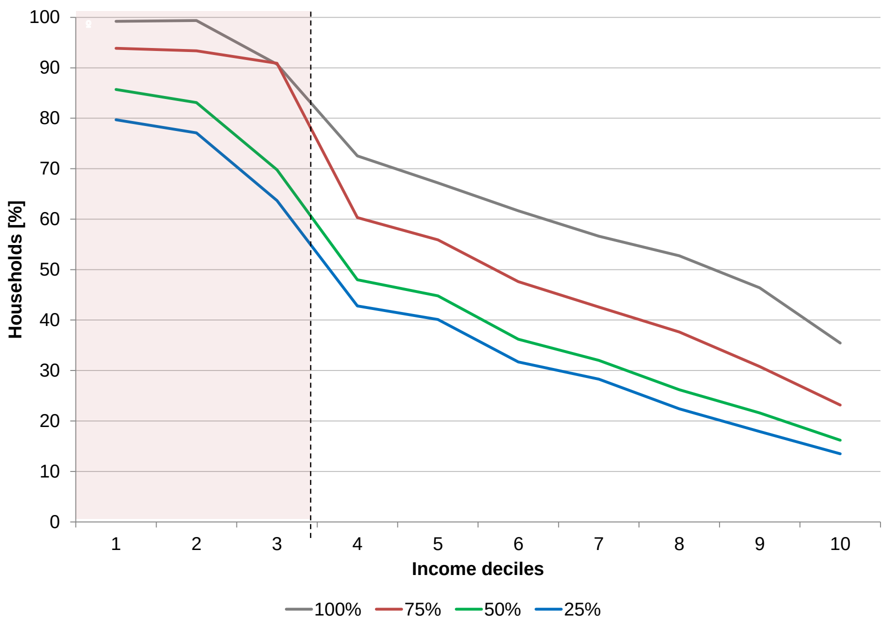


Figure 19 - π_2 indicator results for different income deciles (2014) [Own Elaboration based on INE EPF Base 2006 data]. The red shaded area indicates the incomes affected by monetary poverty

One can conclude that the indicator, including families without heating system, is not efficient at all. If t is chosen equal to 25% of the national median expense on energy services, the indicator would consider as energy vulnerable the 15% of families of the tenth income decile (i.e. with an annual net income between €41,388 and €199,500).

This will result in distorted measurements, and, above all, in policies that will also help families that are absolutely not in stringent financial situations. It is important to remember, once again, that the main aim of the study is to find methodologies for allocating public authorities resources in the fairest and most efficient way.

For this reason we will exclude from γ_4, π_2, and δ_2 the condition of not having any kind of heating system. From now on, we will therefore speak about *adjusted indicators*, to include the aforementioned adaptation. The latter ones will only have two logic conditions.

The adjusted π_2 indicator efficiency performances are showed in Figure 20. We can clearly notice that the blue line drops almost to zero after having passed the general poverty threshold.

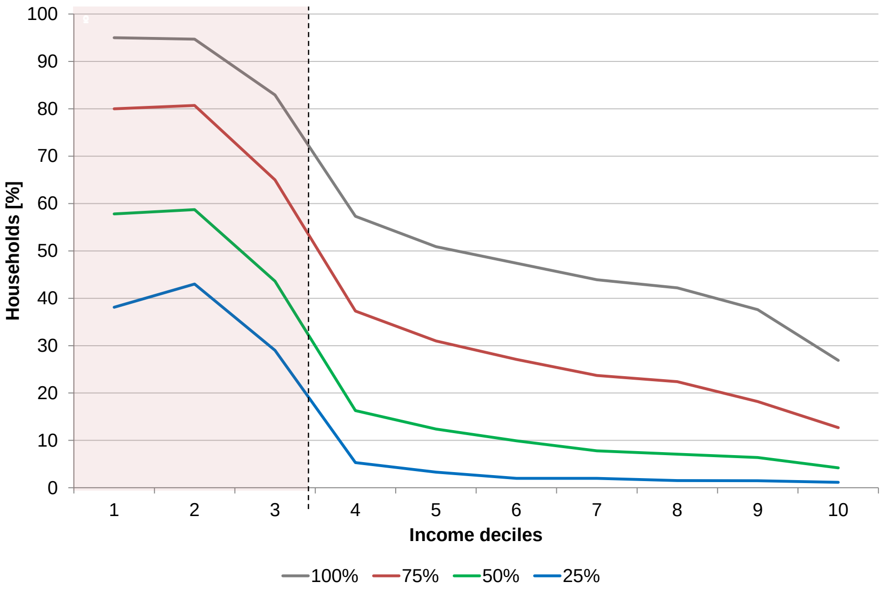


Figure 20 - π_2 indicator results for different income deciles (2014) [Own Elaboration based on INE EPF Base 2006 data]. The red shaded area indicates the incomes affected by monetary poverty

γ_4 and δ_2 do not show such good performances (Figure 21). The former, in the 25% case, converges to zero only for the tenth decile. The latter converges too quickly, including mostly families from the first income class. For this reason, we can state that δ_2 ends up with measuring how many households have extremely serious financial situations. π_2, on the other hand, is not confusing energy poverty with general poverty, but is measuring the latter as a consequence of the former, including information about energy consumption levels, too.

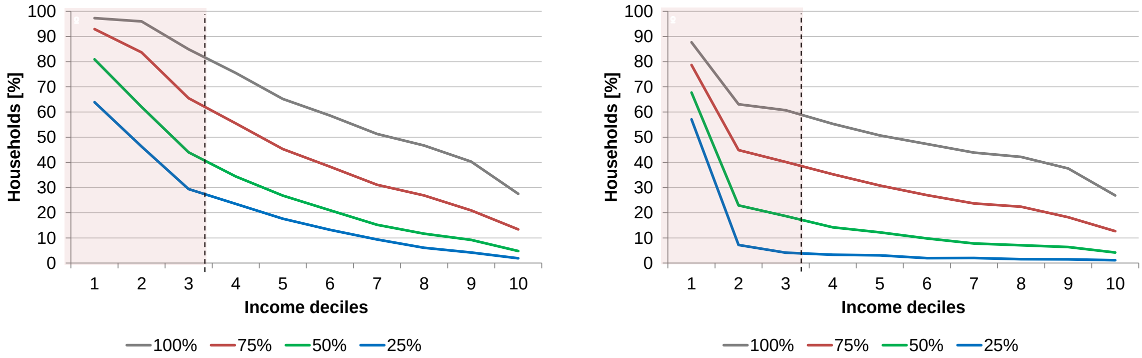


Figure 21 – $\gamma_4$and δ_2 indicator results for different income deciles (2014) [Own Elaboration based on INE EPF Base 2006 data]. The red shaded area indicates the incomes affected by monetary poverty

The previous part has introduced the concept of *false positive* samples. An household, considered energy poor, should be reported as false positive if it has an annual income level not belonging to the first three income deciles. For this reason we can now define the indicator targeting ability as the capacity of keeping to a minimum the number of false positive households. The *EEI* efficiency is perfectly in line with the π_2, indicating that the LIHC family indicators are the best according to this evaluation parameter for the specific Spanish study case.

We will now evaluate, with two Venn diagrams, the targeting intersections between the selected indicators. It is clear, from previous discussion, that γ_4 indicator will individuate an higher share of energy poor families. Figure 22 shows that this is exactly the case: the 7.54% of households is, in fact, signalled by the only DM adjusted indicator (i.e. γ_4). Both the LIHC and MIS adjusted have a lower overestimation risk, focusing only on lower income deciles.

The right graph shows that the 6.1% of poor (i.e. energetically speaking) is reported by both π_2 and *EEI*. This means that there is a high overlapping between these two, as they both belong to LIHC family. The 1.37% identified only by π_2 refers to families with major energy expense caused by high dwelling's surface. On the other hand, *EEI* alone identifies as vulnerable the 2.15% of sampled families. They are characterized by not too high energy expenditures in absolute terms, but by low energy efficiency standards (i.e. high energy requirements per unit of surface $\frac{€}{m^2}$). The most important information is that almost 5% (red arrow in Figure 22) of Spanish households is reported by all three indicators. Thus, at preliminary level we can state that, at least, 190 thousand families in Spain are reported by three heterogeneous indicators.

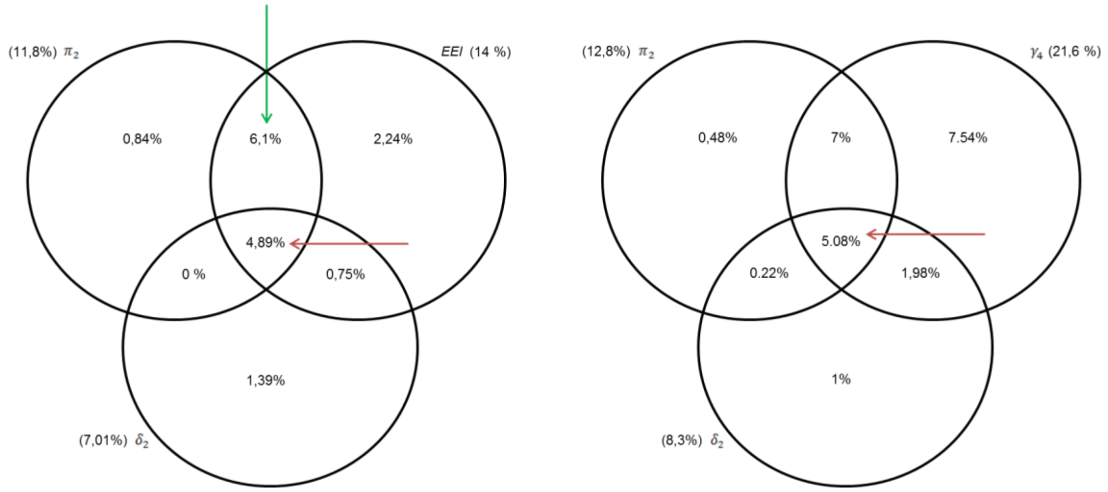


Figure 22 - Intersections evaluation of the indicators γ_4, π_2, δ_2, and EEI (2014) [Own Elaboration based on INE EPF Base 2006 data]

From the right Venn diagram we can also notice that, by using LIHC family indicators, it is possible to cover the majority of families reported by MIS indicators, too.

A γ_4 indicator is not strictly measuring poverty situations, but is rather reporting highly unbalanced expenses on energy products. The δ_2 indicator, as already explained, is converging to “general” poverty, not taking into account the impacts of energy efficiency at all (major expenditure on energy with respect to national median).

Table 4 further demonstrates that γ_4 (DM adjusted) is not suitable for addressing energy poverty in a reliable and socially fair way for the Spanish case.

	γ_4	π_2	δ_2
False Positive Rate [%]	35.35	13.31	17.65
Modified Index [%]	13.88	10.96	6.78
Delta [pp]	21.47	2.35	10.87

Table 4 – False Positive Rate estimation and Targeting Ability assessment for the Spanish case (2014) [Own Elaboration based on INE EPF Base 2006 data]

In the first row, it is possible to see what is the *false positives rate* for the three chosen indicators. The second row shows the modified indicators, in the case of just considering the first three income deciles of population, and, the Delta row indicates the difference, in percentage points, between the original indicator and the modified one.

We can conclude that γ_4 should not be used as energy poverty indicator due to its low Targeting Ability that can lead policy makers to focus their attention on population groups which are not really facing stringent deprivation conditions.

On the other hand, δ_2 has very low heterogeneity failing at including in the analysis the essential energy efficiency component. Moreover, we have demonstrated that LIHC indicators can efficiently cover all the families identified by δ_2.

In conclusion, LIHC indicators, according to this work, should be chosen to study and assess the energy vulnerability issue in Spain. They have good Targeting Ability and high heterogeneity including both energy efficiency and economic considerations.

One of the most common discussions in this field is whether energy poverty should or should not be assessed by “general” poverty measurement tools and strategies.

EPEE (2013) and Hills (2012) [13, 23] consider energy poverty as a phenomenon that can be determined by “general” poverty conditions, but that cannot absolutely be considered the same.

The purpose now is to test if it is possible to reach such conclusions for Spain, comparing the levels of monetary and energy poverty across Spanish regions.

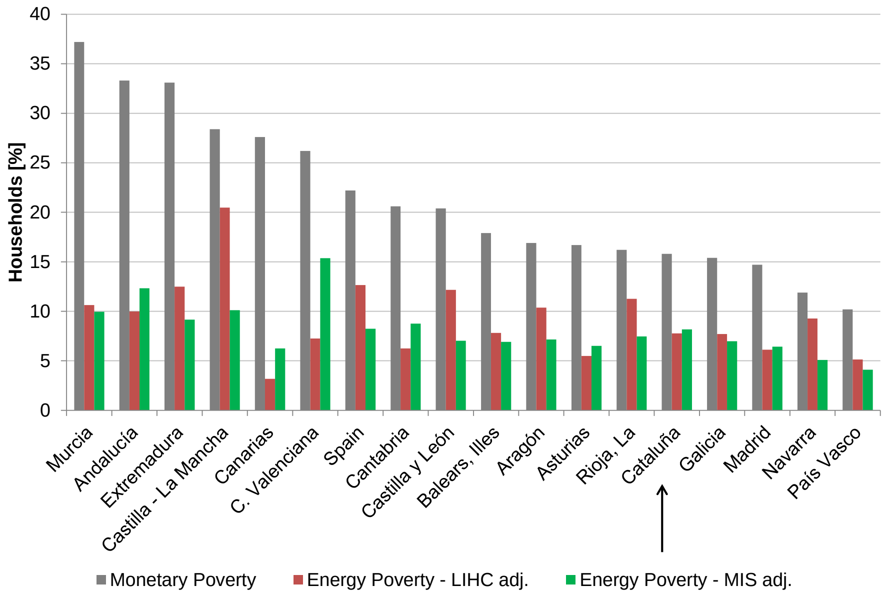


Figure 23 - Comparison between general poverty and energy poverty indicators (2014) [Own Elaboration based on INE EPF Base 2006 and ECV data]

Figure 23 ranks, in black columns, Spanish autonomous regions according to their monetary poverty levels. The data have been taken from the *Encuesta de Condiciones de Vida* (ECV) and indicates the population share living with annual incomes below poverty line (60% of median income). This level for Catalonia is around 15% (black arrow).

The plot proves that energy poverty is a component of general poverty, and that is absolutely not the same phenomenon. Thus, using the same indicators to measure both problems at the same time would provide highly overestimated results. Identification strategy should therefore include other factors and variables other than just families' incomes.

At this point it is interesting to evaluate indicators' chronological evolution to see how their levels have been influenced by both macroeconomic and energy pricing trends. In particular, the former will show the effects of the economic crisis, and the latter the outcomes of Natural Gas and Electricity of last years' rising trends.

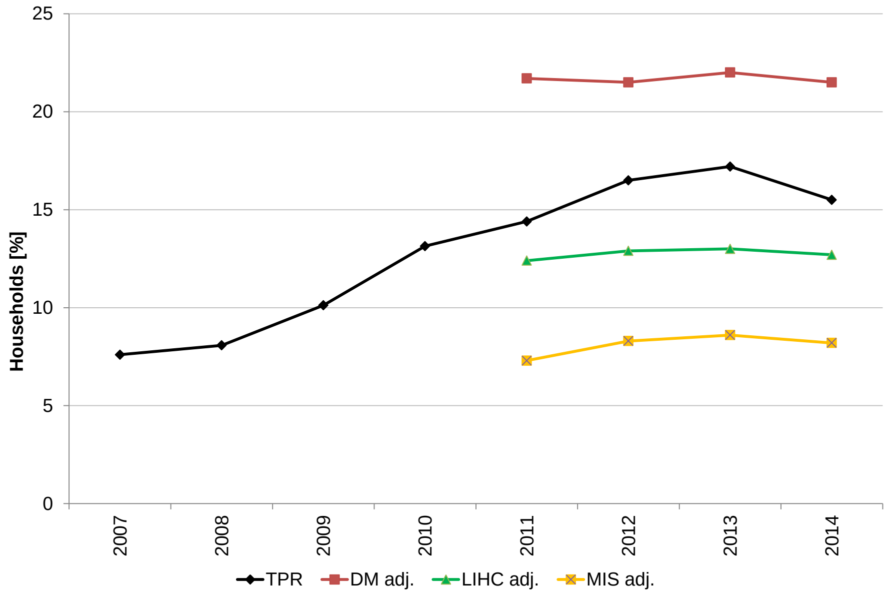


Figure 24 - Indicators evolution in time 2011-2014. The TPR indicator shows the ω evolution starting from 2007 [Own Elaboration based on INE EPF Base 2006 data]

As already anticipated, the black line in Figure 24, representing the TPR indicator, has the scope of assessing the energy poverty levels before 2011. The data for these four years have been taken from Romero's (2014) [20] study on energy poverty in Spain. The TPR trend is important to evaluate how the ω levels have drastically increased from pre-crisis level (2007) to last years' ones. In 2013, in fact, there was a peak for all the indicators considered in this chronological analysis. We can infer a similar trend for all the indicators, clearly demonstrating that the combination of rising energy prices and lower incomes during crisis period has led to a substantial increase from pre-crisis levels. Moreover, from the plot, we can also notice how the LIHC adjusted indicator is more stable in time than the others, as, being a relative indicator, the fluctuations in energy prices are smoothed out. The values, according to this indicator, between 2011 and 2014, remained almost constant, slightly increasing from 12.4% to 12.7%.

We are now interested, moving one step forward, in analysing the problem across Spanish regions. In order to do that, Table 5, gathers the values corresponding to eight of the analysed indicators showing how the situation changes among the nineteen study cases. Moreover, it is also possible to notice the median expenditure per surface ($\frac{€}{m^2\ year}$) area and in absolute terms ($\frac{€}{year}$).

2014	Objective indicators [% of households]								Median expense on energy*	
	γ_1	γ_3	γ_4	π_1	π_2	δ_1	δ_2	EEI	Surface**	Absolute***
Andalucía	14,1	16,1	21,3	10,1	15,2	9	12,9	15,8	8	788,9
Aragón	19,5	21,5	24,6	10,4	13,5	4	6,7	15,6	12,3	1117
Asturias, Principado de	10,2	11,5	14,6	5,5	8,6	4,2	6,5	10,5	10,9	924,1
Balears, Illes	15,5	16,9	21,3	7,8	12,2	3,3	7,1	12,8	9,1	905,1
Canarias	7,2	8,3	19,1	3,2	14	10,1	18,3	17,2	6	568,7
Cantabria	14,1	16,1	18,1	6,3	8,3	4,7	6,3	12,6	10,9	976,3
Castilla y León	22,4	25,1	29,1	12,2	16,2	4,4	8	17,1	11,5	1094,3
Castilla – La Mancha	36,5	39,6	42,5	20,5	24,7	6,4	8,7	20,4	11,3	1269,6
Cataluña	**15,5**	**17,6**	**21,2**	**7,8**	**11,8**	**3,7**	**6,6**	**13,5**	**11,6**	**1025,4**
Comunitat Valenciana	10,7	12,2	16,8	7,3	12	5,6	9,2	12,3	8	817,5
Extremadura	19,5	22,5	26,6	12,5	16,6	5,8	9,3	14,3	8,2	906
Galicia	15,4	17,7	21,5	7,7	11,5	4,3	7,1	12,6	9	919,1
Madrid, Comunidad de	12,2	14,3	17,1	6,1	9	3,5	5,9	12,2	12,7	1093,2
Murcia, Región de	16,6	18,6	22	10,6	14,1	7,6	10,1	12,4	8,2	889,5
Navarra, Comunidad Foral de	18,4	21,8	24	9,3	11,5	2,7	4,8	124,4	12,1	1149,4
País Vasco	7,7	9,2	12	5,1	8	1,4	4	9,8	11,9	995,4
Rioja, La	21,7	24,4	27,4	11,3	14,2	4,3	6,9	14,8	11,8	1140,1
Ceuta	5,8	10	20	3,3	13,3	12,5	19,2	15,8	6,5	507,3
Melilla	12,4	14,7	18,6	5,4	9,3	8,5	12,4	19,4	8,5	724,8
Spain	**15,5**	**17,6**	**21,6**	**8,8**	**12,8**	**5,1**	**8,3**	**13,9**	**10**	**942,2**

*Table 5 – Indicators results per Spanish region for the year 2014 [Data: INE EPF Base 2006]. *Expense on energy products as € per year and per household. ** Annual energy expense per m^2. *** Annual energy expense per household*

The situation displayed in Table 5, is *highly heterogeneous*. The most serious conditions are found in Southern and Central regions: Castilla la Mancha, Castilla y León, Extremadura, and Andalucía. On the other side, the Atlantic area is showing lower risk for energy vulnerability.

Catalonia presents energy poverty levels in line with the national average, even though the absolute median expense on energy products is higher than the national level by 8.8%. From a first look at Table 5, we can see that highly economic focused indicators, like γ_4 and δ_2, are below average, while indicators that take into account the energy efficiency component, are showing performances worse than the average. We can say, according to ACA (2016) [14], that the energy poverty problem in Spain can be divided into *two categories*.

On one side, there are regions experiencing serious financial situations, where households have, in general, low income levels, keeping the ω_i ratio to a minimum.

On the contrary, in Northern regions (like Catalonia) it seems that the energy efficiency component is playing a major role. In the latter case, the LIHC indicators are detecting households which have an expenditure on energy above the national median and a residual income lower than the monetary poverty threshold.

The problem is therefore divided in two groups. The first considers mainly the social and economic effects caused by the economic crisis that has heavily hit Southern and Central regions, while the second is mainly focused on another energy poverty aspect: energy efficiency. In this case, families, originally not in "general" poverty, are brought to energy poverty conditions by major expense on energy products, caused by high dwelling's surfaces and/or poor energy efficiency standards.

This is just a qualitative intuition that needs to be further justified. We will therefore consider and analyse the number of fuel poor households that are also affected by "general" poverty, using the selected π_2 indicator.

We expect that, in Southern and Central regions, the number of non-poor (in monetary terms) households affected by energy vulnerability will be close to zero. On the contrary, in Northern and Atlantic regions the energy poverty phenomenon would be shared by most income groups. In order to do so, we will consider three significant cases: Castilla la Mancha, as it is the region showing most critical performances, Catalonia, not only because it is the subject of this work, but also because it is a region with an average vulnerability situation, and finally País Vasco, as least affected region according to Table 5. Moreover, we will divide the sample into three income groups.

The first will include the three initial income deciles (monetary poverty *danger zone*), the second the income groups between the 60% to 100% of the median, and the third the remaining income deciles.

A Ratio has been calculated evaluating the percentage of identified vulnerable citizens belonging to the third group over the total. The second income group (second row of Table 6) should be still considered financially vulnerable, even though not strictly recognised as in “general” poverty conditions.

The results are shown in Table 6. It is straightforward to notice how, for Catalonia and País Vasco, this ratio is significantly higher than for Castilla la Mancha (by 30%). This means that, using the same indicator, the energy poverty problem in these regions is more differentiated from “general” poverty.

π_2	Castilla la Mancha	Cataluña	País Vasco
<60% median	19,9	8,67	5,9
60% - 100% median	4,15	2,1	1,4
> median	0,66	1,05	0,77
Ratio [%]	**2,7**	**8,9**	**9,5**

Table 6 - Sensitivity study, for the π_2 indicator, between three income groups for the year (2014) [Own Elaboration based on INE EPF Base 2006 data]

From this sensitivity study, we can draw two important conclusions:

1. The energy poverty issue across Spanish regions has two connotations, in relation with the three energy poverty drivers described in Chapter 3. On one side, energy poverty can be mainly caused by intrinsic regional financial (i.e. monetary poverty) and unemployment conditions. On the other side (as for Catalonia), the issue is not only related to social and macroeconomic conditions, but also to higher energy needs. This demonstrates that *energy efficiency* must be evaluated as a crucial driver for energy poverty.

2. The problem in Catalonia, cannot be measured by only socially and economically focused indicators (i.e. δ_2 and γ_4) but with LIHC family ones.

Based on previous discussion, we can conclude that if we consider the impact of energy poverty on non-poor income deciles (Table 6 second and third rows), Catalonia is among the regions with worst performances.

The same can also be said for: Madrid, País Vasco, Baleares, Catalonia, and Navarra. According to Figure 23 the latter ones are showing relatively low levels of “general” poverty.

In these regions energy poverty must be assessed with different methodologies and tackled with broader strategies covering all the three problem’s dimensions (see Paragraph 3.2).

5.3 The Catalan Case

According to the previous analysis, we can state that indicators π_2 and *EEI* are the most suitable for the Catalan study case, and for the regions listed above.

EEI indicator shows that the 13.9% of Catalan households are living in energy poverty conditions, having deficient energy efficiency conditions, (excessive energy consumption in terms of € per m^2), or energy expenditure lower than 25% of the national median. Indicators π_2 reports an 11.8% energy poverty rate. The national levels are, respectively, 13.8% and 12.6%.

Figure 25 represents the interactions among three indicators for the Catalan study case. It provides a comparison with respect to the national situation shown in Figure 22. In previous Paragraph we have concluded that in Catalonia, the energy efficiency component is playing a major role in determining energy poverty situations. High expenditure on energy, caused either by high surface dwellings or lack of energy efficiency standards, can move families into real “general” poverty conditions. To further explain this point, one can consider the case of an household whose income is originally higher than the “general” poverty threshold, but its *residual income*, after the payment of utilities bills, falls below the very same threshold.

Those situations are described by the overlapping regions between *EEI* and π_2 in Figure 25. On the other hand, δ_2 indicator is highly “covered” by the other two. In fact, the overlapping among the three indicators (red arrow in Figure 25) is 4,89% out of a total MIS adjusted rate of 7.03%. This means that, if one between *EEI* and π_2 is used, the 72% of families identified by δ_2 would be considered, too.

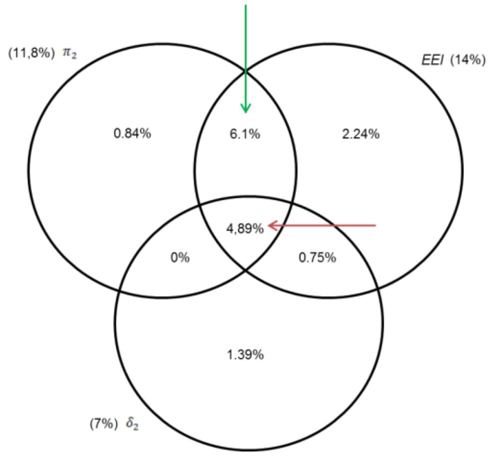


Figure 25 - Intersections evaluation of indicators EEI, π_2, and $\delta_2$for Catalonia (2014)
[Own Elaboration based on INE EPF Base 2006 data]

Moreover, Figure 25 allows us to conclude that 4.89% of Catalan families is identified by all the three considered indicators. If we further elaborate the data, by considering those households' members number, we can conclude that, in 2014, there were 316.6 thousands Catalan citizens at energy poverty risk.

Figure 26 shows the energy poverty chronological trend for Catalonia in the time span 2011-2014. The LIHC adjusted indicator (i.e. π_2) is represented by the green solid line, as it has been selected as the most suitable indicator for both Spain and Catalonia. The remaining indicators, dashed lines, are displayed for sake of completeness.

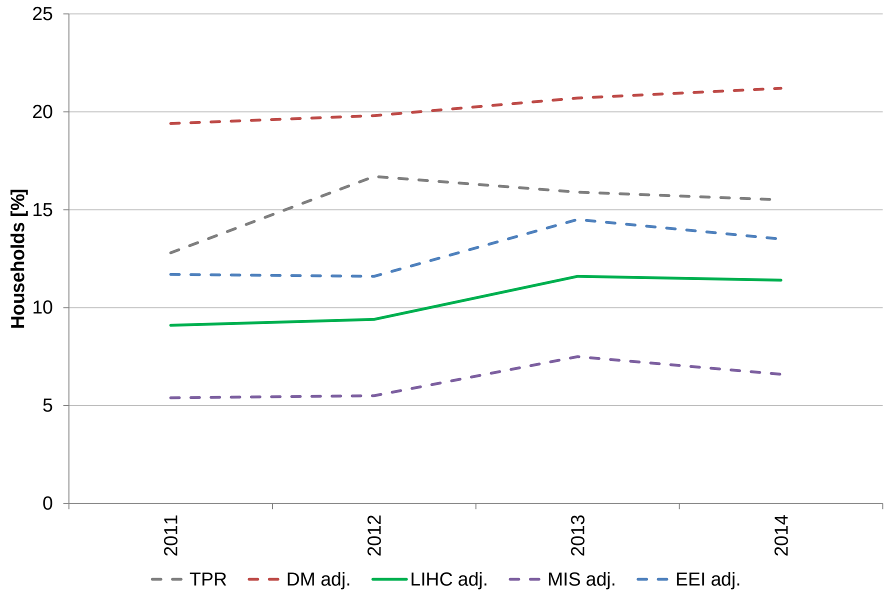


Figure 26 - Energy Poverty historical trend (2011-2014) for the five indicators considered in Chapter 5. . The dashed lines correspond to the indicators that did not show suitable characteristics according to our analysis [Own Elaboration based on INE EPF Base 2006 data]

6 Energy Poverty Modelling

6.1 Introduction to Classification Models

The purpose of this Chapter is to provide a quantitative analysis of the variables that can increase the energy vulnerability risk. The scope is to identify vulnerable groups without explicitly knowing their energy expenses and, initially, their net incomes. This is useful since the latter information are generally hard to retrieve and gather for an entire population at both national and city levels. Moreover, information about energy expenditures have to be gathered from utilities private companies and are not commonly available to public authorities. We are trying to understand what modelling confidence level can be obtained using only the variables already available to local public entities (e.g. City Councils) and publicly shared on Open Data platforms, too.

First, the entire INE EPF database will be considered, and then, for refining the results to our practical application, the analysis will focus on Spanish cities (i.e. with more than 100,000 inhabitants). In order to do so, we will use three statistical and machine learning instruments, identifying which variables are the most relevant according to the most accurate model found.

Chapter 4 defined π_2 and *EEI* (LIHC family) as the most suitable indicators for studying energy poverty in Spain. Due to the higher international comparability (mainly with the British case) and standardization level, we will consider, for the modelling part, the π_2 indicator.

The latter will assign a binary value (y) to every sampled household. If the value is 1 the family is facing energy poverty conditions, while if 0 is not. This approach is similar to medical applications where machine learning methods are widely used (i.e. ill patient=1, healthy patient=0). Equation 6.1 represents this concept, with respect to π_2 definition (see Paragraph 4.2.4).

$$\begin{cases} I\left[s_i^{tot} > Md_2\right] \cdot I\left[\left(y_i^{tot} - s_i^{tot}\right) < Md_T\right] = 1 \quad \rightharpoonup \quad y = 1 \\ \\ I\left[s_i^{tot} > Md_2\right] \cdot I\left[\left(y_i^{tot} - s_i^{tot}\right) < Md_T\right] = 0 \quad \rightharpoonup \quad y = 0 \end{cases} \tag{6.1}$$

We will call y True Condition throughout Chapter 6.

The energy poverty percentage level in Spain for the year 2014 was 12.8% (see Table 5). For this reason the data set will be imbalanced, since the number of non-poor households is almost nine times higher than the number of truly poor ones. Therefore, our problem would be similar to the needle in an haystack one. We will come back to this later on in the Chapter.

Once a suitable indicator is defined and a binary value is assigned to every sampled household, we can start the modelling phase. At this stage, we know the results (True Conditions) but we do not know which are the weights of the considered variable x_i in determining an energy poverty status (i.e. 1 or 0). We are therefore looking for a way of identifying how the variables x_i are driving the problem through a model h_θ. We will therefore follow the backwards process described in Figure 27.

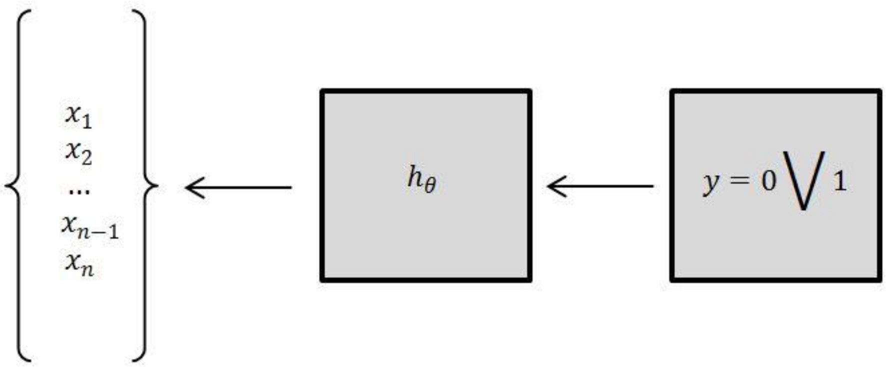


Figure 27 - Explanation chart of the methodology used in Chapter 6. A model h_θ has to be trained in order to obtain results as close as possible to the True Condition (i.e. y) considering the variables $x_1 \cdots x_n$ of the original database EPF. Finally, it will be possible to determine each variable's contribution to model's performance [Own Elaboration]

Once the model will be determined, its outcome, defined as $\hat{y}$ (i.e. Predicted Condition), will be compared to the True Condition y. If the model will be considered valuable, we will be able to determine what are the significant variables and infer their role in the energy poverty issue.

6.2 The Machine Learning Contribution

Machine learning is a subfield of computer science particularly suitable for patterns recognition. The main idea is that, through machine learning methods, the computers can learn how to deal with certain data without being explicitly programmed. It allows the construction of algorithms for the study of data frameworks, and also for predicting outcomes from new data [27].

Due to energy poverty's high drivers number it is straightforward to imagine machine learning as an effective tool to explore the available database (i.e. EPF by INE). We will furthermore demonstrate the possibility of making predictions using the models found.

We will thus follow a *four steps process*:

1. Train three different models.
2. Assess each model's accuracy in correctly identifying energy poor families among all households with $y = 1$ (i.e. True Positivies).
3. Predict outcomes from a new data set and evaluate model's predictive power.
4. Determine what are the most significant variables according to the most accurate implemented model.

6.3 Database Organization

To do this, we need to split the data set in three blocks. The first will help us during the training phase, in which we will estimate model's parameters, the second (i.e. cross-validation) will tune the model identified in the first step, and the final will allow to evaluate the performance of the model in dealing with a new data set (i.e. model's ability to *generalize*).

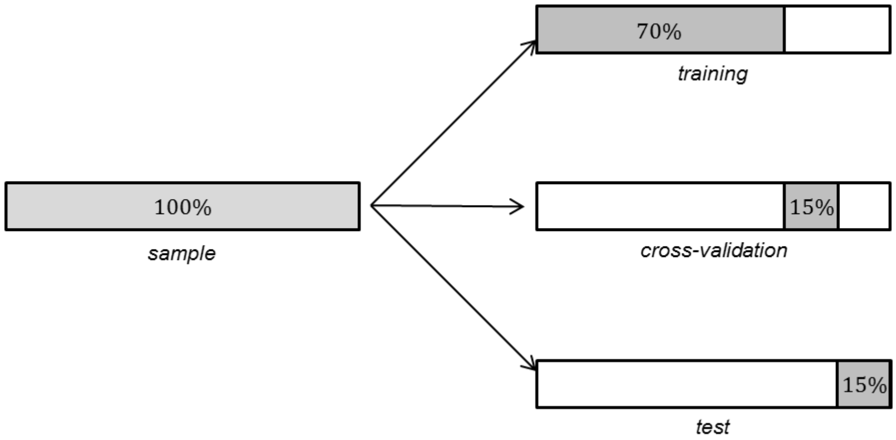


Figure 28 - Database organization for training, cross-validation, and test phases [Own Elaboration]

It is necessary to decide how big should each block be. If too many data are spent in training we will not obtain a good assessment of predictive performance (i.e. fitting). We may find a model that fits the training data very well, but that is not generalizable. The latter concept is called *over-fitting*. On the other hand, if too much is spent on testing, we will not have a good assessment of model's parameters.

In general, it is common practice to use 70% (aleatory chosen) of data for training, 15% for cross-validation and 15% for testing (see Figure 28).

6.4 Definition of the Problem

Once a training set is defined, the following step is to use a *learning algorithm* to define an *hypothesis function* h_θ. This functions takes an input x_i (e.g. in our case a database variable) and outputs an estimated $\hat{y}$. The hypothesis function can assume different forms, based on the method used, for instance, in the case of Linear Regression:

$$h_\theta = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n = \theta^T x \qquad (6.2)$$

In Equation 6.2 the $\theta_i$terms are the function *parameters*, while the n value indicates the number of variables considered by the model.

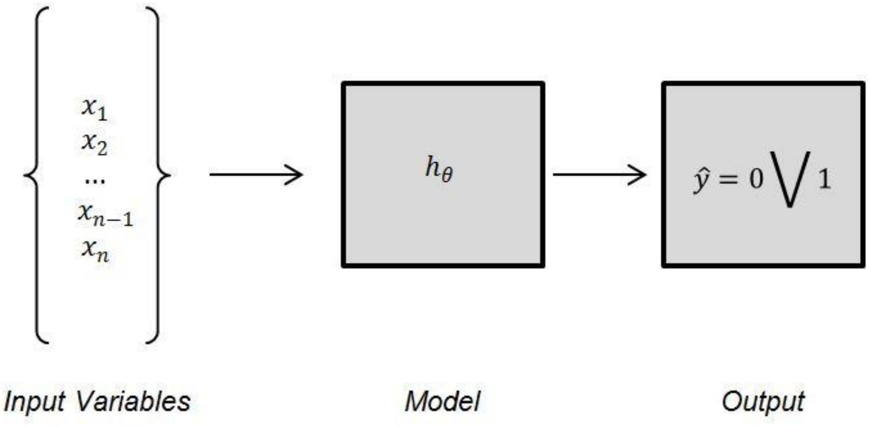


Figure 29 - Work flow process. The model, processing the input variables $x_1 \cdots x_n$, outputs a Predicted Condition(i.e. $\hat{y}$) for each sampled household [Own Elaboration]

We can now evaluate how the model is performing, comparing predicted values (i.e. $\hat{y}$) with real ones (i.e. y). The former are Predicted Conditions, while the latter are the True Conditions. The process is represented in Figure 29. It is important to distinguish it from Figure 27, where True Conditions (y) were instead considered, in the "Output" grey box.

6.5 Model's Evaluation Theory

In Paragraph 6.1 we mentioned the imbalanced characteristic of the data set. In our case the True zeros will be almost nine times more numerous than the True ones. The problem of imbalanced data sets is particularly relevant and hard to solve in machine learning applications. Intuitively, one can imagine that the algorithm would be more precise in categorizing zeros than ones.

For this reason, we will divide the evaluation in *two phases*. In the first phase we will evaluate the model's *accuracy* in *detecting* energy poor families, while in the second phase we will assess the model's ability of *efficiently predicting* families' energy poverty status.

Given a certain ensemble of True Positives and one of True Negatives households, model's *accuracy* will be defined as the number of correctly identified items from each of the two ensembles. On the other hand, the model will have high *predicting Precision* if the majority of the reported families is really facing energy poverty conditions.

The performances' evaluation, due to the binary nature of the problem, will be carried out using the, so called, Confusion Matrix. From the latter we will derive significant parameters, that will be combined in the Receiving Operating Curve (ROC) and Precision vs. Recall analysis.

A “Confusion Matrix” is a cross-tabulation of True and Predicted Conditions:

		Predicted Condition	
		0	1
True Condition	0	A	B
	1	C	D

Table 7 - Confusion Matrix for models evaluation [Own Elaboration]

In Table 7, A is the number of True Negatives identified by the model, or, more specifically the number of predicted zeros (i.e. non energy poor), which are really (i.e. True Condition) non poor.

The concept applies also for D (True Positives): it indicates the number of predicted ones (i.e. energy poor) which are really facing energy poverty conditions.

The other two blocks represent *model’s errors*.

B is the number of False Positives, or, in statistical terms *false alarms*. They are identified by the model as energy poor, but in reality they are not.

In C we see the number of households that the model does not consider vulnerable, while the True Condition row tells us that they should.

We will now introduce four factors, in percentage unit, that are essential for Confusion Matrix and model’s evaluation:

$$Sensitivity = True\ Positives\ Rate\ (TPR) = \frac{D}{C+D} * 100; \tag{6.3}$$

$$Specificity = True\ Negatives\ Rate\ (TNR) = \frac{A}{A+B} * 100; \tag{6.4}$$

$$False\ Positives\ Rate\ (FPR) = \frac{B}{A+B} * 100; \tag{6.5}$$

$$False\ Negatives\ Rate\ (FNR) = \frac{C}{C + D} * 100; \tag{6.6}$$

$$Accuracy\ (ACC) = \frac{A + D}{A + B + C + D} * 100; \tag{6.7}$$

For the problem's characterization phase we are particularly interested in the Sensitivity and Specificity concepts:

- Sensitivity: given that a result is truly an event (i.e. energy poor), what is the probability that the model will predict it as a positive?
- Specificity: given that a result is truly NOT an event, what is the probability that the model will predict it as a negative?

These conditional probabilities are measured by True Positive and True Negative Rates described in Equation 6.3 and Equation 6.4. From now on, we will therefore refer to TPR as Sensitivity and to TNR as Specificity.

A model, in the case of a binary outcome, will output, for each sample, a certain probability $\hat{p}$ of being a 1. What we need to choose is a threshold to decide whether the sample with probability $\hat{p}$ should or should not be considered as a positive (i.e. $\hat{y} = 1$). Randomly, in R, such a threshold is chosen to be 0.5. Thus, if $\hat{p}$ is higher than 0.5 the model's result would be 1, and if $\hat{p}$ is lower than 0.5 the result would be 0.

It is straightforward to imagine that there would be a trade-off in the choice of such a threshold. The purpose of the cross-validation dataset is exactly to tune the trained model to obtain this optimal cut-off threshold. The ROC and the Precision vs. Recall curves provide two possible *optimization methods* to fulfil this tuning objective [27].

The ROC curve allows to maximize Sensitivity (i.e. the benefit) while decreasing to a minimum the cost (i.e. the FPR or Fallout). The latter is computed as:

$$Fallout\ (FPR) = 1 - Specificity\ (TNR); \tag{6.8}$$

For our application, it is perfectly consistent addressing to fallout as "cost". The latter concept is represented by the FPR: number of false alarm over the total number of Condition Negatives (i.e. zeros).

With two classes problems, the ROC curve can be used to find the optimal cut-off threshold. The ROC represents both Sensitivity and Fallout for many possible thresholds (see Figure 30). The best possible classifier (i.e. model) would yield a point the closest possible to the upper left corner (i.e. with coordinates 0,1).

Thus, the following function should be minimised:

$$Distance\ (d) = \sqrt{FPR^2 + (1 - TPR^2)}; \tag{6.9}$$

This is the Euclidian distance from each point of the ROC curve to point (0,1), called point of *Perfect Classification*. The optimal cut-off point would be the one with smallest d.

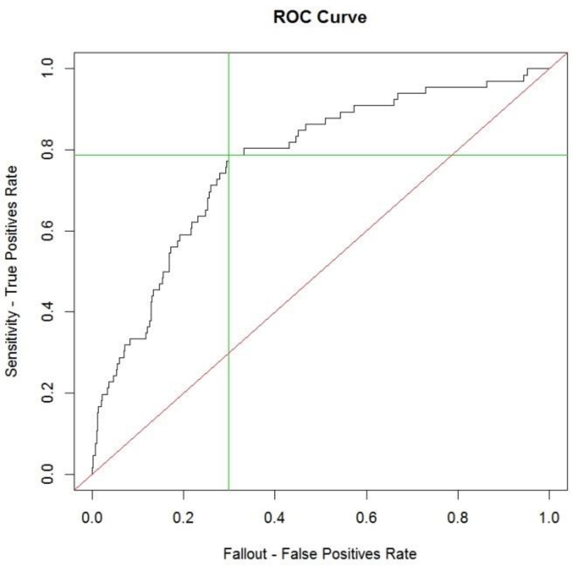


Figure 30 - ROC curve's explanatory representation [Own Elaboration]

A completely random guess would lead to a point on the red line, called line of *no-discrimination*. If we would model the results of a non-rigged flipping coin we would obtain, as optimum, the point with coordinates (0.5,0.5), which basically means that the model is succeeding as many times as it fails (i.e. TPR equals to FPR).

A further optimization method to determine the cut-off threshold is based on Precision vs. Recall curve.

Precision is calculated as:

$$Precision = \frac{D}{B+D}; \tag{6.10}$$

It assesses the measurement tool's *targeting efficiency*. In order to understand that, we should imagine that such a model would be used by a policy maker for introducing a financial aid (for instance a bonus on utility bills) for decreasing the number of energy vulnerable households in a country. The *efficiency* is determined by the number of helped families that are really facing energy poverty conditions.

As for the ROC optimization we will compute the distance from the point of *Perfect Classification*, whose coordinates in this case are (1,1).

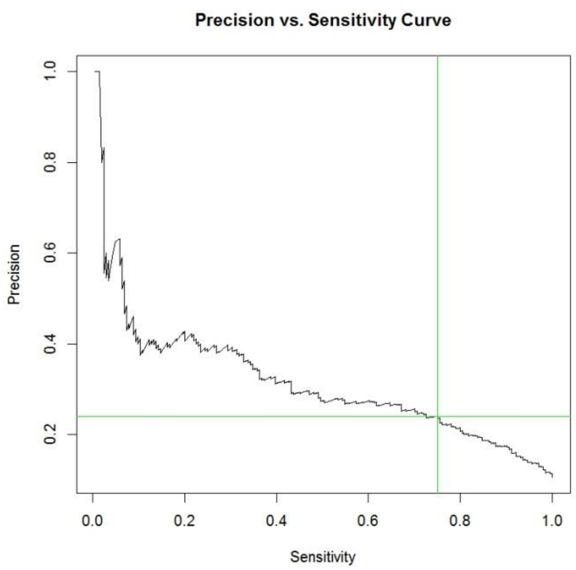


Figure 31 – Precision vs. Sensitivity curve's explanatory representation [Own Elaboration]

Precision is conditioned by the population prevalence (*imbalance ratio*). In other words, it will depend on how many True ones are present in the population over the total number of households. Thus, in our application, this term would be different across all Spanish regions, as characterized by different π_2, and thus by different imbalance situations.

On the other hand, Sensitivity and Fallout (variables of the ROC curve) are not conditioned by the imbalance ratio, making the outcomes of the ROC curve analysis generalizable and stable for different prevalence populations groups.

The ROC curve will be used to assess model's accuracy to get insight over the variables that are driving energy poverty. Precision will be later considered to evaluate model's eligibility as policy makers' support tools. In conclusion, in the following part, the cut-off threshold will be obtained from ROC curve optimization.

6.6 Decision Tree Learning

The Decision Tree Learning method was chosen due to its simplicity and for its high interpretability. It will allow to have a graphical representation of the variables, and categories that determine energy vulnerability situations. They are able to arrange the observations in a very transparent way with a series of if-then statements that will take a typical tree shape with nodes, branches, and leaves [27].

First, the algorithm creates a root node, that divide the data-set in the most effective way: by considering all the observations, it chooses the variables that best splits the data in two blocks. Two nodes are generated. Afterwards the very same method is applied to the latter ones, generating four more nodes. The algorithm is called recursive partitioning, and it will end when the splitting does not add further value to the model.

To understand this stopping criterion we should refer to the concept of Entropy in information theory. It assess the amount of disorder in a set, or, in other words, how mixed a data set is. It is useful to measure how different the outcomes are from each other. If the value is close to zero, it means that the observations are really similar, while if the algorithm measures an high entropy value, it will perform a further split. The goal is reducing the entropy to an agreed minimum.

The lower this minimum is the more complicated the tree would be (i.e. higher number of nodes and branches). Since we have chosen this methodology to have an easy and straightforward representation of energy poverty in the data set, we will accept a low algorithm's predictive power, while having an easily understandable graphical representation.

In Figure 32, we can see an example of a decision tree applied to our dataset. There are 12 terminal nodes that represents, in brackets, the probability of a family of being a 0 or a 1 respectively.

The model uses the following variables: household's main component study level, tenancy, current job situation, building typology, and the magnitude of the city where the sampled household is living. On each branch we can see the logical conditions that are ruling each split step. The number present in the logical conditions, indicates the categorical variables used in the EPF statistical survey. The reader can refer to Annex, where each categorical variable is explained, combining each number with the corresponding verbal expression.

In order to describe the process, we will consider an explanatory case (red path in Figure 32) applying a top-down approach for a randomly chosen household. The first test (root node) asks whether the family's main component has a very high (i.e. university) education level or not. If the condition is true we will move to the right otherwise to the left. We imagine that the sampled family's main member has a lower study level, thus we will follow the red path to the left. The next test asks what is the current job situation. If this categorical variable has a value higher than two, it means that the main member is unemployed, retired, a student, or has permanent handicaps. We imagine that this is the case, moving to the right. The following statement evaluates the magnitude of town where the sampled family is living. The latter will be classified as living in a town with more than 50,000 inhabitants. We have reached a leaf, or a final node, where we can finally get the percentage for the family to be considered energy poor. In this case the risk, expressed in percentage, would be 40%.

Applying the model to a new set of data (i.e. not used neither to train nor for cross-validating the model) we can infer the predictability power of the model. Each test set sample will be therefore assigned to a leaf (i.e. grey boxes in Figure 32) with the same process described above and in the picture by the red path. Afterwards a probability of being a 1 (i.e. energy poor) will be assigned to each sample. Using the cut-off optimal threshold computed with cross validation we can obtain a Confusion Matrix, and all the significant values described from Equation 6.3 to 6.7.

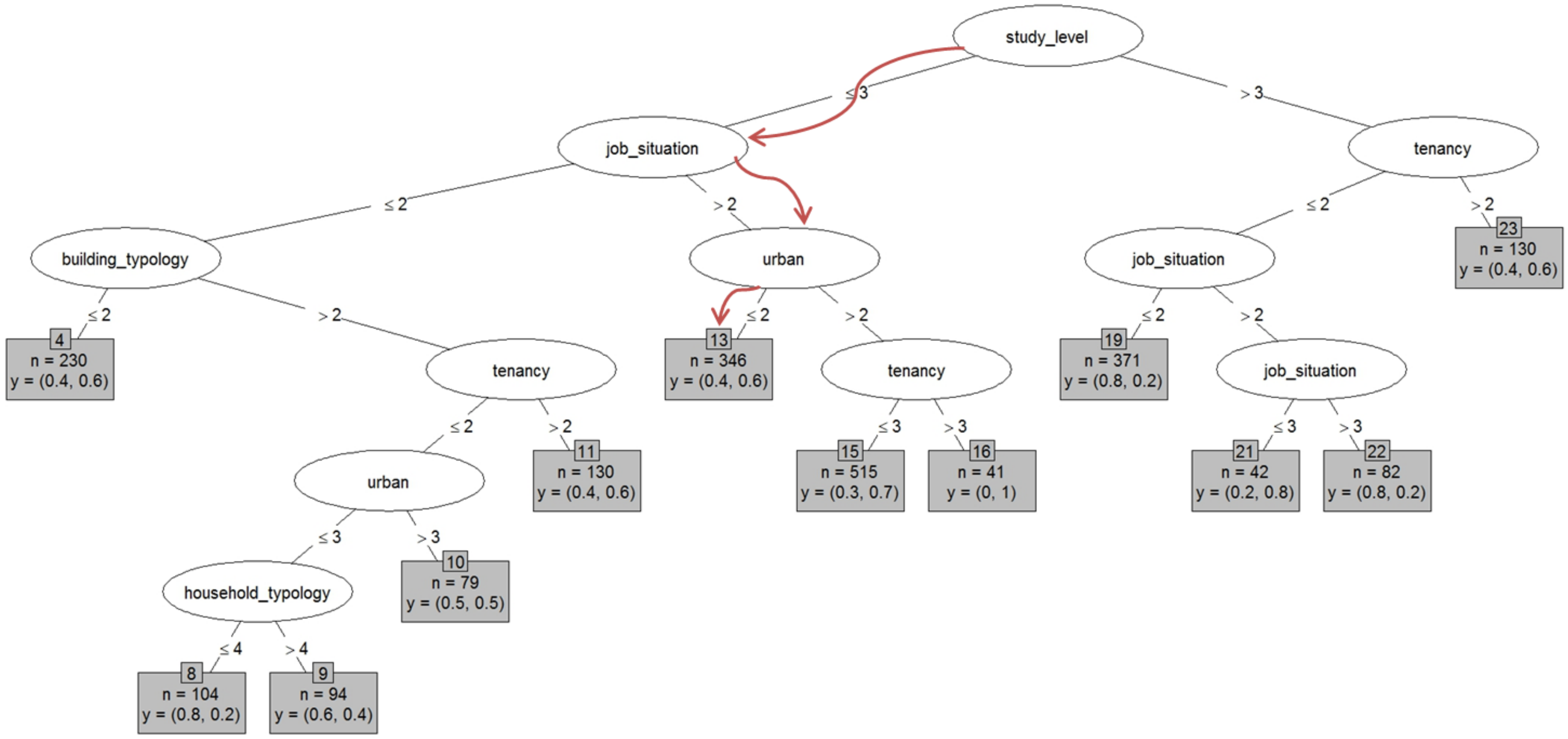


Figure 32 - Decisional Tree (for indicator π_2 in Spain for 2014) implemented with R software. In the grey boxes: the "n" indicates the number of families (of the test set) classified as belonging to a particular leaf; the "y" indicates respectively the probability for a family belonging to a particular node of being a 1 or a 0 [Own Elaboration]

The overall accuracy (ACC) of the model is, in this case, 78%, while the sensitivity (TPR), specificity (TNR), miss-rate (FNR) and fallout (FPR) are represented in Table 8.

TPR = 72%	*FPR = 28%*
FNR = 31%	*TNR = 69%*

Table 8 - Evaluation matrix of the Decisional Tree model for energy poverty in Spain (2014). In the first quadrant (I): False Positives Rate, in second quadrant (II): True Negatives Rate, in third quadrant (III): False Negatives Rate, and in the fourth quadrant: True Positives Rate

The sensitivity value is represented in the fourth quadrant and has a value of 69%. This means that the model identifies as ones (i.e. fuel poor) the 69% of households that are really facing energy poverty condition. The "cost" (i.e. quadrant I of Table 8) indicates that 28 out of 100 non energy poor families are considered as positives while in reality they are not experiencing any energy deprivation condition. The model's ROC curve is displayed in Figure 35 (green line).

6.7 Random Forest

Random Forests involve an ensemble of classification trees that are calculated on random subsets of the original data, using a subset of aleatory restricted and selected predictors for each split in each classification tree [28]. In this way, Random Forests allow to have a valuable and precise estimation of the contribution and behaviour that each predictor has.

Furthermore, according to the same reference, Regression Trees have been shown to produce better predictions than one classification Decisional Tree. They are particularly appropriate to tackle problems with highly heterogeneous predictors, as in our case. In order to "visualize" what a Random Forest looks like, the reader should imagine many (e.g. in our case 500) trees, as in Figure 32, grouped together. The prediction phase from a new data set is carried out, for classification, by aggregating the prediction of the N trees grown. The final model's output (i.e.$\hat{y}$) will be chosen according to the *majority votes criterion*: if the majority of sub-trees will give a 1 this will be the final Forest result.

This allows to consider many possible cases, that cannot be covered by a single tree. Moreover, Random Forests are not only suitable for prediction, but also to assess *variable importance*, a feature that will be of extreme importance in Paragraph 6.10. A further important added value is the possibility, considering variables' importance, to reduce the dimensionality (i.e. number of drivers or variables determining the studied phenomenon) of the treated problem.

The most common Random Forests' drawback is the fair interpretability. In fact, it is impossible in this case to have a clear and straightforward representation (see Figure 32) of the problem as for a single Decisional Trees.

We should add a further specification: since it is based on a truly "random" statistical instrument, the model, and thus the results can vary from run to run.

In Table 9 average significant results for 100 runs are provided while an average Precision of 27.6% has been achieved.

The model ROC curve will be shown in Paragraph 6.9 (see Figure 35) where all the models trained will be compared.

TPR = 70%	*FPR = 30%*
FNR = 25%	*TNR = 75%*

Table 9 - Results of the model based on Random Forest algorithm. In the first quadrant (I): False Positives Rate, in second quadrant (II): True Negatives Rate, in third quadrant (III): False Negatives Rate, and in the fourth quadrant: True Positives Rate

6.8 Logistic Regression

The third methodology used is Logistic Regression. Its approach is similar to Linear Regression, but in this case the output or target variable y can be binary or multiclass. For the latter reason, Logistic Regression has to be identified as a *classification algorithm* for distinguishing it from linear regression that can have other outputs than just zeros and ones.

In logistic regression the hypothesis representation is:

$$g(z) = \mathrm{g}(\theta^{\mathrm{T}}\mathrm{x}) = \mathrm{h}_\theta = \frac{1}{1+\mathrm{e}^{-\theta^{\mathrm{T}}\mathrm{x}}}; \tag{6.11}$$

with θ^{T} being the hypothesis *function parameters vector*.

Such a function is called sigmoid or logistic and its graphical representation is given in Figure 33.

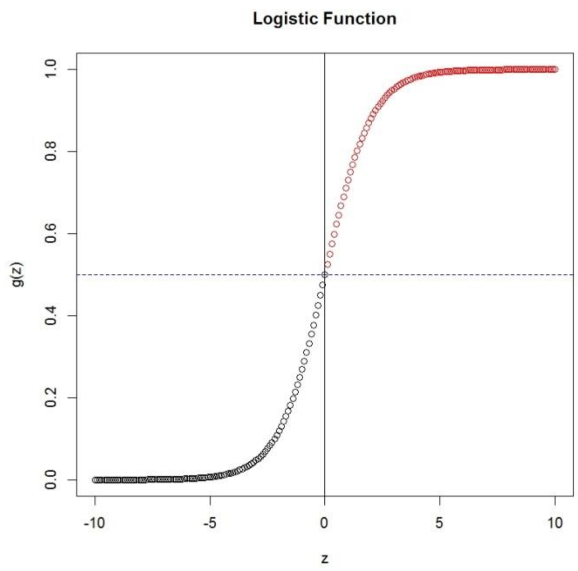


Figure 33 – Logistic or Sigmoid Function in the interval (-10,10) [Own Elaboration]

When the hypothesis function $g(z)$ outputs a number ($\hat{p}$), we treat it as the estimated probability for the sample to be equal to 1 on input x (i.e. vector of categorical values). One way of using the logistic function is: when the probability $g(z)$ is higher than a certain threshold, the model will predict (i.e. $\hat{y}$) a 1, otherwise a 0.

In Figure 33, the reader can notice a dotted line corresponding to $g(z) = 0.5$, this is the decision boundary. It sets the threshold to discriminate between ones (red dots) and zeros (black dots).

As explained before, such a threshold should be adjusted to minimize the model's cost and to choose the optimum ROC curve point. If with a 0.5 decision boundary, the model's performances are poor (i.e. low sensitivity and specificity), the threshold must be moved upward or downwards until the optimal point is found.

For instance if one wants to predict 1 with an high degree of confidence the threshold can be moved, for instance, to 0.8. This way the model will predict 1 only if $g(x) \geq 0.8$. In this case we will be more confident that 1 is a True Positive, but, at the same time, we are increasing the risk of predicting a lower number of samples as ones, increasing the FNR value.

Here the *cut-off* threshold will be chosen through, as previously explained, with a ROC curve optimization methodology.

This is displayed in Figure 34 where the line in magenta represent the distances from the ROC curve's best estimation point, with coordinates (0,1). The minimum of such a curve corresponds to the aforementioned threshold and it corresponds to 0.313.

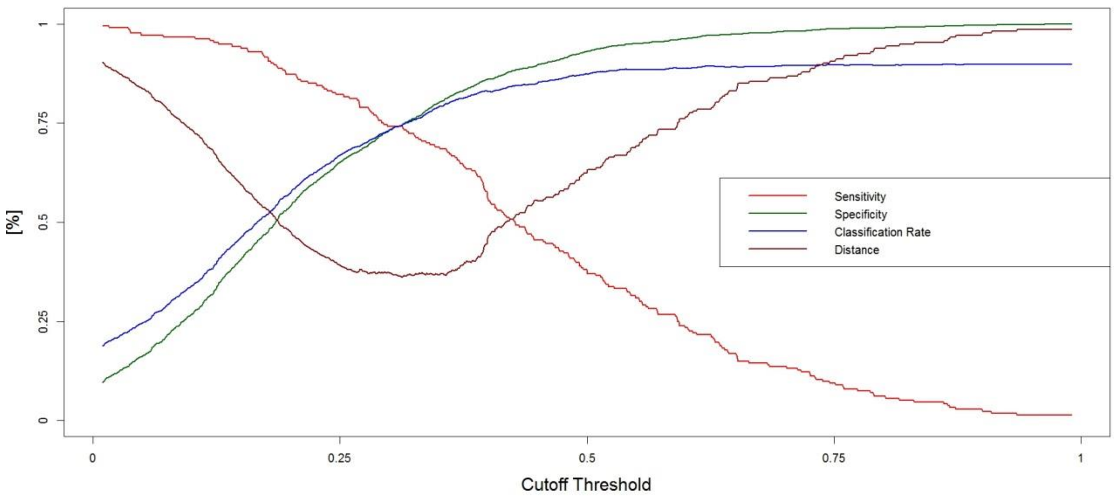


Figure 34- Optimization process for the choice of the optimal cut-off value [Own Elaboration]

After having chosen the best *cut-off*, we prove the model's performances.

TPR = 70%	*FPR = 30%*
FNR = 29%	*TNR = 71%*

Table 10 - Results of the first logistic regression using as threshold g(z)=0.313 from Precision vs. Sensitivity optimization. In the first quadrant (I): False Positives Rate, in second quadrant (II): True Negatives Rate, in third quadrant (III): False Negatives Rate, and in the fourth quadrant: True Positives Rate

In this case the model Precision is 27.4%, while the overall Sensitivity is 71%.

6.9 Model's Selection

We will now compare the models described so far. In order to do this, we will plot their Sensitivity and Fallout values in the ROC space. The curve that will be more skewed towards the upper left corner will correspond to the model with highest accuracy.

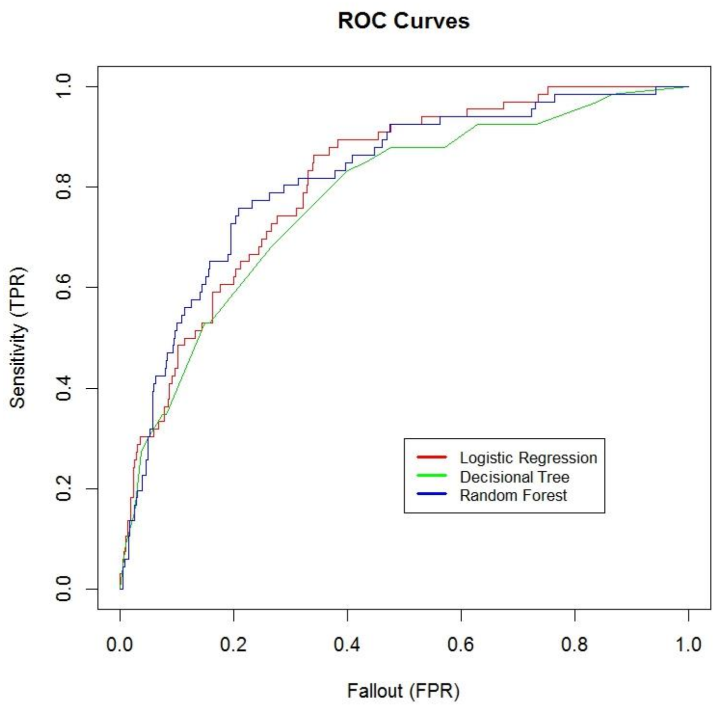


Figure 35 - ROC curves comparison between the three models computed [Own Elaboration]

Figure 35 shows that interesting conclusions can just be drawn with a graphical approach. The green curve is clearly non optimal as, for each Fallout value, corresponds a Sensitivity lower than in the other two cases. The blue curve (i.e. Random Forest) is the optimal since it is the closest to the upper left corner.

As observed, the Precision value obtained with Random Forest is 10% higher than the one obtained with Logistic Regression (choosing a threshold of 0.313).

This means that, according to ROC curves comparison, Random Forest model has to be chosen to have some insight on the variables (x_i) determining energy poverty in Spain. Moreover, it has been already observed how this methodology can be appropriate and efficient for evaluating the variables used by the model. Nevertheless, our model is not suitable for policy implementation support due to its rather low Precision.

6.10 Variables' Importance Evaluation

After the three analysed methodologies we finally want to determine which are the INE EPF variables that determine energy poverty risk. As aim of this work, this task will be initially carried out without any knowledge of sampled household's income or energy expense information.

In order to do so, a method for variables importance evaluation should be chosen. The R package `randomForest` offers specific functions for this task.

The concept used is to evaluate the *mean decrease* of the *Gini coefficient.* The latter is a measure of variable importance based on the *Gini Impurity Index* used for the calculation of splits during Random Forest's training phase.

It is at the base of Random Forests' stopping criterion, as Entropy was for Decisional Trees (see Paragraph 0).

Gini Impurity, for a binary problem, can be computed by adding the probability p of each item being correctly chosen times the probability $1-p$ of a mistake in categorizing that item:

$$G_i = p_i(1-p_i) \tag{6.12}$$

Thus, we can define both a *Gini Impurity Index* for the parental node and for the descendant nodes, respectively: G_{parent}, G_{split1}, and G_{split2}.

The split that would be selected is the one having highest *Gini Information Gain*:

$$IG_i = G_{parent} - G_{split1} - G_{split2} \tag{6.13}$$

The idea is that the *Information Gain* (IG) from parental nodes to the descendants must be positive and increase from node to node. If this is not the case the algorithm will stop, as it is no more possible to perform a further meaningful split (i.e. a split that has a positive *Information Gain*).

The Importance (I) of variable i is then calculated as the Information Gain averaged over all the splits involving the categorical variable in question (i.e. of the specific node):

$$I_i^{x_i} = G_{parent_i}^{x_i} - G_{split1_i}^{x_i} - G_{split2_i}^{x_i} \tag{6.14}$$

The good point here is that, being $I_i^{x_i}$ an *average*, the concept can easily be extended over all splits involving all the variables considered by the model. We therefore know that each variable's (x_i) Importance is an average of all the I_i of all nodes where there is a logical condition involving x_i.

The *Mean Decrease Gini* of the group would just be the mean of all the I_i weighted by the usage share of each variable (i.e. how many times the variable is considered in the Forest).

With the function `varImpPlot` it is possible to visualize the $I_i^{x_i}$ of each of the variable considered by the model, showing a variables' ranking. It displays the capability of each variable i in terms of average Information Gain. The numbers represented in Figure 36 are particularly useful in relative terms: we can determine how a variable is performing with respect to the others in explaining the phenomenon under study.

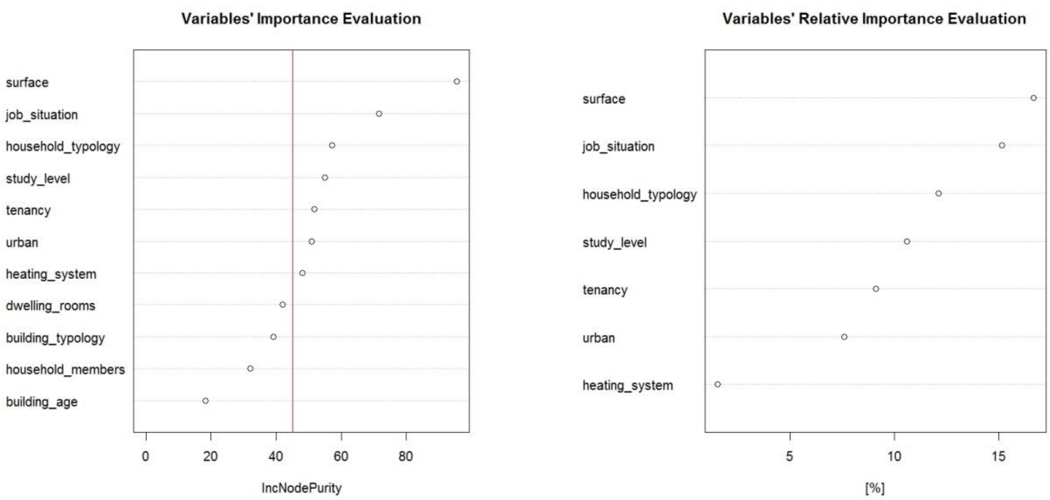


Figure 36 - Random Forest Model variables importance in relative terms and Gini Information Gain for the entire Spanish data set [Own Elaboration]

Due to the high number of variables considered, it would be interesting to determine if some of them might be excluded from the problem analysis and estimation. In order to do that we should compute model's performances, subtracting variables, one by one, and evaluating how the model's evaluation parameters (i.e. Sensitivity, Fallout…) change.

Eleven different models have been trained, starting from the complete one (i.e. considering all the variables) and then applying a *backward variables subtraction process* starting from the one having lowest *Mean Decrease Gini* value in Figure 36. Moreover, due to the fact that Random Forests have different results from run to run, the process have been repeated 100 times.

At the end, the results have been averaged determining for each one of the eleven models: Sensitivity, Precision, Fallout and Distance (see Equation 6.7).

In order to exclude some variables from our analysis we should define a model *minimum acceptability criterion*.

In previous pages (see Paragraph 6.1) we stated that, according to ROC curve optimization method, the Euclidian Distance is particularly suitable to describe model's accuracy. On the other hand, to evaluate the model's predictability power, one should control Precision, too.

Based on this consideration, we set as minimum model's standard a Precision higher than 20% [19], and a Sensitivity higher than 70%. The plot in Figure 37 shows on the x-axis the eleven models considered, while on the y-axis the values for Fallout, Sensitivity, Distance, and Precision. Model 11 represents the one that considers only the most significant variable (according to Figure 36), while Model 1 is the complete model.

The dotted lines indicates the two set minimum standards: 20% for the Precision and 70% for the Sensitivity.

We can notice that Sensitivity decreases significantly starting from Model 5, dropping by 10 percentage points per each subtracted variable. It is interesting to notice how the Precision value is above 20% for almost all the considered models. For this reason, we can state that the most stringent condition is, in this case, having a Sensitivity level higher than 70%. Sensitivity is above the standard until Model 5. We can infer that the four variables with lowest Average Information Gain values, in Figure 36, can be discarded from the analysis, as, without considering them, it is still possible to achieve results above the defined standards. In other words, they are not bringing significant improvements to model's *Information Gain*. As further proof, the blue line is keeping almost constant for the first five models, meaning that they are almost equally accurate, according to the ROC optimization.

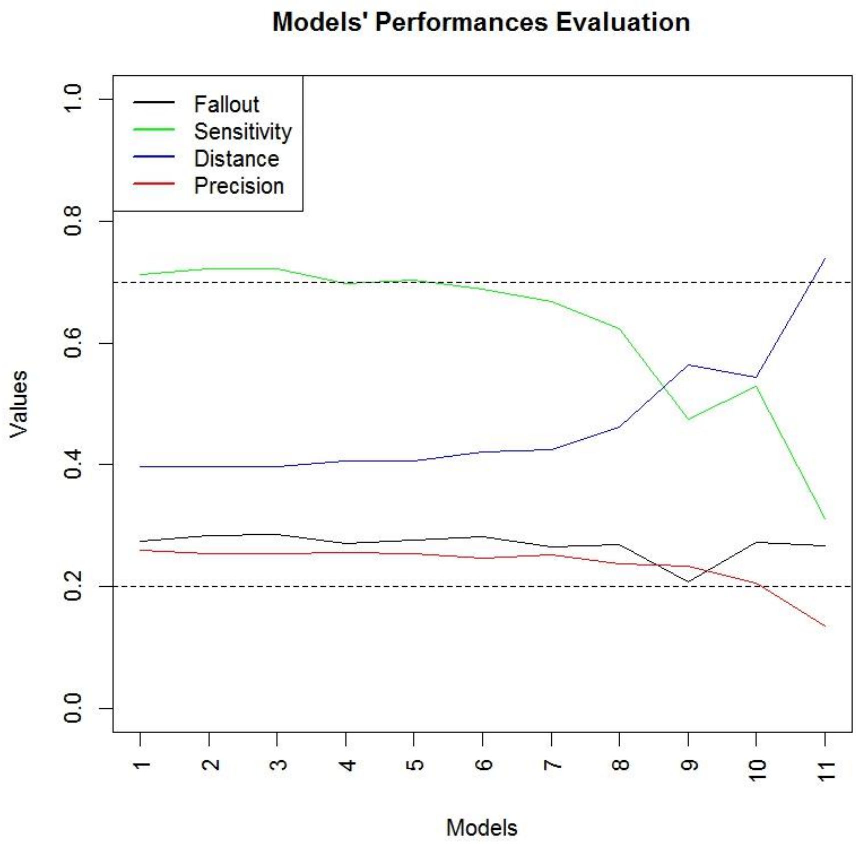


Figure 37 – Models' Performance Evaluation for variables selection [Own Elaboration]

The significant variables are: *surface*, *job situation*, *household typology*, *study level*, *tenancy*, *magnitude of the town of residence*, and *heating system typology*. They are listed, according to their relative importance in Figure 36 (b).

The reader can notice that, overall, the Precision has quite low values, reaching, as a maximum, 27.6%. This basically means that for each True Positive identified there would be almost three false alarms. For this reason, we can state that, given the set of variables used, it is not possible to achieve high Precision in terms of energy poverty predictability. However, considering the efficiencies of other policies implemented throughout Europe, our result does not appear to be so low, as it is, on average 40% higher than other European use cases [19].

Thanks to the use of the ROC Curve optimization, we are sure that the same accuracy can also be reached applying the models to other population with different imbalance ratios.

6.11 Urban Modelling Analysis

The analysis performed so far is not directly applicable to the urban context. The results achieved, in terms of variables' relative importance, have been obtained from the analysis of the whole INE EPF database.

The next step would be to apply the very same methodology to sampled households which are living in towns with more than 100,000 inhabitants. This will guarantee that the results obtained would be more explicative of the energy poverty drivers in city context and it will result in more precise and targeted results in line with this work main aim.

In cities with more than 100,000 inhabitants the energy poverty prevalence, according to π_2, is 10.1%. In this case there is a new unbalance ratio, higher than the original case (12.8%). As observed in Paragraph 6.5, this will result in a lower Precision of the model.

We demonstrated that, also considering this "new" subset, the best characterizing and predictability performances are achieved through a Random Forest analysis.

TPR = 70%	*FPR = 30%*
FNR = 27%	*TNR = 73%*

Table 11 - Results of the model (using Random Forest) trained on the urban sub-set. In the first quadrant (I): False Positives Rate, in second quadrant (II): True Negatives Rate, in third quadrant (III): False Negatives Rate, and in the fourth quadrant: True Positives Rate

As observed in Table 11, both Sensitivity and Fallout are close to the ones obtained in the original case (see Table 9).

The model is therefore proving as accurate as the original one in detecting ones among all True Positives and zeros among all True Negatives (respectively Sensitivity and Specificity). The conclusion, in terms of variables' importance, will be characterized by a *similar degree of confidence* with respect to the original case.

The *variables' importance* results for the urban case are showed in Figure 38.

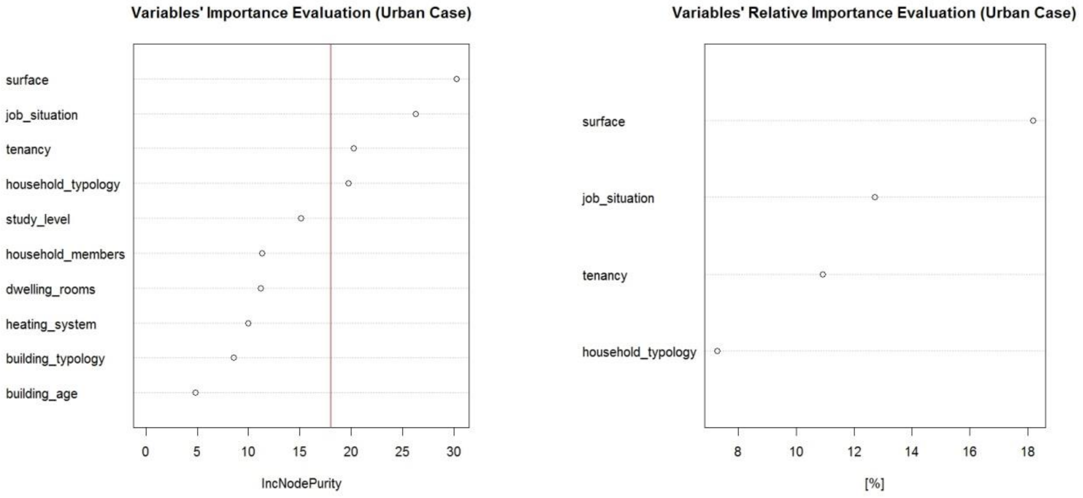


Figure 38 - Random Forest Model variables importance in relative terms and Gini Information Gain for the Spanish cities study case (more than 100,000 inhabitants) case [Own Elaboration]

Also in this case we can evaluate to eliminate some variables from the analysis, using the same *backward subtraction methodology* of the original case (see Figure 39).

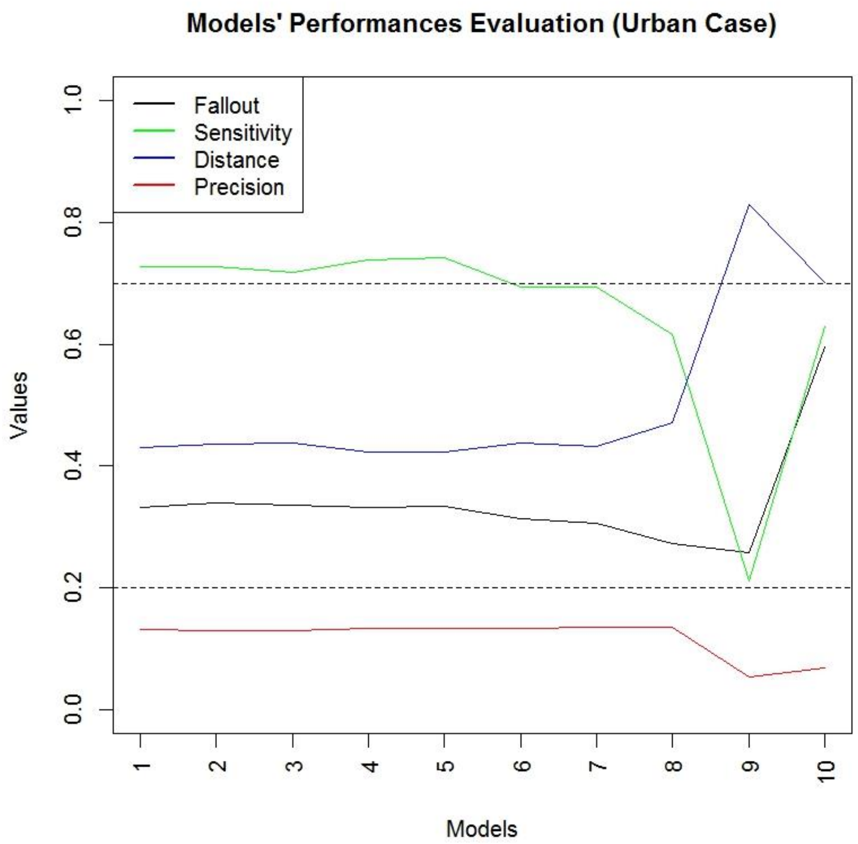


Figure 39 – Models' Performance Evaluation for variables selection for the Urban Case [Own Elaboration]

Here, the considered Models (on the x-axis) are ten since the *urban* variable has been eliminated from the computation. In this case, in fact, all the analysed families live in cities with more than 100,000 inhabitants.

Model's accuracy is expressed by the Euclidian Distance from the ROC optimization (the blue line in the graph) that is almost constant for the first seven Models considered and significantly higher for the remaining ones. Nevertheless, the overall performances of the model, concerning Precision are *unacceptably low*, signifying that relying just on the considered variables is not sufficient for achieving good model's predictability power and to support policies' implementation.

For this reason, we are now going to relax the hypothesis of not taking into account the households' *income decile* variable.

New model performances are shown in Figure 40. It is possible to notice significant improvements in terms of both Sensitivity and Precision (respectively, green and red lines in Figure 40).

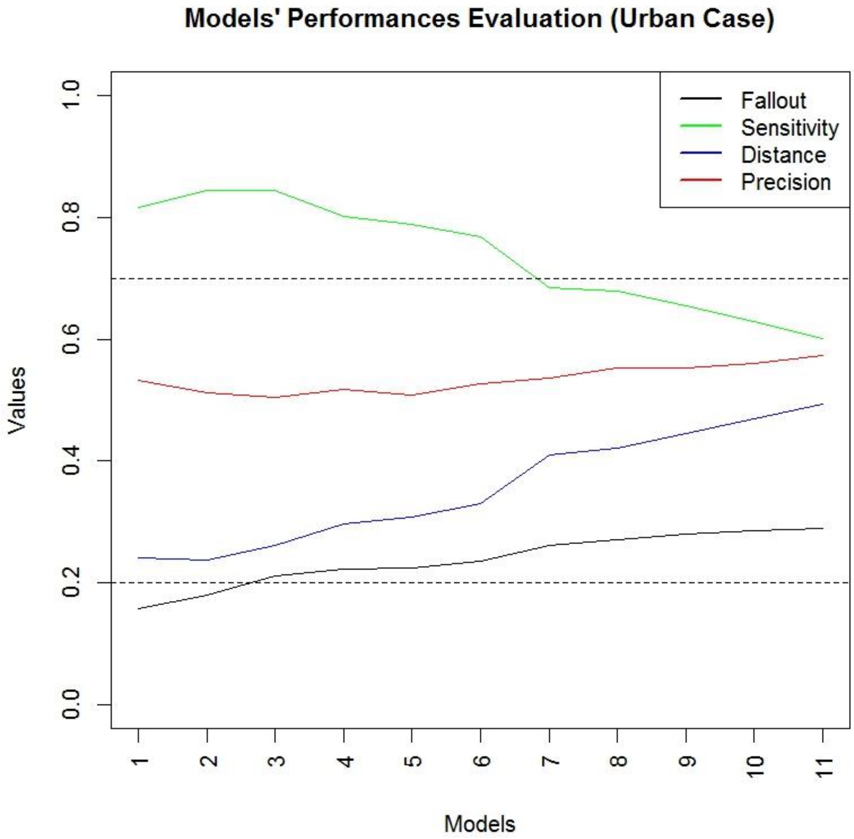


Figure 40 - Models' Performance Evaluation with the inclusion of the Income Deciles variable [Own Elaboration]

In this case, on the x-axis, eleven models' performances are reported. This is due to the fact that a further significant variable has been introduced in the computation: *household's income decile*.

The reader can notice that the Distance (blue line) is starting to steadily increase from Model 6. For this reason, according to the set requirements, we can state that the first five variables, in importance order, should be considered. This will guarantee to have a model with a Sensitivity higher than 70% and a precision above 20%. In the complete case (Model 1) the average Sensitivity is 85%, while the Precision is 56.3%.

From this analysis, for the urban case with the inclusion of the income decile variable, five are the variables to be considered (in decreasing order of importance): *income decile*, *surface*, *job situation, tenancy*, and *household typology*.

As observed in Paragraph 6.7, the variable analysis in Random Forest is mainly significant in relative terms. For this reason, the results of Figure 38 are *normalized* over the variable having maximum Average Information Gain.

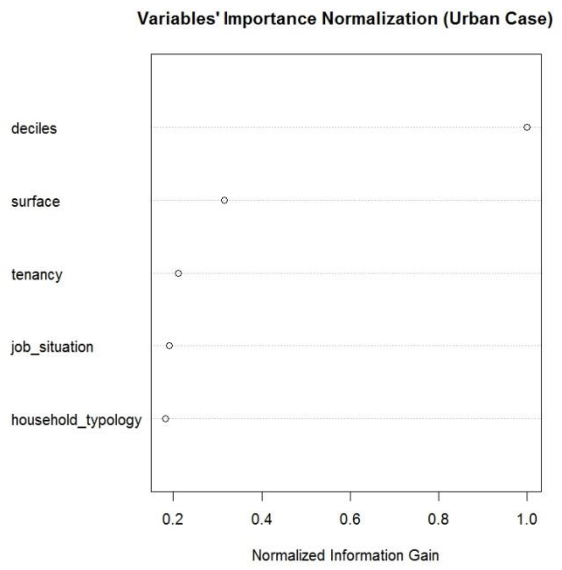


Figure 41 - Normalized Variables' Importance evaluation for the Urban Case with the inclusion of the income decile variable [Own Elaboration]

Figure 41 shows that, according to the model, the *household's deciles* variable is contributing to the Mean Information Gain as 5 times the *household typology* one.

We can finally state [29] that the values shown in Figure 41 represent the selected variables' contribution in explaining the problem, and what are their *weights* in determining an energy poverty status.

The `randomForest` package gives the possibility of evaluating which are, for each of the selected variables, the most determinant labels for having $\hat{y} = 1$. This allows to get a sense of the partial effect of each label. This is done by holding each value of the predictor of interest constant (while all the other predictors can vary at their original values), passing it to the Random Forest, and predicting the responses. The average probability of $\hat{y}$ being equal to 1 is plotted against each value of the predictor of interest. The latter can be related to what we previously defined as *cut-off* threshold. In the Random Forest simulation for the urban case the latter was set, after ROC optimization, to 0.081. In this way, $\hat{y}$ was 1 only if its probability $\hat{p}$ was higher than 0.081.

Figure 43 reports the average success probability (i.e. $\hat{y} = 1$) for each label of the five considered variables. At preliminary level, one can state that, if a label's average probability is higher than the *cut-off* threshold, a label can be considered relevant to identify the sample family as energy poor.

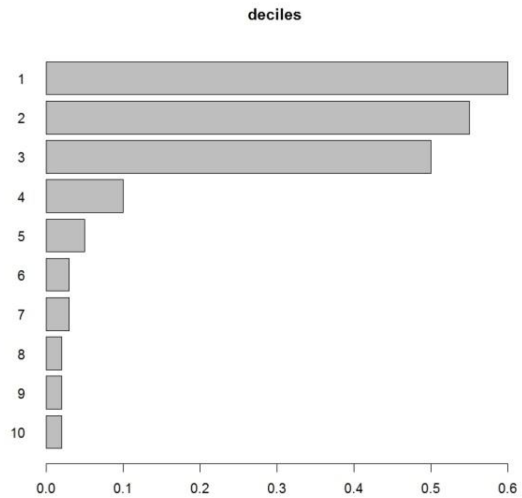


Figure 42 - Partial Dependence plots of the income decile (categorical) variable for the study of the energy poverty problem in Spanish cities with the inclusion of the income decile variable [Own Elaboration]

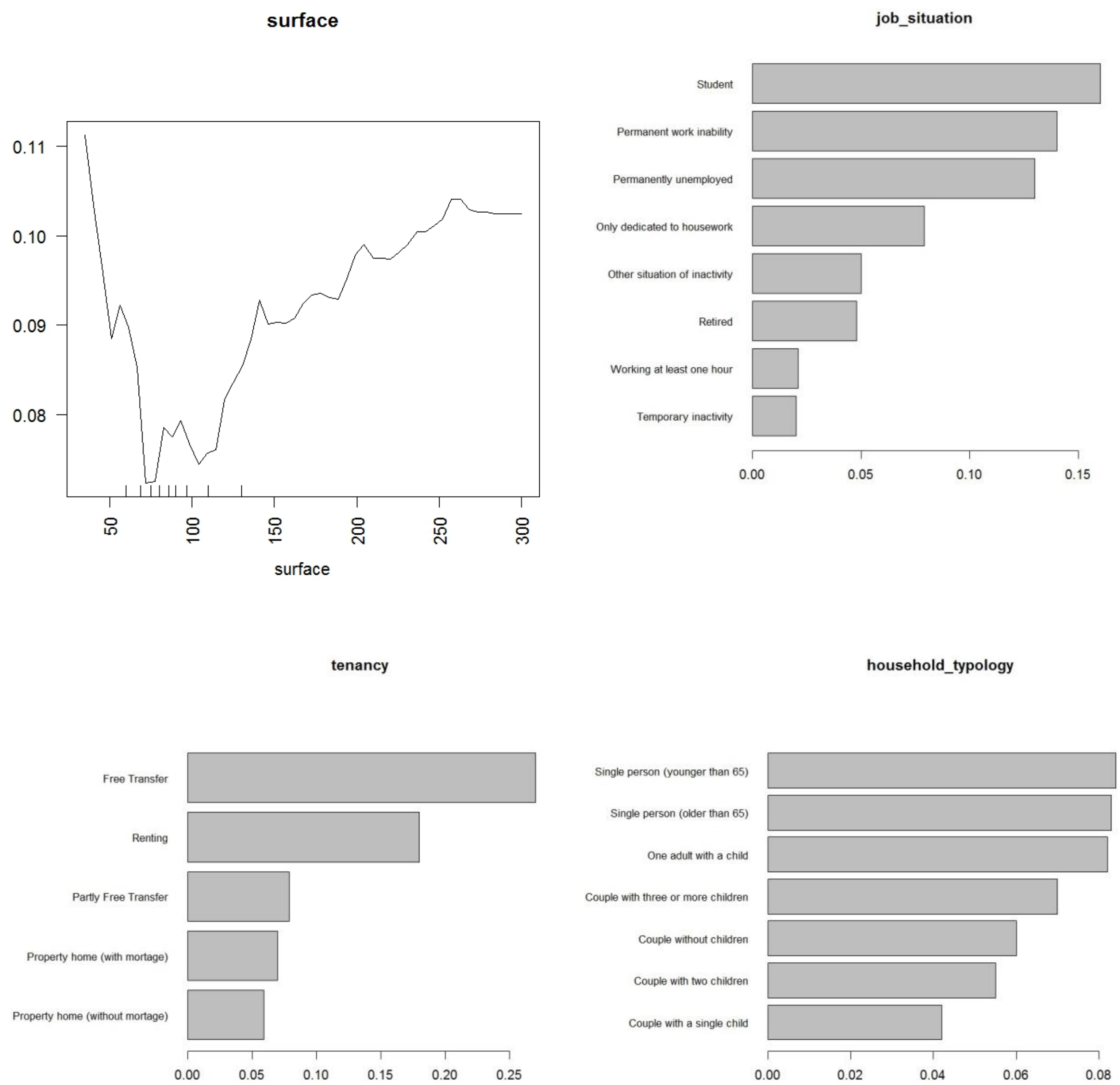


Figure 43 – Partial Dependence plots of surface (continuous), job situation (categorical), tenancy (categorical), and household's typology (categorical) variables for the study of the energy poverty problem in Spanish cities with the inclusion of the income decile variable [Own Elaboration]

The reader should remember that the second variable (*surface*) is continuous while the remaining categorical.

Concerning the *job situation*, three are the labels at major risk: households where the main family member is a student, unemployed or permanently unable to work.

There is evidence that the renting condition is also playing a major role in increasing the probability for energy poverty in the urban context.

On the other hand, three household's typologies labels have average success probabilities higher than the cut-off threshold: single person (younger than 65), single person (older than 65), and one adult with a child.

6.12 Summary

Chapter 6 has applied three statistical and machine learning tools to the INE EPF database for the year 2014.

Two important results have been obtained:

- Random Forest, among three tested machine learning algorithms, has shown most suitable performances either in terms of model's Accuracy and Precision.
- Five variables have been selected for characterizing energy poverty households for the Spanish urban context and can be applied to a specific city application.

Chapter 6 has given the tools for evaluating energy poverty throughout Spanish cities. Due to the inherent lack of data that characterize the subject of this work we will add a further hypothesis, considering that the results of Chapter 6 could be applied to Barcelona specific case.

In the following Chapter, we will use the obtained information to estimate energy poverty conditions in city's neighbourhoods (*barrio*). The analysis will be only qualitative and its scope would be to identify which parts of the city are facing higher risk according to the proposed model and where public authorities should start to

7 Results Application

The purpose here is to demonstrate how the results obtained in Chapter 6 can be used by public authorities to evaluate the energy vulnerability issue in a specific real-case application. The aim of this work is to go beyond the common practice of considering energy vulnerability as only typical of poorest (in monetary terms) neighbourhoods, but to use the information obtained, in terms of variables importance, for reaching *more systematic and targeted* results.

It is clear that, to keep implementation costs to a minimum, it is desirable and useful for public authorities to select a limited amount of significant variables to be considered and what are the parts of the city where to focus their attention on.

This Chapter proposes an *active energy poverty detection process*. This means that, using such an approach, vulnerable groups can be found in the areas of the city where the identified conditions, according to the model trained in Chapter 6, are most common with respect to city's average risk.

The logic that lies behind this methodology is to individuate particularly vulnerable groups characterized by lower than average imbalance ratios (i.e. high number of fuel poor). Formally, we are looking for ensembles where the number of $y = 1$ over the total sampled households is particularly relevant. This will allow, by achieving higher Precision, to use the model for quantitative policy support at urban level.

The process is described in Figure 44.

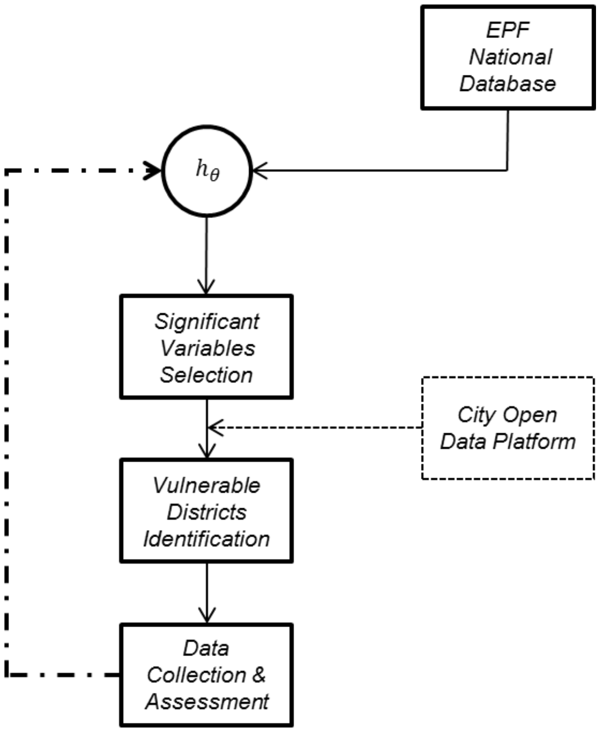


Figure 44 – Result's implementation framework and refinement for the active energy poverty detection process [Own Elaboration]

The Model implemented and described in the previous Chapter (i.e. h_θ) responded basically to two major requirements: determine the importance of the used variable and determine their contribution to Model's performances.

In the city context of Barcelona, it is not possible, as far as this work is concerned, to retrieve a disaggregated data set in line with the EPF by INE. For this reason we are not directly capable of replicating the approach used in Chapter 6. However, it is possible to combine the results obtained for Spanish cities with the information available on the Barcelona Open Data Platform. As in Figure 44, the two results combined will identify particularly vulnerable districts and neighbourhoods.

The methodology will allow to focus policy makers' efforts and financial resources on specific areas of the city. Moreover, public authorities would be encouraged to start and promote more targeted statistical surveys, significantly cheaper as focusing only on specifically selected areas and on certain important variables.

Once collected and evaluated the new data, the Model h_θ can be trained, refined and adapted to city's specific case. At this point the process can restart with the selection of a new suite of significant variables.

After the definition of the problem and the discussion of the proposed strategy we are going to present a practical application of what, in Figure 44, is defined as "*Vulnerable Districts Identification*".

7.1 The Barcelona Case

The purpose is to evaluate how *income decile, surface*, *job situation*, *tenancy*, and *household's typology* variables are distributed across city districts, highlighting situations of major *deviation* from city's average values.

The analysis will focus on the variables' labels selected (see Figure 43), where the average probability for the model (Random Forest based) of giving $\hat{y} = 1$ was given.

It is important to further specify that this is a major and important *hypothesis* since we are using the results obtained by an analysis conducted at Spanish level and not specifically centred on the city of Barcelona for the inherent lack of available data.

Each city district will be assigned with the percentage (l) of households living in the conditions described by the selected *vulnerability labels*. Those are shown in Table 12:

Variables	Vulnerability Labels (l)
surface	Rate of households living in dwellings with floor area lower than 60 m^2 and higher than 150 m^2.
tenancy	Rate of renting households or that has received the dwellings through a free transfer (e.g. will).
job_situation	Rate of households where the main member is a student, is facing unemployment or permanent work inability conditions.
household_typology	Rate of households composed by a lonely person (older or younger than 65) or by a single adult with a child.

Table 12 - Variables label selection for selecting the data from the Barcelona Open Data platform

It is not possible to get information about the percentage of households' according to their income deciles. However, the *Renta Familiar Disponible* (RFD) index is assigned to all city's neighbourhoods. The latter assesses the average family's income available for expenditures and/or savings. It is expressed in relative terms where the city average is set to 100. We will therefore evaluate, for each single *barrio*, how its RFD ratio is displaced from city average to get an insight over the average households' income level in that specific part of the city.

Thus, for each of the considered five variables, a deviation D_i is then calculated. It represents the neighbourhood's displacement from city average and it is calculated as:

$$D_i = \frac{l - \bar{l}}{\bar{l}} \cdot 100\ [\%] \tag{7.1}$$

In Equation 7.1, l indicates the neighbourhood's vulnerability rate (see Table 12) for a specific variable, while $\bar{l}$ indicates the city average. For each neighbourhood we will calculate five *deviations*, as the number of the selected variables.

At this point, from the urban case modelling (see Paragraph 6.10), we know which is, in *Information Gain* terms, each variable's partial contribution (i.e. *importance*). This has to be interpreted as the average decrease in Gini coefficient (see Equation 6.12) from a paternal node to its derivate. Thus, one can state that *household's income*, the variable with highest average Information Gain, is the best in splitting the dataset in two different subsets (i.e. derivate nodes) whose Gini Coefficient would be significantly lower than their paternal nodes. For this reason, we can assign, to each of the five selected variables, a *weight* (ε_i) to evaluate variable's ability in splitting the data in two heterogeneous ensembles. The *weights* will correspond to the normalised *importance* values obtained from the Random Forest model applied to the Spanish urban case (see Figure 41).

At this point it is possible to aggregate all the five deviations for all city neighbourhoods as a *weighted mean*:

$$D_{tot} = \frac{\varepsilon_{RFD} \cdot D_{RFD} + \varepsilon_{surf} \cdot D_{surf} + \varepsilon_{job} \cdot D_{job} + \varepsilon_{ten} \cdot D_{ten} + \varepsilon_{hous} \cdot D_{hous}}{\varepsilon_{RFD} + \varepsilon_{surf} + \varepsilon_{job} + \varepsilon_{ten} + \varepsilon_{hous}} \tag{7.2}$$

A positive D_{tot} indicates that the families living in the analysed neighbourhood have, on average, an higher probability of being identified as energy vulnerable (i.e. $\hat{y} = 1$) with respect to city's average risk, according to the model proposed and its approximations.

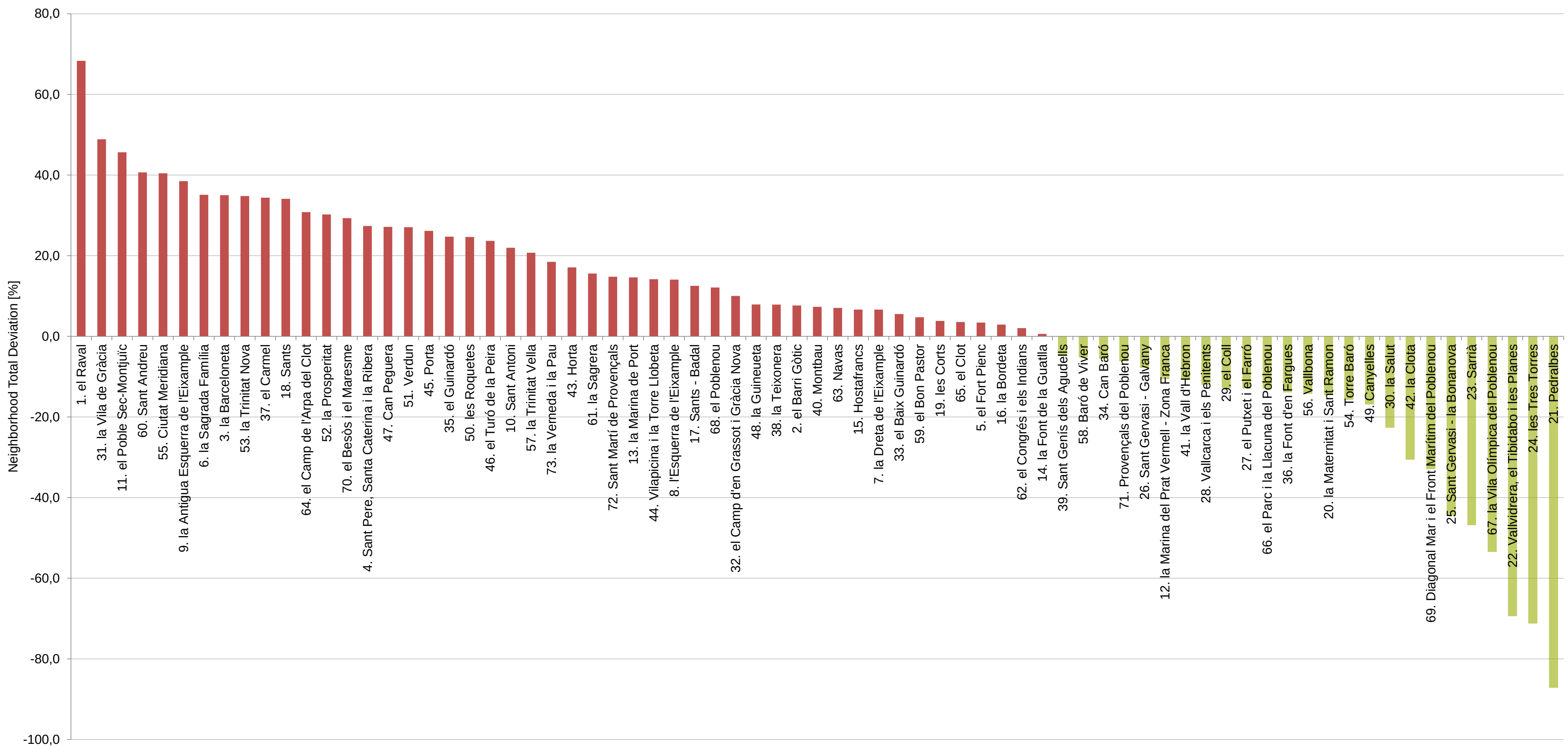


Figure 45 – Barcelona's neighbourhoods at major energy vulnerability risk, according to national urban modelling. The picture shows neighbourhoods' total deviations (D_{tot}) as reported in Equation 7.2 [Own Elaboration]

Figure 45 shows the results, in terms of neighbourhoods *total deviations* from city' average. It represents a first attempt of vulnerable neighbourhood identification, according to the Spanish urban case modelling. The ranking in the plot should not be interpreted as an energy vulnerability absolute measurement across Barcelona *barrios*, but as a demonstrated and objective suggestion to policy makers. They should, according to our model's results, start by collecting data and analysing the situation in the identified neighbourhoods to fulfil the last "block" of Figure 44: "*Data Collection & Assessment*". Once the latter is done, a new Model can be trained based on more targeted data providing a more precise and coherent evaluation of energy poverty in the city of Barcelona. However, from the neighbourhoods' ranking displayed in Figure 45 all the energy poverty attention offices of the Barcelona city councils are located in *barrios* identified as risky by our model (red columns in Figure 45). Those are: *la Marina de Port*, *San Martí de Provençals*, *Sant Andreu*, and *el Turó de la Peira.* According to La Vanguardia (2016) they have been located by the City Council (see Annex A and G), in *barrios* with the highest number of forced disconnections.

The trained model and the active identification methodology can also contribute at identify which are the partial contribution to the energy poverty risk case by case.

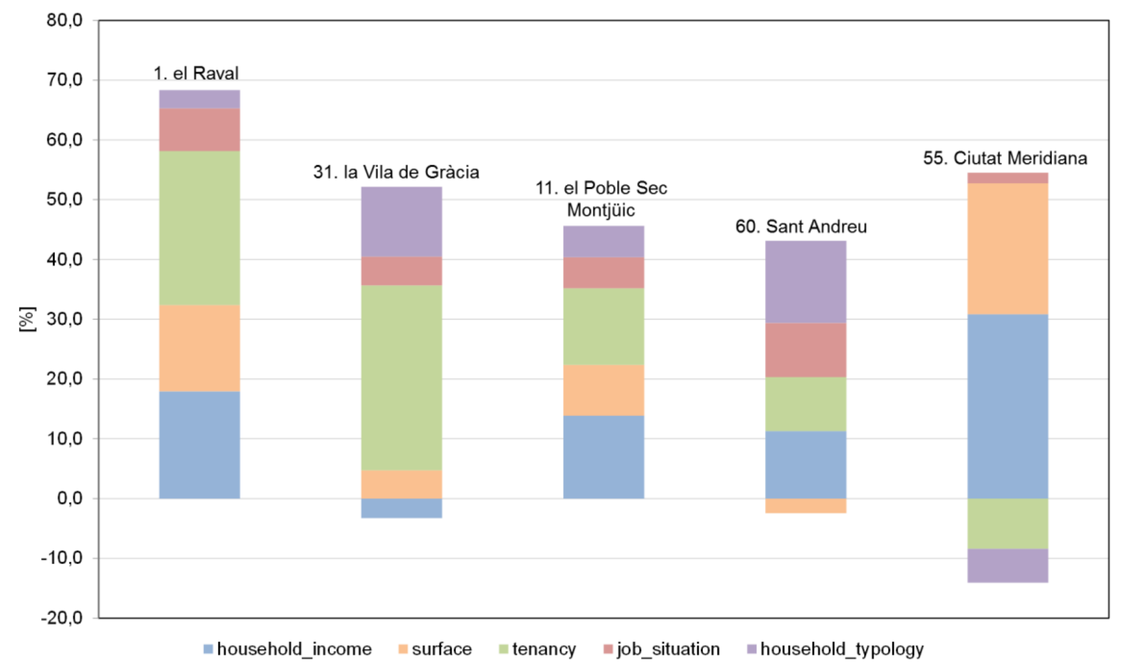


Figure 46 – Partial contribution of the considered variables (RFD, job situation, surface, tenancy, and household typology) to total neighbourhoods' deviations [Own Elaboration]

The five variables contributions to total neighbourhood's deviation are shown in Figure 46 and represents Equation 7.2 summation.

If one variable partial contribution is positive, it means that the risk for energy poverty, related to that specific variable, is higher than the city average. On the contrary, if negative, it means that the variable is not contributing to the *barrio* energy poverty risk, but is decreasing the overall probability for it.

In the case of *Ciutat Meridiana*, for instance, the reader can notice that the major component to neighbourhood's risk is related to household's *income* conditions and to *surface* (significant share of dwellings smaller than 60 m^2 and bigger than 150 m^2).

The situation is different for *la Vila de Gràcia*, where the most significant contribution is given by tenancy (majority of renting households') and by household typology (number of households belonging to the vulnerable labels identified in Figure 43) [8, 13].

The information obtained from Figure 46 can be useful to understand what are the drivers that determine energy vulnerability situations in different parts of the city in a systematic way, allowing policy makers to choose more targeted measures.

In Paragraph 3.4 some policies and "best practices" have been analysed, underlining the essential difference between *short-term* and *long-term* solutions.

For instance, in the case of *Ciutat Meridiana*, where the main identified driver is household's income, it would advisable to address the problem through social and financial oriented policies in a *short-time* horizon. In this case, referring to Table 1, a policy maker should reduce the impact of energy products' prices or use strong "canonical" social actions.

On the contrary, energy efficiency improvements might be needed in *la Vila de Gràcia* and in *el Raval*, as it has been proved that low energy efficiency standards are directly linked to a renting tenancy status [8, 13].

8 Project's Budget

The project's budget has been calculated considering the total amount of worked hours, the licenses of the software used, and the travel expenses.

Work Typology	Dedicated Time [hours]	Cost [€/hour]	Total Cost [€]
Energy Poverty Literauture Review	85,0	25,0	2.125,0
R Programming	151,0	25,0	3.775,0
Results Elaboration	170,0	25,0	4.250,0
Software Licenses			
R Project for Statistical Computing			0,0
Microsoft Office 2010			400,0
Travel Expenses			
Barcelona Urban Transport			189,3
Project Total Cost (*before taxes)*			10.739,3
Project Total Cost (*including 21% VAT*)			12.994,5

Matteo Farinoni,

Barcelona, 07/07/2016

The estimated budget will be valid for one month after the reported date of delivery of the project..

9 Conclusions

Energy poverty is a phenomenon driven by technical and economic causes that results in major social deprivation and health effects. This work has developed an evaluation framework and a tool for supporting policy makers in identifying what are the vulnerable groups to be considered and how to identify them considering the available statistical data.

The European Union recognised the problem's priority asking, in 2009, to all Member States to start policies for identifying and solving the issue. Spain has not an official definition of energy poverty, yet.

It must be stressed the fact that this work has faced and tackled a very actual issue which is at the centre of both Spanish and Catalan political and economic discussion.

In the first part the drivers and causes of the problem have been explored in detail, concluding that, for the Spanish case, the combination of rising energy price and economic and financial instability has considerably contributed to the increase of the number of families experiencing energy poor situations. The study has taken into consideration all the measurement tools (i.e. *indicators*), at European level, with the purpose of identifying strengths and weakness for each case. The indicators belonging to the LIHC family, currently applied and used in the United Kingdom, show particularly suitable features for being applied to the Spanish case. Moreover, a new and innovative indicator, strongly centred on *energy efficiency*, has been proposed and tested.

Thanks to the official database (*Encuesta de Presupuestos Familiares*) provided, for the year 2014, by the Spanish Statistical Institute, the work has been able to determine what is the number of households experiencing energy poverty conditions. According to the LIHC indicator (π_2) the percentage of needy family is Spain was around 12.8%, in 2014. In Catalonia, it was close to 11.4%.

It has also been possible to demonstrate that the number of non-poor families, in monetary terms, affected by energy poverty, in Catalonia, is higher than the national average due to a substantial expense on energy product. This demonstrates that the *energy efficiency* driver is essential, besides household's income and energy prices, for distinguishing energy vulnerable families.

The problem has also been modelled, with a completely new and innovative approach, applying three machine learning instrument to the original data set. In this part the energy expenditure variables has not considered with the scope of implementing a tool fully independent of utilities or privately owned information. This is a recognised need for many public authorities in order to face the information asymmetry existing between private and public entities and companies.

The main aim has been to evaluate and quantify which are the variables that determine an energy poverty situation, highlighting their *weights* in driving the phenomenon in Spanish cities. Five variables have been proven particularly significant in this sense: the household's income, the dwelling's surface, the tenancy status, the job situation, and the typology of the households. This gives to policy makers the possibility of objectively quantifying the issue, while optimizing and prioritizing governmental financial resources. It has been demonstrated that current policies are not showing suitable features and targeting precision.

The study has combined this results, to obtain a practical application for the city of Barcelona. As data source, the city Open Data Platform has been considered in full detail to assess the issue in the city. The characteristics of each *barrio* (i.e. city's neighbourhood), according to the considered variables, have been compared with city's average level and it has been possible to draw up a ranking of the areas that are most likely to be affected by energy poverty, according to the trained model. The same approach allows also to identify what are, case by case, the driver of the phenomenon, increasing considerably policy makers' targeting ability and effectiveness.

This study has been carried out with the scope of proposing to a local public authority a tool for assessing and studying the problem in a specific city. The methodology and approach used in this work can set the base for real and effective social, financial, and *energy efficiency* policies in the European perspective. Thanks to the possibility of ranking neighbourhoods according to their energy poverty vulnerability levels, the results of the study can be further improved and refined as more detailed and targeted are collected.

10 Bibliography

[1] European Commission, “Energy Union and Climate,” European Commission, 2016. [Online]. Available: http://ec.europa.eu/priorities/energy-union-and-climate_en. [Accessed 2016 May 16].

[2] International Energy Agency (IEA), Energy Policies of IEA Countries - European Union, Paris, 2014.

[3] EUROSTAT, 2016. [Online]. Available: http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:EU_statistics_on_income_and_living_conditions_(EU-SILC). [Accessed 23 April 2016].

[4] Building Performance Institute Europe, Alleviating Fuel Poverty in the European Union, Brussels, 2014.

[5] Öko-Institut, “How to end energy poverty? Scrutiny of current EU and Member States instruments,” Brussels, 2015.

[6] D. Healy, Housing, Fuel Poverty and Health: A Pan-European Analysis, Aldershot: Ashgate, 2004.

[7] S. Bouzarovski, “Energy poverty in the European Union: landscapes of vulnerability,” *Energy and Environment,* vol. 3, no. 3, 2014.

[8] INSIGHT_E, “Energy poverty and vulnerable consumers in the energy sector across the EU: analysis of policies and measures,” 2015.

[9] Instituto para la Diversificación y Ahorro de la Energía (IDAE), «Análisis del consumo energético del sector residencial en España,» Madrid, 2011.

[10] EUROSTAT, “Energy Price Statistics,” May 2015. [Online]. Available: http://ec.europa.eu/eurostat/statistics-explained/index.php/Energy_price_statistics. [Accessed 06 April 2016].

[11] European Commission, “Quarterly Report on European Electricity Markets,” Market Observatory for Energy, Brussels, 2015.

[12] Asociación de Ciencias Ambientales (ACA), «Pobreza Energética en España - Analisis de Tendencias,» Madrid, 2014.

[13] EPEE project, "European Fuel Poverty and Energy Efficiency," Intellingent Energy Europe, 2009.

[14] Asociación de Ciencias Ambientales, «Pobreza, vulnerabilidad y desigualdad energética,» Madrid, 2016.

[15] J. Hills, "Fuel Poverty: the problem and its measurement," London, 2011.

[16] D. Healy, "Excess winter mortality in Europe: a cross country analysis identyfing key risk factors," *Journal of epidemiology and community health,* 2003.

[17] S. Roberts, "Energy, equity and the future of the fuel poor," *Energy Policy,* 2008.

[18] J. Robine, "Death toll exceeded 70,000 in Europe during the summer of 2003," *Comptes rendus biologies,* pp. 171-178, 2003.

[19] I. Faiella, "La povertà energetica in Italia," *Questioni di economia e finanza,* 2014.

[20] J. Romero, «Pobreza Energética en España - Análisis económico y propuestas de actuación,» Economics for Energy, Vigo, 2014.

[21] D. Ürge-Vorsatz, "Building synergies between climate change mitigation and energy poverty alleviation," *Energy Policy,* 2011.

[22] BREDEM , "A technical description of the BRE Domestic Energy Model (Version 1.1)," Watford, 2012.

[23] J. Hills, "Getting the measure of fuel poverty: Final Report of the Fuel Poverty Review," London, 2012.

[24] R. Moore, "Definitions of fuel poverty. Implications for Policy," *Energy POlicy,* vol. 49, pp. 19-26, 2012.

[25] Instituto Nacional de Estadística (INE), «Encuesta de Presupuestos Familiares. Base 2006,» 2014. [En línea]. Available: http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176806&menu=ultiDatos&idp=1254735976608. [Último acceso: 1 March 2016].

[26] C. Thomson, "Energy Poverty in the EU," 2013.

[27] Stanford University, [Online]. Available: https://www.coursera.org/learn/machine-learning. [Accessed 14 March 2016].

[28] S. Shih, "Random Forests for Classification Trees and Categorical Dependent Variables: an informal Quick Start R Guide," *Stanford University,* 2011.

[29] G. Louppe, “Understanding Random Forests - from theory to practice,” Liège, 2014.

[30] Instituto Nacional de Estadistica, «Encuesta de Presupuestos Familiares - Fichero de Usuario,» Madrid, 2014.

[31] G. Federico, “The Spanish Gas and Electricity Market,” IESE, 2010.

[32] Ajuntament de Barcelona, “OpenDataBCN,” [Online]. Available: http://opendata.bcn.cat/opendata/es. [Accessed February-June 2016].

Energy Poverty:

Measurement Strategies and Solutions

ANNEXES

Master Thesis defended by:

Matteo Farinoni

Universitat Politècnica de Catalunya

KIC InnoEnergy, MSc Energy for Smart Cities

Directors:

Andreas Sumper

CITCEA – Universitat Politècnica de Catalunya

Adriana Anguera Jordà

everis

July 2016

A - Context and Motivation

CincoDías

El bono social, para el que lo necesita

17-05-2016 07:24

El debate de la pobreza energética, como el de la exclusión de la vivienda, era necesario en este país, pero no siempre se enfoca bien. Hay familias incapaces de pagarse la calefacción en invierno, lo que resulta insoportable, del mismo modo que el drama social de los desahucios estaba pasando desapercibido hasta que una movilización ciudadana lo puso en el primer plano del debate público. Es cierto que hay situaciones de emergencia social que las autoridades deben atender. Tuvo que llegar la más grave crisis en décadas, y recrudecerse la protesta social, para que los grupos políticos y los gestores públicos se dieran cuenta.

Ahora bien, dando por sentada esta sensibilidad hacia los problemas de los más desfavorecidos, es sabido que no todas las personas que dejan de pagar la luz o la hipoteca lo hacen por no tener recursos suficientes. No tiene lógica ni casaría con la seguridad jurídica que dejar de pagar por cualquier motivo implicara sistemáticamente que el moroso siga recibiendo el servicio o bien contratado. La clave, entonces, estará en medir bien quién necesita de la protección legal y quién no.

Un debate vinculado a la pobreza energética es el del bono social, un descuento en la tarifa eléctrica, que Endesa ha llamado a replantear. El actual modelo beneficia de forma indiscriminada a distintos colectivos:todos los que contraten una potencia inferior a 3 kW (aunque tengan un estudio de lujo);todas las familias numerosas (aunque sean millonarias);los que cobran la pensión mínima o familias con todos sus miembros en paro. Sin discutir la cobertura a estos dos últimos colectivos, la eléctrica considera que se benefician del bono social medio millón de hogares con rentas altas. Lo que propone es vincular el bono a la renta, crear un fondo que se haga cargo al 100% del suministro a los verdaderos necesitados (identificados por los servicios sociales de los ayuntamientos), y una tercera iniciativa para mejorar la eficiencia energética de las viviendas. Además, la compañía pide que se ponga orden en las competencias sobre este asunto entre las distintas Administraciones.

La de Endesa es una propuesta interesante para un debate necesario, aunque en términos políticos resultaría delicado para cualquier Gobierno revisar los beneficios para pensionistas o familias numerosas. Otro elemento para el debate:en los colectivos que hoy pueden acogerse al bono social no figuran personas efectivamente pobres, como serían familias no numerosas que dependen de empleos precarios y mal pagados. Tiene sentido estudiar cómo vincular esta ayuda a la situación real. Otro asunto es cómo se reparten los costes de estos mecanismos de solidaridad entre los consumidores, las compañías y la Administración:entre los tres parece más que asumible.

http://cincodias.com/cincodias/2016/05/16/empresas/1463426211_224187.html

CincoDías

Electricidad

El 80% de los usuarios en pobreza energética carece del bono social

- Endesa propone a Industria y la CNMC un modelo más equitativo
- Supondría dejar sin bono eléctrico a 700.000 de los beneficiarios actuales, hasta 1,8 millones

17-05-2016 07:24

En su última junta de accionistas Endesa informó que había elaborado una propuesta de bono social de la tarifa eléctrica que iba a remitir a Industria y la CNMC. La compañía avanzó que su propuesta ligaba el derecho a un descuento en la factura de la luz a los ingresos familiares y subrayaba la necesidad de utilizar el actual fondo de eficiencia energética en el aislamiento de las viviendas.

En el informe, al que ha tenido acceso **CincoDías**, la compañía hace suya una parte de la propuesta de real decreto del Gobierno sobre el bono social (una norma de la vasta reforma eléctrica que nunca se aplicó), de mantener los actuales criterios que dan derecho al bono eléctrico, pero (esta era la novedad de la nonata propuesta del PP) **siempre que no se superasen ciertos umbrales de renta**. Dichos criterios son los mismos que estableció el exministro de Industria Miguel Sebastián, **inventor del bono**: ser familia numerosa;tener más de 60 años y pensión mínima; menos de 3 kW de potencia contratada o ser una familia con todos los miembros en paro.

En esta línea, Endesa propone mantener el bono social para titulares de pensiones mínimas y familias con todos los miembros en paro, pero fija **límites de renta a las familias numerosas y los usuarios con menos de 3 kW** de potencia, que suponen un 75% del total. El coste del bono (que podría ser un descuento del 25% en el precio regulado de la energía (PVPC) o un 60% del término fijo de los peajes, lo que incentivaría el ahorro) deberían financiarlo, según Endesa, el Estado y el resto de consumidores en la tarifa, y no las eléctricas como ocurre ahora.

Además, para los clientes en verdadera situación de pobreza, plantea crear otro fondo para **costearles el 100% de la factura** (se financiaría igual que el anterior y participarían los Servicios Sociales de los ayuntamientos), además de un tercer fondo de eficiencia (200 millones de la tarifa)para mejorar el aislamiento de las viviendas de estas familias, que gestionaría el IDAE.

La nueva definición del bono social parte de la base de que, al no estar ligado a la renta, el actual lo reciben familias **no precisamente vulnerables** (aproximadamente, medio millón). Endesa, que tiene firmados un centenar de convenios con comunidades autónomas y ayuntamientos que dan cobertura a 11,7 millones de hogares de 27 provincias, estima que del número de hogares que acude a los servicios sociales, **casi el 80% de los mismos no tiene bono social**. A falta de estadísticas, según extrapola Endesa, este tipo de clientes sumaría 116.600 con una facturación de 18,9 millones y una medio en los recibos aplazados de 230 euros.

El bono de Endesa, si bien resulta más equitativo, también **es mucho más restrictivo**. De hecho, reduciría el actual número de beneficiarios de **2,5 millones a 1,6 millones** (los que tendrían derecho al descuento) que se sumarían a **los 161.000 en pobreza extrema** a los que se financiaría el 100% del recibo. También el coste se recortaría de los 188 millones anuales del actual modelo, que pagan las eléctricas verticalmente integradas, a 150 millones. Pese a la crisis, el número de beneficiarios del bono eléctrico ha caído un 18% entre 2009 y 2014, en parte por el trasvase de clientes al mercado libre, donde desaparece el derecho al bono.

http://cincodias.com/cincodias/2016/05/16/empresas/1463421027_796981.html

EL PAÍS

27 MAR 2014 - 20:55 CET

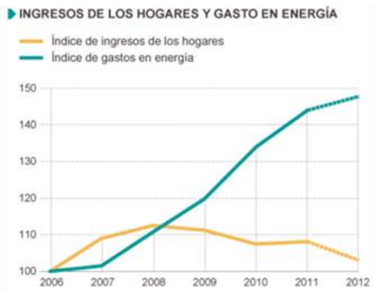
La pobreza energética se dispara

La tasa de familias que destinan más del 10% de su renta a la luz y el gas alcanza el 16,6%
El 9% no puede mantener una temperatura adecuada

El número de españoles que pueden estar en riesgo de pobreza energética ha aumentado en dos millones en solo dos años. Es la conclusión más destacada que ofrece la segunda edición del estudio de referencia en España, editado por la Asociación de Ciencias Ambientales (ACA), que actualiza datos de 2010 a 2012. Según el principal indicador del informe —basado en la metodología instaurada por Reino Unido, país pionero en el análisis de este problema—, el porcentaje de hogares que tienen que destinar una cantidad desproporcionada de sus ingresos a pagar facturas de luz y gas subió en 2012 hasta el 16,6%, lo que supone unos siete millones de personas, frente al 12,4% registrado en 2010, equivalente a cinco millones. Esto se traduce en familias que pasan frío en invierno y calor en verano, viviendas con moho y humedad, cortes de suministro por impago (1,4 millones en 2012, más del doble que en 2006, según cálculos de este periódico), menos dinero para satisfacer otras necesidades básicas y, lo más grave, muertes prematuras en invierno. Hasta 7.200 fallecimientos podrían evitarse si se erradicara el problema, según el sistema de medición de la Organización Mundial de la Salud.

La culpa de este aumento la tienen sobre todo dos fenómenos paralelos: mientras los ingresos de los españoles se reducen por la crisis, el precio de la energía se dispara. La factura de la luz, principal responsable de este aumento, subió un 60% desde 2007, mientras las rentas bajaron un 8,5%, según el Instituto Nacional de Estadística (INE). En consecuencia, los hogares deben dedicarle un porcentaje cada vez más alto de sus ingresos. De una media del 4,3% en 2007 se ha pasado al 6,5% en 2012. "Siguiendo la metodología británica, consideramos que el porcentaje empieza a ser desproporcionado cuando supera el 10%. Ahí es donde empieza el riesgo de pobreza energética porque pueden aparecer dificultades para cubrir ese gasto", explica José Luis López, coordinador del estudio. "En este grupo podrían estar incluidas familias con altos ingresos que tienen mucho gasto energético, por lo que la cifra real de riesgo podría ser más baja, aunque no demasiado. Posiblemente en próximos análisis introduciremos algún parámetro de corrección para ser más exactos, pero en esta ocasión hemos querido mantener la metodología que usamos en la primera edición para ver la evolución", aclara López.

Cuando ACA publicó la primera edición de este estudio, en 2012, casi nadie hablaba de pobreza energética en España. Solo la organización catalana Ecoserveis había hecho algunas exploraciones en años anteriores, aunque no tan detalladas. Pero como no deja de crecer, el problema ha acabado emergiendo hasta convertirse en objeto de movilización social (con colectivos muy activos como la Plataforma por un Nuevo Modelo Energético o la Alianza contra la Pobreza Energética) y colarse de manera destacada en el debate político. En el último año se han presentado en el Congreso una moción (Izquierda Plural), dos proposiciones no de ley (Izquierda Plural y Grupo Mixto) y una proposición de ley (PSOE) con medidas para mitigar su incidencia, aunque todas han sido rechazadas por el PP, que tiene mayoría parlamentaria. Y la Defensora del Pueblo, además de advertir de la gravedad del problema en su memoria anual, ha pedido al Gobierno información transparente sobre el número y las causas de los cortes de luz.



Las comunidades autónomas también han abordado la cuestión, pero solo Cataluña logró aprobar en diciembre un plan de mínimos que impide a las eléctricas cortar el suministro a hogares vulnerables hasta abril, aunque eso no significa que se les perdone el pago de los recibos, sino que les permite aplazarlo. Andalucía ha anunciado una fórmula parecida, pero aún no la ha concretado.

Mientras tanto, como medida a corto plazo, el estudio de ACA subraya que debe redefinirse el único mecanismo que existe en España para mitigar el problema: el bono social. Esta medida, que congeló el precio de la luz con el importe de julio de 2009 y a día de hoy ofrece a los beneficiarios en torno a un 25% de descuento sobre la tarifa regulada, se considera insuficiente porque solo subvenciona la calefacción eléctrica —cuando muchos utilizan sistemas de gas— y además los hogares que pueden solicitarlo no siempre son los más vulnerables. “El mero hecho de tener una potencia contratada inferior a tres kilovatios [uno de los criterios que se aplican para seleccionar a los beneficiarios] no indica que se trate de un hogar vulnerable. Viviendas que en muchas ocasiones están vacías, pero que mantienen dado de alta el suministro eléctrico con la potencia mínima se están beneficiando del bono social”, apunta el informe.

http://sociedad.elpais.com/sociedad/2014/03/27/actualidad/1395947956_321445.html

elPeriódico BARCELONA LUNES, 25 DE ENERO DEL 2016

Barcelona forma a 100 personas para combatir la pobreza energética

- **El convenio entre ayuntamiento y entidades sociales permitirá auditar 5.000 viviendas**
- **El plan está destinado a personas en riesgo de exclusión social que trabajaron en la construcción**

Una astuta unión de fuerzas permitirá asestar sendos pequeños golpes a dos de las lacras de la sociedad contemporánea: el paro y la **pobreza energética**. Un acuerdo entre el Ayuntamiento de Barcelona y las **Entidades Catalanas de Acción Social** (ECAS) facilitará un empleo a 100 personas en riesgo de **exclusión social** para que revisen el consumo energético en unas **5.000 viviendas** de los 10 distritos de la capital catalana. Para todo ello, se invertirán **2,5 millones de euros**, el 80% de los cuales saldrán del superávit municipal del 2015, y el 20% restante, de las organizaciones del tercer sector que se impliquen en el proyecto.

Los elegidos, mayoritariamente personas que se dedicaron a la **construcción** y que un buen día de la crisis se vieron en la calle, se someterán a una formación que está previsto que empiece en febrero. Constará de 128 horas de aprendizaje que se alargarán durante tres semanas. A partir de ese momento, el centenar de elegidos para esta tarea se repartirán durante cinco meses por el territorio para auditar un total de 5.000 viviendas, en las que está previsto que diagnostiquen la situación del piso, el gasto y el uso de la energía.

"PROGRAMA TRANSFORMADOR"

La responsable de las entidades se ha felicitado por el acuerdo, que ha calificado de "**programa transformador**". Nada que ver, ha señalado, con "colaboraciones puntuales quizás no demasiado importantes". "No es un proyecto piloto, sino un nuevo modelo en la lucha contra la pobreza energética", ha señalado.

De las 20 personas contratadas para cada ámbito (cinco grupos y dos distritos por equipo), cinco trabajarán como informadores energéticos y los otros 15 se encargarán del trabajo de mono azul y herramientas. El proyecto se enmarca dentro del **programa Labora** del Ayuntamiento de Barcelona, creado en marzo del 2015 y gestionado por el **Instituto Municipal de Servicios Sociales**, que, en colaboración con las entidades sociales y el tejido empresarial de la ciudad, genera una bolsa de trabajo reservada a personas en riesgo de exclusión social.

Preguntada sobre las ayudas municipales –Barcelona puso en marcha en noviembre tres puntos de **atención a la pobreza energética**-, Ortiz ha explicado que el año pasado se destinaron 780.000 euros al pago de facturas de familias que no podían asumir los suministros del hogar, una cifra muy baja, ha admitido. La razón, que muchos hogares no quieren que se conozca su situación de precariedad y no recurren a los servicios sociales. Y también, que muchas personas mayores que viven solas no saben cómo pedir asistencia.

Aunque tampoco ayudan las **empresas eléctricas**, que no informan sobre los contribuyentes a los que se **corta la luz**, lo que facilitaría la acción municipal. Cuestionada por si las compañías cumplen y no cortan el **suministro** sin consultar si se trata de casos de **vulnerabilidad**, Ortiz ha asegurado que solo han tenido conocimiento de tres casos, dos de los que se solucionaron con rapidez y un tercero del que han tenido noticia este mismo fin de semana y en el que están trabajando. La teniente de alcalde ha avanzado que el ayuntamiento denunciará si se producen **cortes injustificados**.

http://www.elperiodico.com/es/noticias/barcelona/barcelona-forma-100-personas-para-combatir-pobreza-energetica-los-hogares-4843096

LA VANGUARDIA

Barcelona "reorienta" las ayudas y abrirá más oficinas contra la pobreza energética

- Firma un convenio un convenio con entidades sociales para "detectar" más casos
- Con un presupuesto de 2,5 millones, creará 100 puestos de trabajo para personas vulnerables

25/01/2016 10:22

El Ayuntamiento de Barcelona "reorientará" las ayudas que hasta ahora ofrecía para impagos de los subministros básicos del hogar, a fin de luchar contra la pobreza energética en especial en los meses más fríos. A partir de ahora, ha anunciado la teniente de alcalde de Derechos Sociales, Laia Ortiz, el consistorio "reorientará" parte de la partida presupuestaria para abrir más Oficinas de atención energética en los barrios con especiales dificultades. Según Ortiz, las 5.062 ayudas otorgadas en 2015 son pocas en relación a la estimación de población destinataria -un 10 % de las familias de la ciudad- y quedan lejos del presupuesto disponible para este fin. "No hemos de esperar a que nos vengan a buscar, hemos de ver quiénes son", ha señalado.

La teniente de alcalde ha anunciado este cambio en la presentación del convenio entre el consistorio y la plataforma Entidades Catalanas de Acción Social (ECAS), que persigue "tener más ojos" para detectar los casos de pobreza energética en Barcelona. La iniciativa cuenta con un presupuesto de 2,5 millones y creará 100 puestos de trabajo para personas con dificultades de acceso al mercado laboral, preferentemente parados de larga duración y mayores de 45 años.

Cinco mil pisos

El programa tiene una duración de seis meses, durante los que prevén intervenir en unas 5.000 viviendas, con el objetivo de mejorar sus condiciones para contribuir a reducir el consumo energético y las facturas, en un programa del que el Ayuntamiento asume el 80% del presupuesto y en el que entidades de Ecas aportan el 20% restante. La iniciativa, en la que también colaboran Labcoop y el Programa Làbora, se desarrollará en los distritos de Sant Andreu y Sant Martí mediante la entidad ABD -que lleva a cabo la coordinación operativa del programa-; en Nou Barris y Horta-Guinardó a través de la Fundació Salut i Comunitat; en Ciutat Vella y Sarrià-Sant Gervasi con Suara y Surt, y en Les Corts y Sants-Monjuïc con Àmbit Prevenció e Iniciatives Solidàries.

Prevé intervenir en cada uno de los cinco territorios en unos 1.000 hogares, en los que primero se hará una visita para diagnosticar la situación de la vivienda, luego se ofrecerán medidas de ahorro -con un 'kit energético' con bombillas y aislantes, entre otros elementos-, y finalmente se analizará la situación tras las mejoras. “El papel del ayuntamiento no es pagar antes las facturas a las suministradoras, sino que se reduzcan estas facturas y exigir a las eléctricas que cumplan su función social” , ha criticado Ortiz.

Inserción laboral

El programa creará empleo mediante la recalificación profesional de personas que son beneficiarias del programa Làbora y que tienen dificultades de inserción en el mercado laboral -especialmente que tengan formación y experiencia en el ámbito de la construcción- para dirigirlos hacia nuevos nichos de empleo que contribuyen a una mejora ambiental. Empleará a 20 personas en cada uno de los cinco territorios, con 15 agentes energéticos y cinco informadores en cada uno, puestos para los que se han valorado 319 currículum, de los que se han preseleccionado 154, 130 de los cuales se han mantenido en el proceso tras una acción grupal, de los que 85 presentan un perfil de agentes y 45, de informadores.

Los seleccionados serán contratados el 1 de febrero e iniciarán el proceso de formación -de 128 horas-, y serán empleados siguiendo el Conveni Català d'Acció Social a 32 horas semanales y con un sueldo de 1.070 euros brutos en la categoría de administrativo, ha detallado Ortiz preguntada por los periodistas.

Los agentes se encargarán de la diagnosis de las viviendas, de detectar situaciones de vulnerabilidad, de asesoramiento y gestión tarifaria, de detectar eventuales irregularidades en los servicios energéticos, de instalar medidas de aislamiento y de derivar casos a servicios sociales para posibles reformas y cortes de suministro, entre otras funciones. Los informadores llevarán a cabo la gestión administrativa, la planificación de intervenciones, la atención telefónica, la recepción de casos y de contacto con los usuarios y la gestión de incidencias, entre otras.

Preguntada sobre la situación de las personas tras los seis meses que dura el trabajo, Ortiz ha señalado que el gobierno de Ada Colau exige a las empresas suministradoras incluir en el protocolo en el que están trabajando que presten servicios de información y asesoramiento, funciones en las que asegura que encajarían estos perfiles.

780.000 euros en ayudas

Sobre las ayudas municipales para combatir la pobreza energética, ha señalado que el Ayuntamiento ha concedido 780.000 euros repartidos en 5.062 ayudas en este ámbito en 2015, cifra que representa un aumento del 22,71% respecto al año anterior.

Preguntada por si las compañías cumplen y no cortan el suministro sin consultar si se trata de casos de vulnerabilidad, Ortiz ha asegurado que sólo han tenido conocimiento de tres casos, dos de los cuales se solucionaron con rapidez y un tercero del que han tenido conocimiento este mismo fin de semana y en el que están trabajando, tras lo que ha avisado de que el Ayuntamiento denunciará si se producen casos que lo incumplan.

Teresa Crespo, presidenta de ECAS, ha resaltado que la iniciativa es un "programa transformador, no un proyecto pequeño o una prueba piloto, sino que quiere ser un nuevo modelo de lucha contra la pobreza energética" que busca nuevos perfiles profesionales para crear nuevos puestos de trabajo.

http://www.lavanguardia.com/local/barcelona/20160125/301656294640/barcelona-pobreza-energetica-oficinas-insercion.html

B - The "*Encuesta de Presupuestos Familiares*" Database

In the following part we are going to explain in further detail the structure of the database used throughout this research work. In particular, we are interested in listing the *categorical variables* considered and their *labels* (i.e. states). The structure of this document takes as a reference the Manual attached to the database, available on the INE website for the year 2014 (last available survey). [30]

Database Variables

Variable Original Code	Variable Name	Variable Description	Variable Labels with Identification Number
CCAA	region	Household's residence region	1 Andalucía 2 Aragón 3 Asturias, Principado de 4 Baleares, Illes 5 Canarias 6 Cantabria 7 Castilla y León 8 Castilla – La Mancha 9 Catalonia 10 Comunitat Valenciana 11 Extremadura 12 Galicia 13 Madrid, Comunidad de 14 Murcia, Región de 15 Navarra, Comunidad Foral 16 País Vasco 17 Rioja, La 18 Ceuta 19 Melilla

TAMAMU	urban	Number of inhabitant of the town where the sampled household is living	1 Town with 100,000 or more inhabitants 2 Town with 50,000-100,000 inhabitants 3 Town with 20,000-50,000 inhabitants 4 Town with 10,000-20,000 inhabitants 5 Town with less than 10,000 inhabitants
NMIEMB	household_members	Number of family members	1-20
TIPHOGAR7	household_typology	Household's typology	1 Person living alone older than 65 2 Person living alone younger than 65 3 Couple without one child 4 Couple with one child 5 Couple with two children 6 Couple with three or more children 7 One adult with children 8 Other household's typology
ESTUDREDSP	study_level	Household main member study level (based on Spanish schooling system)	1 No schooling experience or basic level (first grade) 2 Secondary Education (First Cycle) 3 Secondary Education (Second Cycle) 5 Higher Education (i.e. University)
SITUACTSP	job_situation	Household main member job situation	1 Working at least one hour 2 Temporary inactivity 3 Unemployed 4 Retired 5 Student 6 Only dedicated to housework 7 Permanent work inability 8 Other situation of inactivity

REGTEN	tenancy	Household's dwelling tenancy status	1 Owned property (without mortgages) 2 Owned property (with mortgages) 3 Renting 4 Partly Free Transfer 5 Free Transfer
TIPOEDIF	building_typology	Typology of the building where household's dwelling is located	1 Detached house 2 Semi- detached house 3 Buildings with less than ten dwellings 4 Buildings with more than ten dwellings
NHABIT	dwelling_rooms	Number of rooms present in the dwelling	1-7 Rooms number (less than eight rooms) 8 More than seven rooms
ANNOCON	building_age	Building construction period	1 Less than 25 years ago (in 2014) 2 More than 25 years ago (in 2014)
SUPERF	surface	Dwelling's available surface	35 35 m^2 or less 36-299 m^2 300 300 or more m^2
CALEF	heating	Availability of heating system	1 Yes 2 No -9 No answer
FUENCALE	heating_system	Dwelling's heating system typology	1 Electricity 2 Natural Gas 3 Liquid Gas 4 Liquid Fuels 5 Solid Fuels 6 Solar Energy
NUMERO	number	Household's database identification number	1 - 22146
IMPEXAC	income	Family net monthly income	0 - 99999

C - Energy Poverty Drivers

This Annex Section gathers information about the *energy efficiency* energy poverty's driver for the United Kingdom. It is useful, at this point, to say that the Government-recommended measure for assessing the energy performance of dwellings is the Standard Assessment Procedure (SAP). The latter is an indicator of energy consumption per unit of floor space and includes the costs associated with space heating, water heating, ventilation and lighting, less any cost savings from self-generated energy. The rating is adjusted to the floor are so that the rating is independent of the dwelling size. It is expressed on a scale of 1 to 100, where higher numbers denote greater thermal efficiency and lower energy costs. To put this into perspective, a semi-detached property with no insulation and no central heating system would have a SAP rating of 1. The same property with loft and cavity wall insulation, double glazing and gas central heating would have a SAP rating of 73.

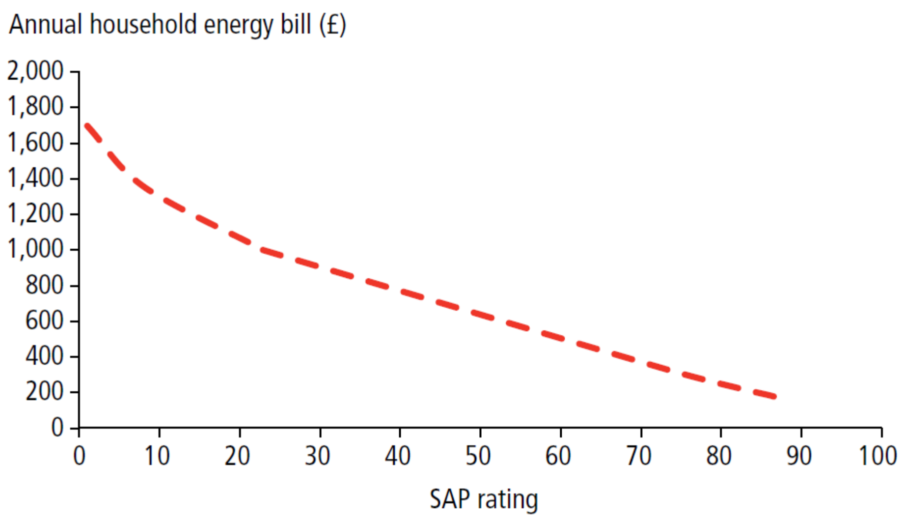


Figure 47 - Relationship between the SAP rating and annual energy bill (on 12/12/2011 1 1GBP=1.1829 EUR) for a typical semi-detached, cavity wall dwelling that is attached to the gas grid. The plot clearly shows the strong relationship between expense on energy and the energy efficiency level of the dwelling [23].

Improving energy efficiency standards implies significant upfront investment costs.

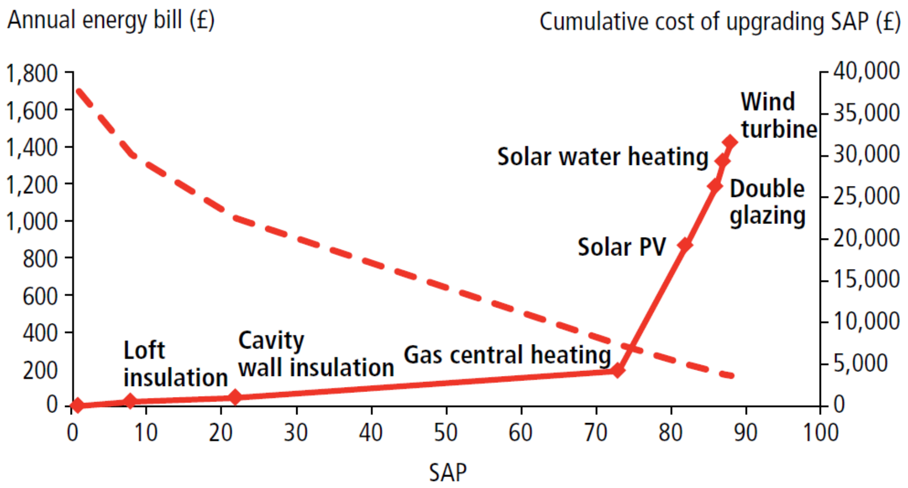


Figure 48 – Cumulative Capital investment (right vertical axis) for different improvements in energy efficiency standards (in order of cost-effectiveness). For instance a basic insulation will cost around £4,000 with very significant improvements in SAP [23].

D - Electricity and Natural Gas Markets

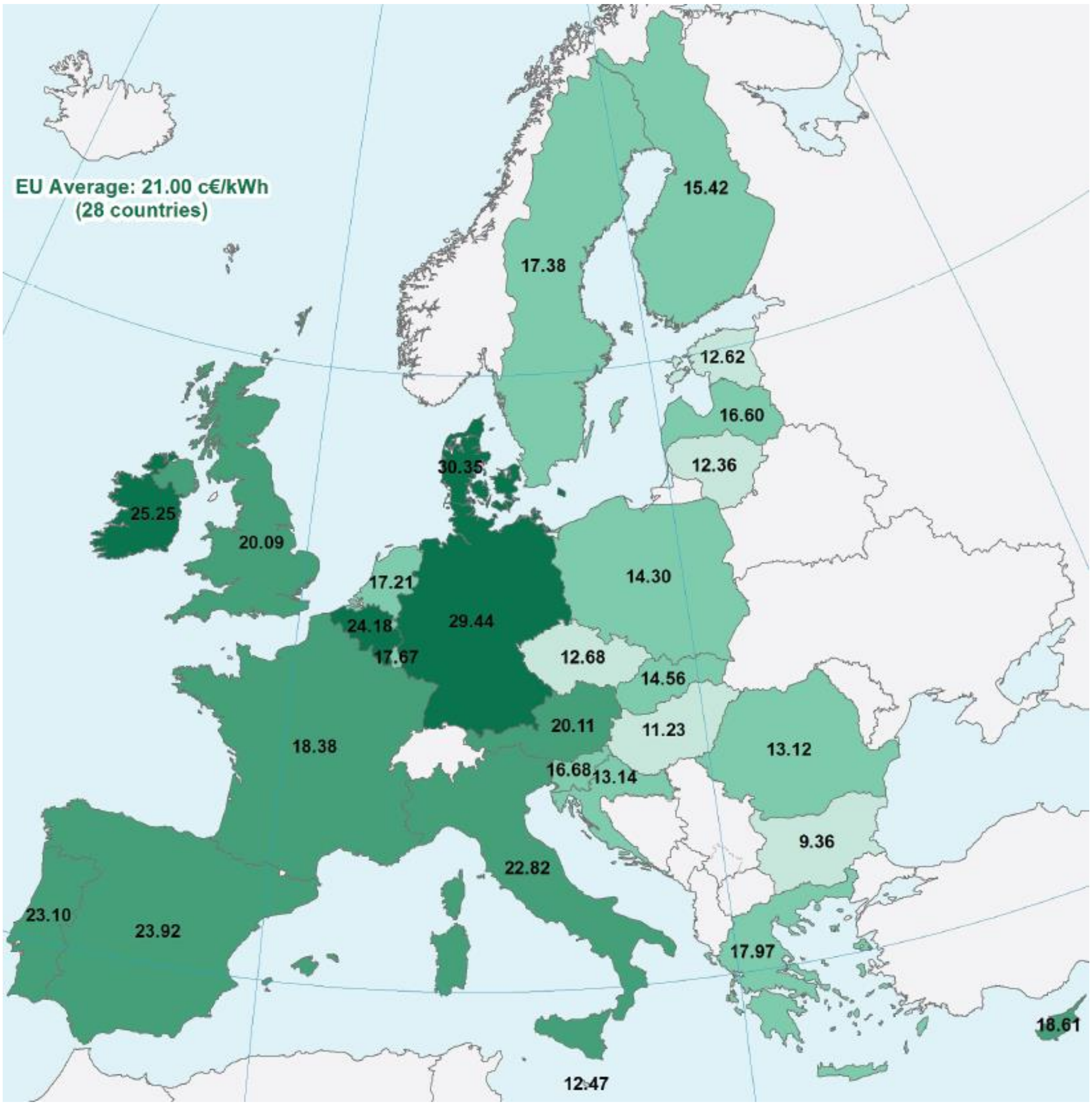


Figure 49 – Electricity retail prices (in €cent/kWh)for the domestic sector throughout European Member States [European Commission Quarterly Report on Electricity Markets, 2015]

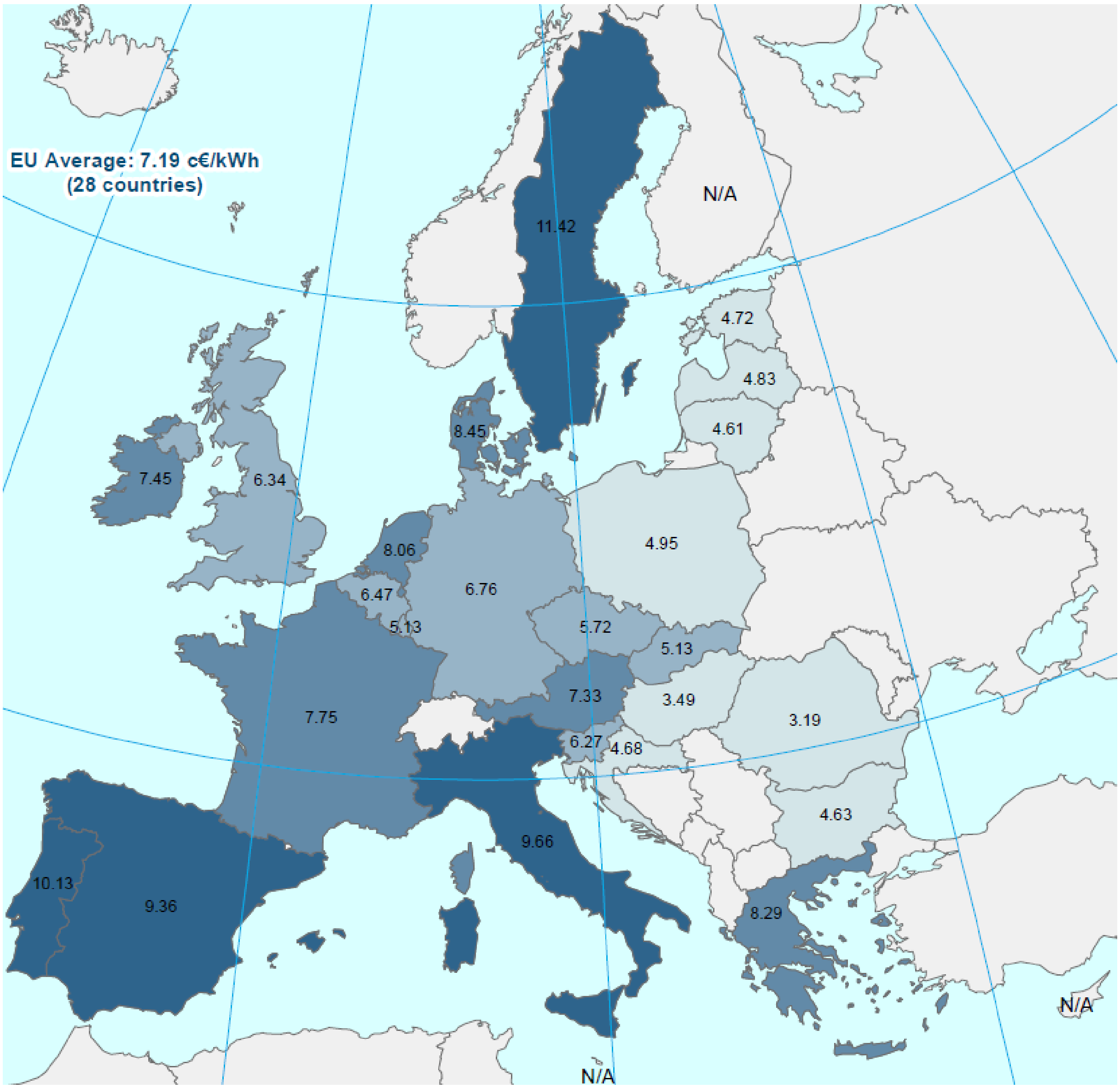


Figure 50 – Natural Gas retail prices(in €cent/kWh) for the domestic sector throughout European Member States [European Commission Quarterly Report on Natural Gas Markets, 2015]

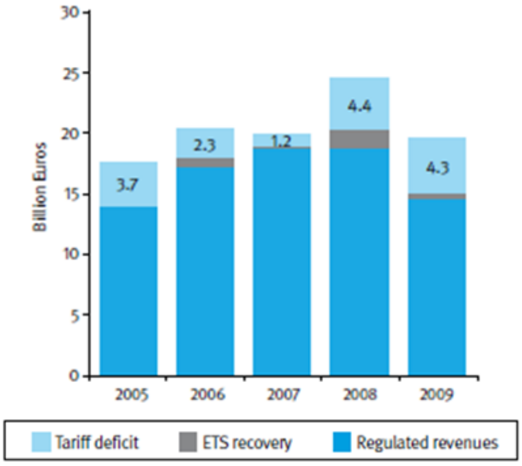


Figure 51 - Annual Utilities deficits (Electricity). It calculates the shortfall between overall regulated revenues (collected from regulated tariffs and access charges) and corresponding costs. This shortfall reached the maximum level in the 2008-2009 period, with annual amounts of roughly €4.3 - €4.4 billion in each of the two years. The high annual deficit was associated with exceptionally high wholesale electricity costs [31]

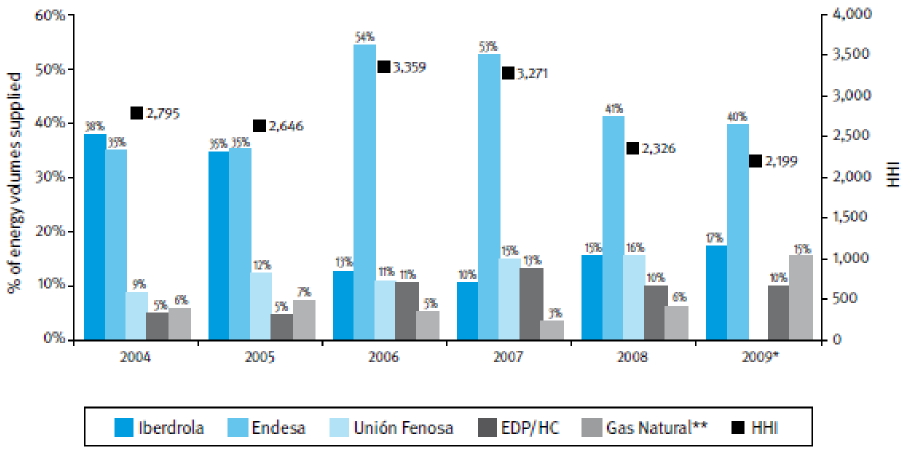


Figure 52 – Retail shares in the Liberalised Spanish Electricity market. Endesa acquired a leading position in the Electricity market in 2006)when other competitors scaled back due to their unbearable tariff deficits). Endesa's market share has gradually declined from 54% to just below 40% in 2009. In the right axis the Herfindhal-Hirschman Index is reported to evaluate the electricity market concentration [31]

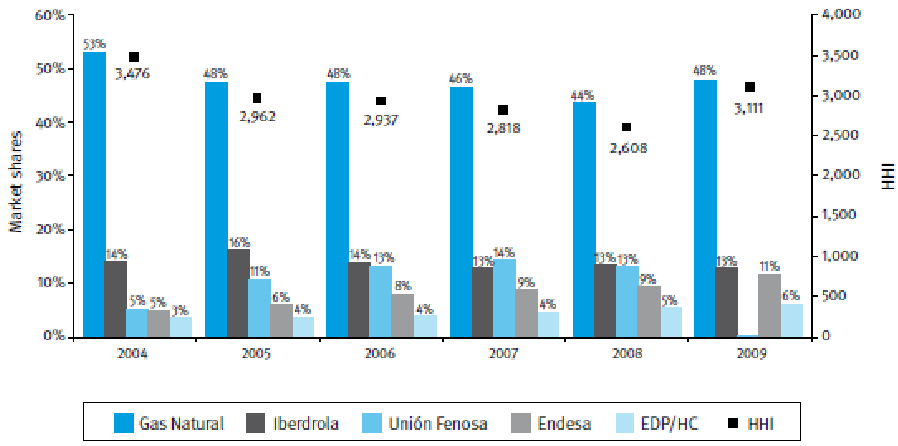


Figure 53 - Retail shares in the Liberalised Spanish Natural Gas market. Endesa acquired a leading position in the Electricity market in 2006)when other competitors scaled back due to their unbearable tariff deficits). The Spanish gas market is characterized by an high concentration degree and the presence of a strong retail player (Gas Natural). The Gas Natural market share increased in 2009 due to the acquisition of Union Fenosa Gas (UFG). In the right axis the Herfindhal-Hirschman Index is reported to evaluate the electricity market concentration [31]

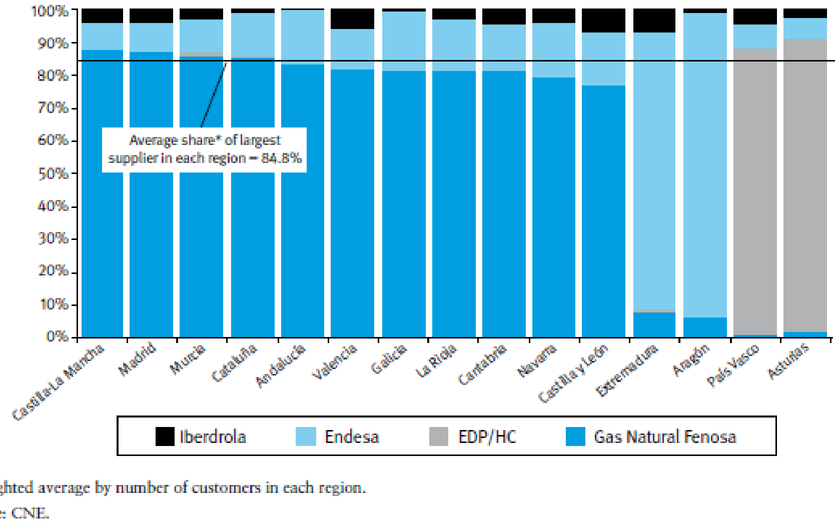


Figure 54 – Regional Shares in the Retail Gas Market by number of customers in 2009 [31]

E - European Regulatory Framework

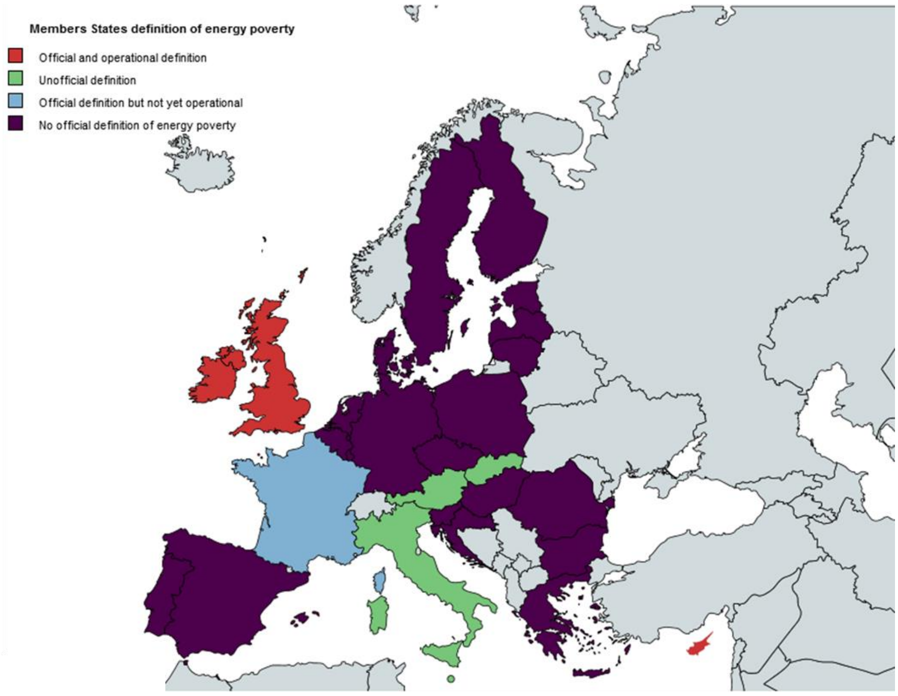


Figure 55 - Members States Definition of energy poverty [Own Elaboration based on INSIGHT-E (2015)]

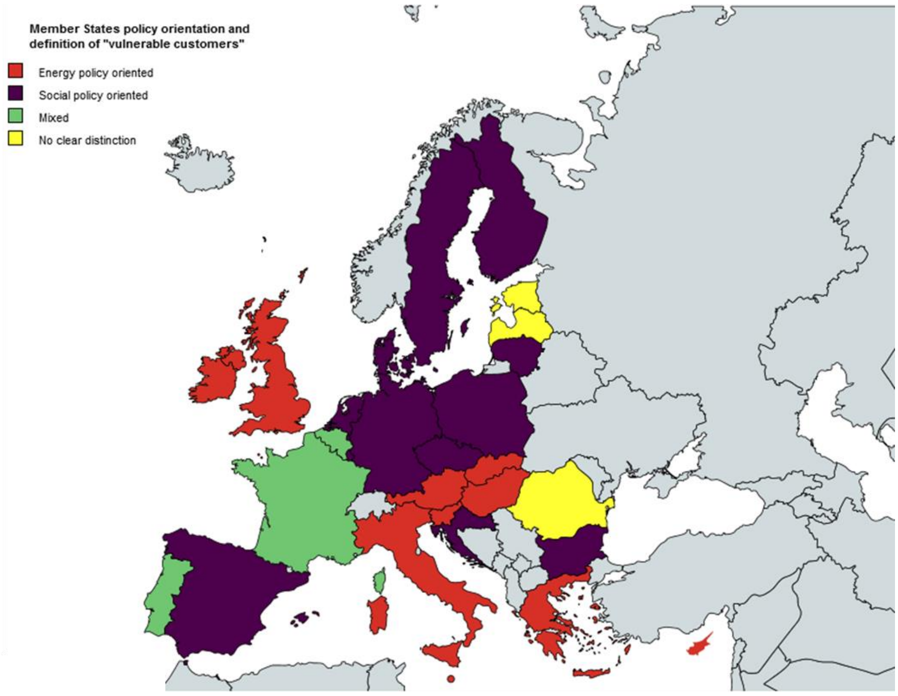


Figure 56 - Member States policy orientation and definition of "vulnerable customer" [Own Elaboration based on INSIGHT-E (2015)]

F - Energy Poverty Tables (2011 – 2013)

2011	Objective indicators [% of households]								Median expense on energy*	
	γ_1	γ_3	γ_4	π_1	π_2	δ_1	δ_2	EEI	Surface**	Absolute***
Andalucía	14,5	18,1	22,5	10	14,4	5,5	8,9	14,9	8,8	886,9
Aragón	18,2	22	25,8	9,9	13,8	3	6,4	14,6	12,1	1060
Asturias, Principado de	8,7	10,6	15,8	5,1	10,3	2,6	7,5	11,4	9,9	855
Balears, Illes	10,1	13,7	19,1	5,3	10,7	2,5	7,4	11,9	9,4	934,2
Canarias	6,5	8	16,8	3,4	12,1	8,9	15,3	15,2	6,3	595,4
Cantabria	13,7	17,5	21,9	7,2	11,6	3,3	7	14,3	10,6	977
Castilla y León	21	24,1	30	12,5	17,9	3,9	8,7	19,6	10,7	1018,2
Castilla – La Mancha	31,7	38,2	41,8	18,1	21,7	3,1	6,3	20,3	11,4	1264,8
Cataluña	**12,8**	**16,4**	**19,4**	**6,1**	**9,1**	**2,8**	**5,4**	**11,7**	**12**	**1034,7**
Comunitat Valenciana	12,6	15,7	20,1	7,4	11,8	5,3	9,3	12,2	8,3	840,2
Extremadura	20,5	23,5	27,7	12,7	17	4,5	8	16	8,2	888,1
Galicia	13,6	17,1	20,9	8,2	12,1	3,6	6,7	11,4	9,2	946
Madrid, Comunidad de	11,2	13,2	16,2	6,5	9,4	2,2	4,9	11,4	12,5	1093,9
Murcia, Región de	18	21,6	27,8	10,5	16,7	7,3	12,7	16,5	8,7	946
Navarra, Comunidad Foral de	15,3	18,3	20	7,5	9,3	1,5	3,1	10,3	11,3	1137,6
País Vasco	7,3	9,6	12,7	3,8	6,9	0,8	3,8	8,3	11,1	947,9
Rioja, La	17,7	22,1	25,2	9,7	12,8	2,8	5,4	12,5	11,4	1075,5
Ceuta	1,7	1,7	13,7	1,5	12	3,4	13,7	13,7	6,5	544
Melilla	9,2	13,8	21,5	4,6	12,3	11,5	17,7	23,9	8,4	628,3
Spain	**14,4**	**17,6**	**21,9**	**8,3**	**15,5**	**3,7**	**7,4**	**13,5**	**9,8**	**930,2**

*Figure 57 - Indicators results per Spanish region for the year 2013 [Data: INE EPF Base 2006]. *Expense on energy products as € per year and per household. ** Annual energy expense per m^2. *** Annual energy expense per household [Own Elaboration]*

2012	Objective indicators [% of households]								Median expense on energy*	
	γ_1	γ_3	γ_4	π_1	π_2	δ_1	δ_2	EEI	Surface**	Absolute***
Andalucía	17,5	18,3	24,1	10,6	16,4	8,3	12,5	16,9	9,3	921
Aragón	21,5	22,3	26,7	9,7	14,1	2,7	6,5	15,4	12,7	1173,5
Asturias, Principado de	8,7	9,5	14,2	5,7	10,4	3,3	7,3	12,4	10,7	907,3
Balears, Illes	11,5	12	19,1	5,6	12,6	3,5	9,8	12,7	9,3	970,1
Canarias	8,6	9,2	21,9	4,3	17	9,2	19,6	18,5	6,5	628,7
Cantabria	16	17,1	21,5	8	12,4	4,3	7,9	15	11,7	1066,4
Castilla y León	25,1	25,5	29	13,3	16,7	3,7	6,6	17,1	11,7	1174,2
Castilla – La Mancha	35,7	36,9	40,5	21,1	24,8	4	6,9	24	12,5	1366,7
Cataluña	**16,7**	**17,2**	**19,8**	**6,9**	**9,4**	**3,5**	**5,5**	**11,6**	**12,6**	**1118,9**
Comunitat Valenciana	12,6	13,1	17,2	6	10,2	4,7	8,2	11,1	8,9	886,3
Extremadura	20,5	21,8	27,2	11,8	17,3	6,8	11,2	16,1	8,2	951,6
Galicia	16,9	17,6	21,3	7,2	10,8	3,3	6,6	10,4	9,7	1006,8
Madrid, Comunidad de	12,7	12,9	16	5,3	8,4	2,9	5,5	11,6	13,6	1189,5
Murcia, Región de	17,5	18,1	25,2	9,5	16,6	5,8	11,7	16,2	9,2	969,5
Navarra, Comunidad Foral de	18,7	19,1	21,3	9,4	11,6	3	5,1	12,3	12,5	1229,1
País Vasco	7,3	7,9	11,4	3,5	7,1	1,5	4,9	8,7	11,7	988,6
Rioja, La	17,4	18,5	21,8	9,8	13,1	4,4	7,4	12,2	11,9	1135,8
Ceuta	8,6	8,6	21,6	2,6	15,5	4,3	16,4	19,8	8	632,9
Melilla	19,4	19,4	23,3	3,9	7,8	20,2	23,3	20,9	9,4	713,7
Spain	**16,5**	**17,1**	**21,7**	**8,5**	**13,1**	**4,5**	**8,4**	**14,1**	**10,5**	**1001,6**

*Figure 58 - Indicators results per Spanish region for the year 2012 [Data: INE EPF Base 2006]. *Expense on energy products as € per year and per household. ** Annual energy expense per m^2. *** Annual energy expense per household [Own Elaboration]*

2013	Objective indicators [% of households]								Median expense on energy*	
	γ_1	γ_3	γ_4	π_1	π_2	δ_1	δ_2	EEI	Surface**	Absolute***
Andalucía	17	17,6	23,1	10,2	15,6	10	14	16,5	8,7	863,2
Aragón	22,3	23,4	25,7	9,3	11,6	3,2	5,2	13,9	12,9	1225,3
Asturias, Principado de	11,1	11,5	15,1	6,1	9,7	3	6,1	11,1	10,6	913,4
Balears, Illes	15,1	15,2	20,5	7,2	12,4	5	9,5	13	9	971,4
Canarias	9,4	9,7	22	4,3	16,7	11,6	20,4	19,1	6	589,5
Cantabria	16,9	17,3	21,1	8,3	12,1	4,4	7,7	14,8	11,6	1070,2
Castilla y León	26,5	27,3	30,9	13,2	16,7	3,8	6,9	18,9	11,8	1175,2
Castilla – La Mancha	35,5	36,2	39,5	19,4	22,7	6,2	9,1	20,1	11,7	1232,3
Cataluña	**15,9**	**16,7**	**20,7**	**7,7**	**11,6**	**3,9**	**7,5**	**14,5**	**12,4**	**1052,7**
Comunitat Valenciana	12,9	13,5	18,3	6,6	11,4	4,8	8,9	12,5	8,6	866,2
Extremadura	21,1	21,5	27,4	11,9	17,7	7	12,6	16,2	8	865,6
Galicia	15,7	16,2	21,6	7,1	12,5	4,7	8,7	13,6	8,9	961,6
Madrid, Comunidad de	15,2	16,2	18,7	7,3	9,8	3,2	5,4	12,8	13,4	1174,8
Murcia, Región de	18	18,9	24,6	10,8	16,6	7	11,7	16,6	8,5	932,4
Navarra, Comunidad Foral de	19,4	20	22,3	9,8	12,2	3,6	5,7	12,3	12,7	1226,1
País Vasco	9	9,5	12,4	4,4	7,4	1,5	4,2	9,7	12	1015,1
Rioja, La	22	22,5	25	10,6	13,1	3,7	5,5	12,4	12,4	1203,1
Ceuta	7,6	7,6	16,8	3,4	12,6	10,1	16	15,1	7,1	601,6
Melilla	16,9	16,9	25,8	8,9	17,7	15,3	22,6	29,9	9	689,7
Spain	**17,2**	**17,8**	**22,3**	**8,8**	**13,3**	**5,3**	**8,9**	**14,6**	**10,3**	**980,5**

*Figure 59 - Indicators results per Spanish region for the year 2012 [Data: INE EPF Base 2006]. *Expense on energy products as € per year and per household. ** Annual energy expense per m^2. *** Annual energy expense per household [Own Elaboration]*

G - Energy Poverty Detection Algorithm for the Barcelona Study Case

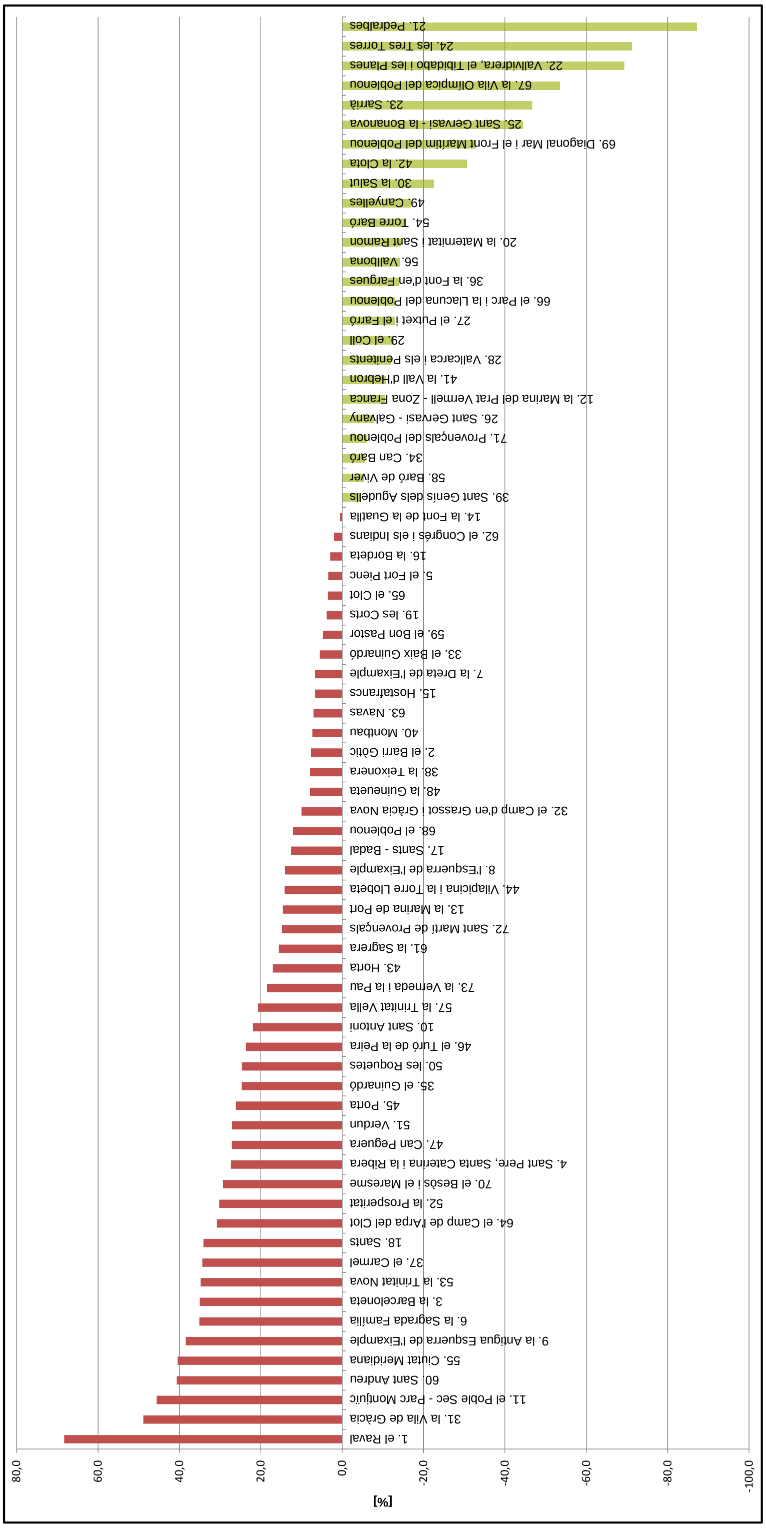


Figure 60 - Barcelona neighbourhoods at major energy vulnerability risk, according to national urban modelling. The picture shows neighbourhoods' total deviations (from city's average levels of surface, job situation, tenancy, and household typology variables) [Own Elaboration]

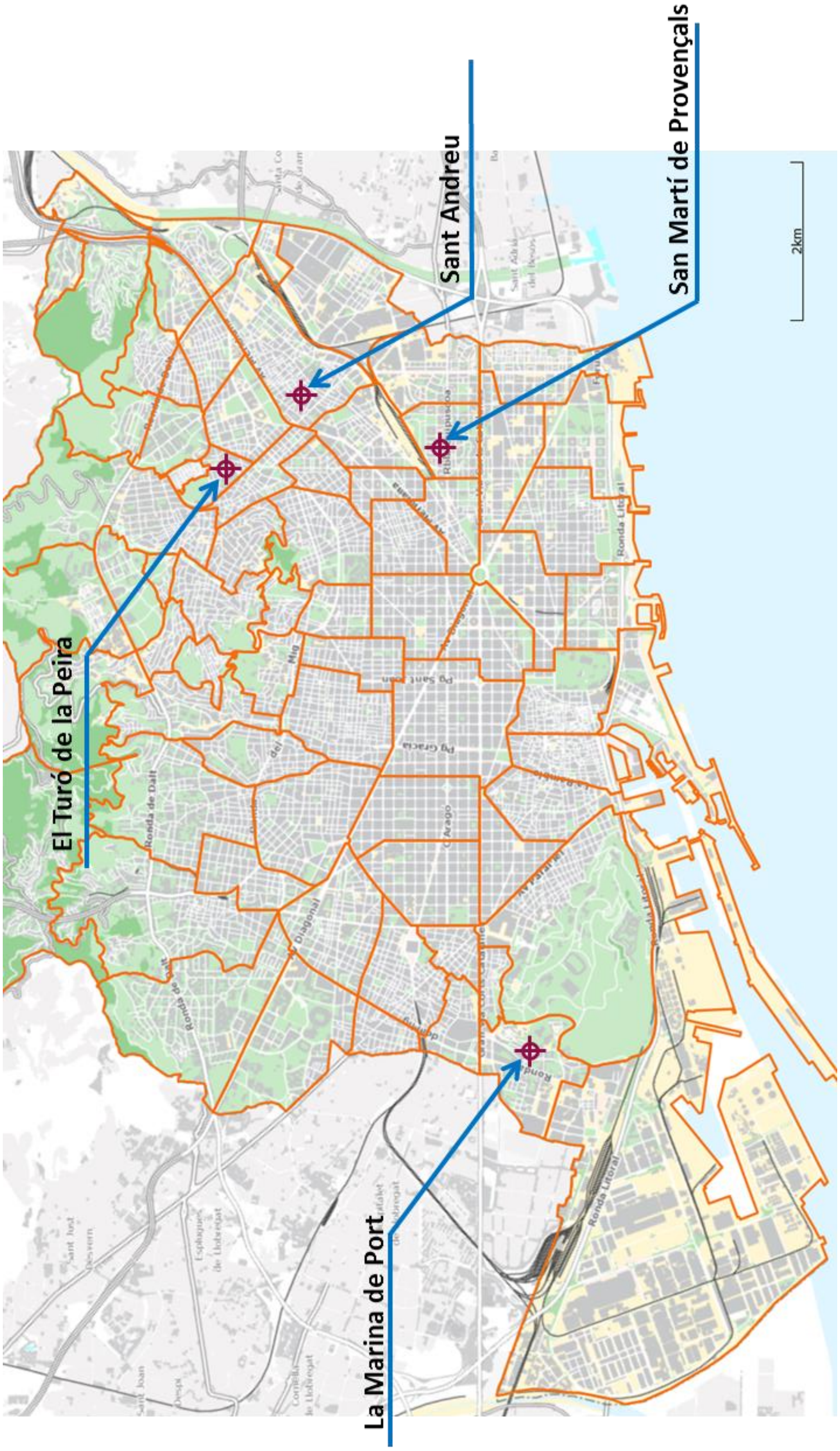


Figure 61 - Map of the city of Barcelona, highlighting the positions of the energy poverty offices opened so far by the City Council (Ajuntament de Barcelona)

http://w110.bcn.cat/portal/site/ServeisSocials/menuitem.931633495bcd6167b4f7b4f7a2ef8a0c/index1d19.html?vgnextoid=d90a1cc72cda6410VgnVCM1000001947900aRCRD&vgnextchannel=d90a1cc72cda6410VgnVCM1000001947900aRCRD&lang=es_ES

H - Energy Poverty Mortality

	Víctimas de accidentes de tráfico en carretera (1996-2014)	Mortalidad asociada a la pobreza energética (1996-2013)		
		10% de la TMAI absoluta	30% de la TMAI absoluta	40% de la TMAI absoluta
ESPAÑA	4.082	2.400	7.100	9.500
Andalucía		500	1.400	1.900
Aragón		100	200	300
Asturias		100	200	300
Baleares		100	200	200
Canarias		100	200	300
Cantabria		0	100	100
Castilla y León		100	400	500
Castilla-La Mancha		100	300	400
Cataluña		400	1.200	1.600
C. Valenciana		300	800	1.100
Extremadura		100	200	300
Galicia		200	500	600
Madrid		200	600	800
Murcia		100	200	300
Navarra		0	100	100
País Vasco		100	300	500
Rioja, La		0	0	100

Figure 62 - Winter Mortality across Spanish regions [14]

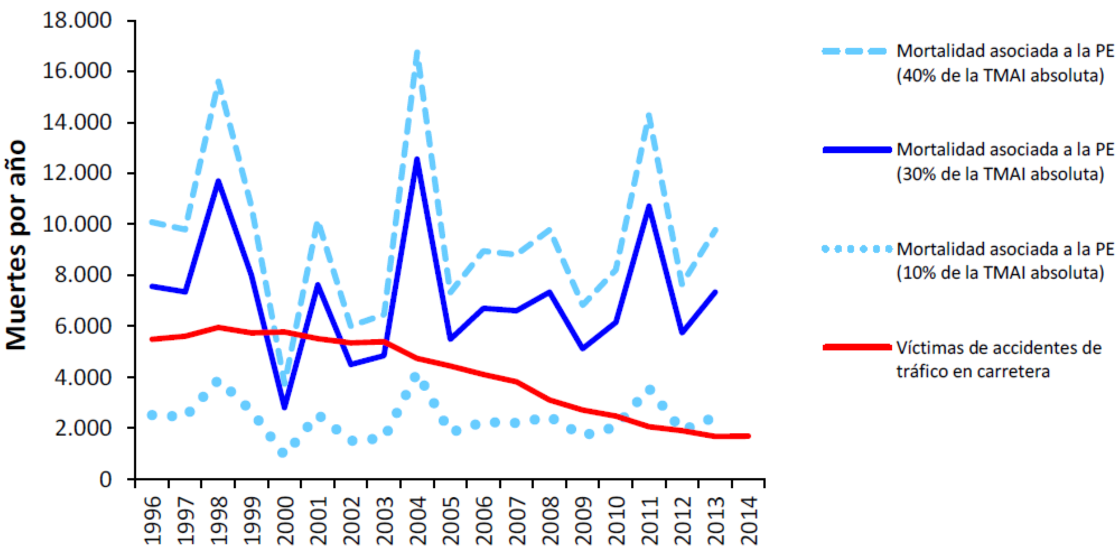


Figure 63 - Winter Mortality per year for Spain between 1996 and 2014 [14]