Università degli Studi di Padova

dSEA

Riccardo Camboni,
DSEA, University of Padova and OIPE

Alberto Corsini,
Université Côte d'Azur - CNRS, GREDEG and OIPE

Raffaele Miniaci,
DEM, University of Brescia and OIPE

Paola Valbonesi
DSEA, University of Padova and OIPE

MAPPING FUEL POVERTY RISK AT THE MUNICIPAL LEVEL: A SMALL-SCALE ANALYSIS OF ITALIAN ENERGY PERFORMANCE CERTIFICATE, CENSUS AND SURVEY DATA

June 2020

Marco Fanno Working Papers - 252

Mapping fuel poverty risk at the municipal level:

A Small-Scale Analysis of Italian Energy Performance Certificate, Census and Survey data[*]

Riccardo Camboni,[°] Alberto Corsini,[×] Raffaele Miniaci,[#] and Paola Valbonesi[⊗]

ABSTRACT: We use the *nearest neighbour propensity score matching* to link dwellings holding Energy Performance Certificates (EPCs) in the Italian province of Treviso with information on the socio-economic characteristics of households most likely to inhabit them. We construct a database of 17,405 dwellings for which information on standardized energy needs is matched to data on (potential) inhabitants and their imputed income, based respectively on census records and survey data. Our analysis shows that EPC registers can be exploited to investigate how income and housing conditions affect fuel poverty and to identify municipal areas with higher fuel poverty risk. Our findings highlight that when designing interventions to reduce fuel poverty, policymakers should target households based not only on their income but also on type of heating fuel, and on efficiency and the size of their accommodation. (135 words)

JEL CODE: C21; I32; Q48.

Keywords: Fuel Poverty; Energy Performance Certificates (EPCs); Building and dwelling Efficiency; Energy: Government Residential Policy.

---

[*] We thank you participants in the XVIII SIEPI Workshop 2020, Venice (Italy), CUOA Workshop 2018, Vicenza (Italy); 7th IAERE Annual Conference 2019, Udine (Italy); 3rd International Conference on Smart Planning for Cities and Regions 2019, EURAC Bozen (Italy); Bank of Italy seminar, Rome (Italy); IAEE Conference in Paris, 2020. We are indebted with Alberto Brunetti, Giuliano Vendrame, Francesca Zuliani (Direzione Ricerca, Innovazione ed Energia – Regione Veneto), Tania Ferrotti (Direzione ICT – Regione Veneto), Nedda Visentini (Ufficio Statistico, Regione Veneto) for data provision and their qualified cooperation in related discussion. This work has received the financial support of the *Interdeparmental Center Giorgio Levi-Cases on Energy Economics and Technology* (University of Padova); and of the French government, through the *UCAJEDI Investments in the Future* project managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-01.

[°] Department of Economics and Management – University of Padova, Italy, riccardo.camboni@unipd.it, and OIPE.

[×] Université Côte d'Azur - CNRS, GREDEG, France, alberto.corsini@univ-cotedazur.fr, and OIPE.

[#] Department of Economics and Management - University of Brescia, Italy, raffaele.miniaci@unibs.it, and OIPE.

[⊗] Department of Economics and Management – University of Padova, Italy, paola.valbonesi@unipd.it, and OIPE.

1 - Introduction

The first step in the attempt to fight fuel poverty is mapping the problem. This is not straightforward. Fuel poverty is a culturally-sensitive issue whose extent can vary over time and location and whose measurement requires a multi-dimensional approach. Recent research proposes a range of empirical approaches to the identification and measurement of energy poverty in a number of European countries. Many of these approaches are driven by data availability, and these different measurements make comparison of energy poverty levels among European countries very difficult. [1]

In the present paper, we propose a new empirical approach to map fuel poverty at the municipal level. Our strategy is mainly based on the information included in building Energy Performance Certificates (EPCs). EPCs were introduced in the European Union (EU) in 2002 as a way of achieving energy efficiency targets based on collecting and sharing information on the energy consumption of buildings. [2] The EPC for a certified dwelling provides data on energy efficiency and an estimate of standardized energy consumption (Pasichnyi et al. 2019). However, EPCs provide no information on building occupants and their characteristics, which is a piece of relevant information for policymakers trying to design tools to fight fuel poverty effectively. Our novel approach enriches the EPCs with census and survey data. Specifically, we consider about 20,000 EPCs for dwellings located in Treviso province (north-east Italy) and match them to the census data on the 280,000 inhabited dwellings of the province. Accommodations with EPC can be systematically different from those without certification. Therefore, combining these two sources of data calls for attention to linking dwellings with similar characteristics. To this aim, we employ a *nearest neighbour statistical matching procedure*, which links every EPC to a dwelling recorded in the census dataset, in the same municipality and with the same heating system, and whose characteristics make its probability of being certified similar to that of a dwelling with EPC. This statistical matching procedure resulted in

[1] For a survey of energy poverty indicators, see: Miniaci et al. (2008); Tirrado Herrero (2017).
[2] Information included in EPCs should provide incentives towards energy efficiency also via the housing markets: see Fuerst and McAllister (2011), Fuerst, et al. (2016) for investigations on the effect of EPCs rating on residential prices in UK and in Wales, respectively.

a dataset of 17,405 (geo-referenced) records of certified dwellings, which include information on their energy efficiency, standardized consumption and household occupant characteristics (e.g., age, education, occupational status, homeownership). To obtain insights into both the subpopulation living in certified houses and the entire population, we set appropriate post-stratification weights.

We use this matched dataset to investigate the probability that a household is living in an energy inefficient accommodation. Our results show that, after controlling for some dwelling characteristics, the probability of living in an inefficient accommodation does not depend on the characteristics of its occupants. We obtain similar results using a measure of standardized heating costs. Our findings show that standardized heating needs are determined largely by the dwelling's characteristics (and only marginally by the household's ones): this result is particularly relevant for considerations on the short-run, a period during which households are unlikely to be able to adjust the type and technological endowments of their dwellings. Then, we add data on incomes, imputed from the Survey on Income and Living Conditions (SILC), to study the probability of being in fuel poverty. Following Hills (2012), we define a household to be in fuel poverty if its income is below the relative poverty line, while its standardized heating costs are above the median ones. We find that both the socio-demographic characteristics, such as income, and the type of accommodation, affect the probability of being in fuel poverty.

Our results have three main implications for policymakers. First, information in EPCs, combined with data from other sources, can be used to identify the areas at higher risk of fuel poverty. This would allow to design local and area-specific interventions, and address them to the main source of the problem (e.g., poor housing conditions and insufficient income). Second, with the goal of allocating public resources more efficiently to ease fuel poverty, the eligibility of vulnerable households at risk of fuel poverty should be integrated with information on their housing conditions. Finally, our approach also shows how information included in EPCs can be exploited to provide a basis for the design of European wide policies to deal with fuel poverty.[3]

The paper contributes to two strands of literature. First, we add to the fuel poverty literature. Our new methodology, that matches individual-level data on building efficiency (EPCs) to the socio-economic and income information on residents, allows a more accurate identification of the problem. Thomson et al. (2017) suggest that there are three main approaches to the estimation of residential energy deprivation in Europe:[4] i) the expenditure approach, which is based on the household income to energy expenditure ratio; ii) the consensual approach, which develops measures based on self-reported survey data on dwellings; iii) the direct measurement approach, which records indoor thermal conditions (and electricity use) to check whether households enjoy temperatures (and electricity consumption) adequate for their well-being. Our methodology builds on the expenditure approach,[5] and proposes an empirical strategy to determine the household's necessary fuel costs and to relate them to the household's socio-economic characteristics. In particular, our fuel poverty measure mimics the Low Income High Cost (LIHC) indicator,[6] and uses the fuel consumption required to maintain a standard indoor temperature, given the observed energy efficiency of the dwellings. In so doing, our methodology refers to dwelling standard consumption data which is better than using households' actual fuel expenditure:[7] indeed, standard consumption refers to a degree of comfort which the policymaker deems a merit good, whereas actual expenditure inevitably depends on the resources available and on the household's behaviour and choices (more or less virtuous).

[3] Concerns have been raised about the quality of the information included in EPCs by Harsman et al. (2016), Jenkins et al. (2017), Las-Heras-Casas, et al. (2018) and Pasichnyi et al. (2019). EU Directive 31/2010 established a quality control regime for EPCs which was implemented in 2014 and which calls explicitly for a statistically significant random sample of the EPCs issued annually to be monitored. Although we use data from EPCs since 2015, i.e., after the quality control regime was implemented, we are aware that EPC data quality requires further validation.

[4] See also European Energy Network (2019).

[5] See Thomson et al. (2019, Table 4, p. 886) for a summary of research papers that provide expenditure based assessments of energy poverty across the EU.

[6] The LIHC indicator of fuel poverty classifies households as fuel poor if they (a) have (required) heating costs above the national median, and (b) income net of energy costs below 60% of the equivalent median (Hills, 2012).

[7] See among others Moore (2012) and Liddell et al. (2012) for discussions on the limits to using households' actual energy expenditure to measure fuel poverty.

Second, our empirical results add to literature that uses EPC data to develop policies addressing buildings' energy demand/efficiency (Pasichnysi et al., 2019). Buildings are among the largest consumers of energy and one of the most cost-effective sectors to target for reducing energy consumption. Both decreasing buildings' energy demand and improving buildings' energy efficiency lead to emissions reductions (IEA, 2010). However, the technological investments required to reach such environmental goal will be rarely implemented by vulnerable households. Thus, mapping the probability of being in fuel poverty provides a relevant tool for local policymakers to design actions addressed to the specific target of low-income households living in inefficient buildings. All in all, these actions - on the one hand - will fight energy poverty and - on the other - will provide incentives for energy-saving regeneration of buildings, urban renovation and environmental improvement measures. We are not the first in exploiting EPCs to pinpoint households living in fuel poverty. Fabbri (2015) uses information from EPCs to estimate the household income threshold below which the occupants of the building would spend more than about 7% of their income for domestic energy consumption. He estimates such thresholds for an Italian region and different types of buildings; he uses aggregate statistics to assess the risk of fuel poverty in the population by income quantile. Our approach is different in that we link each dwelling holding the EPC to the characteristics of its (most likely) occupants: in so doing we can investigate the determinants of fuel poverty at the micro-level rather than at the aggregate level.

The rest of the paper is organized as follows. Section 2 presents our data sources and Section 3 describes our methodological approach. Section 4 illustrates and discusses the results. Section 5 provides some policy implications and conclusions.

2 - EPC, census and SILC data to measure fuel poverty

We describe in detail the datasets used to implement our strategy to measure fuel poverty, which relies on matching the standardized heating costs of each observed certified dwelling to the socio-economic characteristics of their most likely inhabitants. Standardized heating costs are estimated on data contained in the EPCs, which are part of an EU-wide rating scheme aimed at obtaining and sharing information on the energy efficiency of buildings. Household socio-economic information is derived from the ten-year general population and building census and the EU SILC.

2.1 - Energy Performance Certificates

EPCs were introduced by the 2002 Energy Performance of Buildings Directive (EPBD).[8] The information contained in EPCs are consistent across EU member states, and 19 states provide open access to their local EPC registers.[9] An EPC is required for every new, renovated, sold or rented residential dwelling. Energy auditors provide the information needed to assess the building’s/dwelling’s energy efficiency, based on a 10-point rating scheme; they are also responsible for suggesting ways to reduce energy consumption (Perez-Lombard et al., 2009).

We have access to the EPCs for around 25,000 dwellings (corresponding to about 6.3% of the residential stock) in the province of Treviso, a densely populated county in the Veneto region in the north-east of Italy, which has fairly homogeneous climatic conditions. The certificates were issued between September 2015 and December 2017 and conform to the format adopted in Italy from Q4-2015. They include information on the heated surface and volume of the individual dwelling, its construction date, its geo-location and the characteristics of the main building in which it is located. In particular, each EPC provides information on the energy source mix available for heating, domestic hot water and lighting; use of renewable energy sources; insulation and the exposure of the dwelling/building. Based on these data, each certificate provides an estimate of the energy required for *normal use* during a reference year. Dwelling then are rated in ten levels, as A4-A3-A2-A1-B-C-D-E-F-G, ranked from the most to the least efficient.

[8] The EPCs in our dataset are consistent with Article 20(2) of the 2010 updated EPBD. In July 2018, the new EPBD (2018/844) came into force: it was implemented in Italy on October, 4, 2019.
[9] A list of those states with public EPC registers is available at: https://ec.europa.eu/energy/en/content/public-epc-registers.

The EPC heating data are rated based on the energy required to maintain the dwelling at a constant 20° C temperature, 24 hours a day. However, Italian regulation limits domestic heating to a maximum of 14 hours a day. To account for this limit and to provide more realistic estimates of heating costs, we correct consumption for each energy vector reported in the EPCs and multiply total consumption by a scale factor which takes values between 0.75 and 0.9, depending on the age of the building.[10]

Finally, we construct a dwelling-specific standardized measure of the heating cost, defined as the sum, for all energy vectors $v = 1, \dots, V$, of the unitary cost of the fuel p_v multiplied by the scaled consumption for heating C_{iv} in dwelling i:[11]

$$CS_i = \sum_{v=1}^{V} p_v C_{iv} \qquad (1)$$

After cleaning the data, we are left with 20,278 certified dwellings (see Table 1): 34.3% were constructed less than 20 years earlier; 17.1% were larger than 140 sqm; 82.9% had an independent central heating system; 83.5% used natural gas as their main energy source for heating; and only 5.8% had renewable energy systems. Table 2 shows the differences between certified and non-certified dwellings.

Table 1 shows that the median standardized heating cost is €708.09/year and that the median heating cost per square metre (sqm) is €8.30/year. The standardized heating costs depend on: i) the size of the dwelling - median costs range from €379.32/year (less than 60 sqm) to over €1,500/year (larger than 140 sqm); ii) the primary heating fuel (median cost €668.93/year for natural gas and €1,368/year euro otherwise); iii) date of construction (newer dwellings cost half as much to heat); and iv) EPC class. A comparison between EPC classes highlights that the median standardized heating cost is

[10] In so doing, we adopt the scale factor for housing efficiency defined in the *Energy Report for the Veneto Region* (2017, p.187).
[11] According to our EPC data, 5.2% of dwellings use electricity for both heating and cooling purposes. To exclude cooling from our standardized measure of heating costs, we set an upper bound on electricity consumption. We based this upper bound on dwellings with electric heating systems (but no cooling systems) and we computed electricity consumption per sqm. Then, conditional on the energy efficiency class and for each quartile, we derived the median value of that ratio. We set this value as the maximum electrical consumption/sqm even.

€419.65/year (€3.62/sqm per year) and €1,139.36/year (€13.01/sqm per year) for dwellings rated A-B (14.4% of the dataset) and for those rated F-G (37,6% of the dataset), respectively.

Table 1: Descriptive statistics of EPC data. EPCs for Treviso province, issued Sep-2015 – Dec-2017.

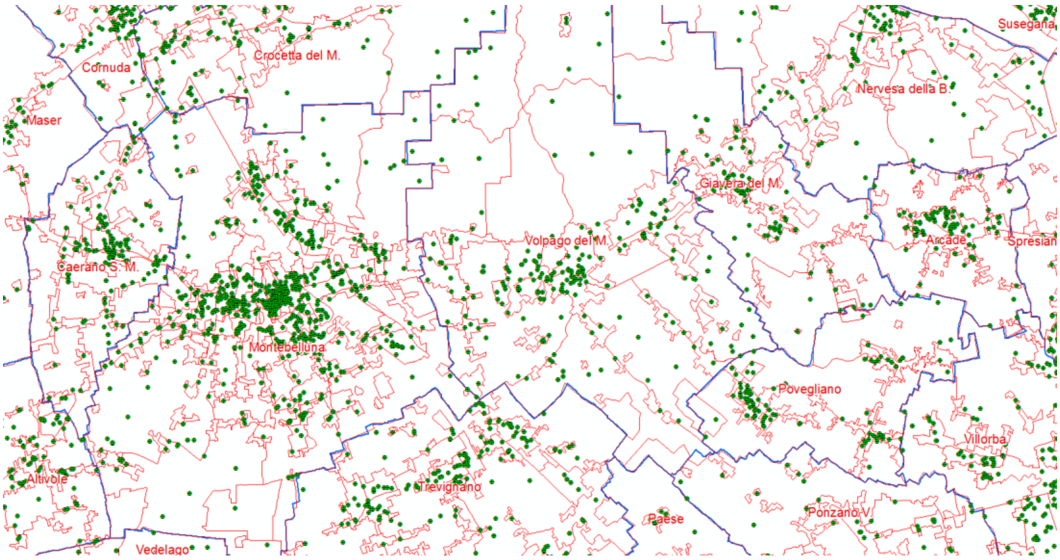
	N. of dwellings	Percentage	Median annual heating cost	
			Euro	Euro/sqm
Total	20,278	100%	708.09	8.30
Construction period				
pre-1960	2,614	12.9%	1189.31	13.00
1960-1969	2,926	14.4%	1011.47	11.27
1970-1979	2,750	13.6%	991.47	10.96
1980-1989	2,039	10.1%	824.82	9.39
1990-1999	2,995	14.8%	573.37	7.76
From 2000	6,954	34.3%	464.81	5.72
Surface				
up to 60 sqm	4,116	20.3%	379.32	7.91
60-80 sqm	4,547	22.4%	562.07	8.02
80-100 sqm	3,933	19.4%	789.75	8.89
100-120 sqm	2,638	13.0%	978.37	9.14
120-140 sqm	1,574	7.8%	1113.35	8.55
140+ sqm	3,470	17.1%	1507.03	8.13
Primary heating fuel				
Natural gas	16,932	83.5%	668.93	8.10
Other	3,346	16.5%	1368.00	12.83
Central heating				
Yes	3,469	17.1%	706.77	8.16
No	16,809	82.9%	708.25	8.34
Renewable resources				
Yes	1,169	5.8%	520.83	3.97
No	19,109	94.2%	720.58	8.53
Number of dwellings in the building:				
1	4,234	20.9%	886.59	9.50
2	3,259	16.1%	816.65	8.86
3-4	2,638	13.0%	717.07	8.32
5-8	3,154	15.6%	668.31	8.26
9 or more	4,809	23.7%	666.97	8.25
EPC class				
A-B	2,918	14.4%	419.65	3.62
C-E	9,726	48.0%	557.92	7.10
F-G	7,634	37.6%	1139.36	13.01

2.2 - Census data

We use the most recent (2011) general population and housing census data at the individual level for information on households' socio-demographic characteristics. Treviso province includes a total of

876,790 individuals, 347,833 households and 399,815 residential dwellings.[12] Each household is associated to a census tract, corresponding to a small contiguous area which, in this province, includes 75 households on average. The EPC register provides information on the geo-location of each certified dwelling, allowing us to match each dwelling in the EPC register to the corresponding census tract.[13] Figure 1 depicts the output of this exercise for a small part of the province of Treviso. Municipality boundaries are in purple and the much smaller census tracts are in red. The green points identify certified dwellings. We exploit census tract information to improve the quality of the matching between EPC and census data. This is relevant, particularly, for large municipalities where households in different neighbourhoods may have heterogeneous socioeconomic characteristics.

Figure 1: EPCs and census tracts.



Note: Municipality (purple) and census tract (red) boundaries for a small part of the province of Treviso. Exact locations of dwellings identified using EPC data are in green.

Table 2 below compares the characteristics of the housing stock surveyed by the 2011 census with the certified dwellings in the EPC register.[14]

[12] For further information on the Italian 2011 Population and Housing Census see http://dati-censimentopopolazione.istat.it/Index.aspx.
[13] Geo-localized positions typically include a 20 m. error. Addresses are not included in our data for privacy reasons.
[14] More statistics on province housing and populations are available at http://dati-censimentopopolazione.istat.it/Index.aspx.

2.2 - European Union Statistics on Income and Living Conditions (EU-SILC)

The EU-SILC is an EU harmonized survey of household income and living conditions.[15] In Italy, the sampling design allows for statistics on incomes at the (NUTS 2) regional level.[16] Publicly available microdata do not include respondents' exact locations and most of the socio-demographic and housing condition descriptors are in line with the general population and housing census. Therefore, we use the representative sample of the population of the Veneto region surveyed in the 2015 EU-SILC, to estimate a household income function and then use it to impute incomes to the households matched to the certified dwellings. The procedure is described below.

3 - Combining different data sources at the micro-level.

To get an idea of how energy needs (*Y*) covary with dwelling and household characteristics (*Z* and *X*, respectively), we need data on the joint distribution of (*Y*, *X*, *Z*). As already mentioned, we do not have direct observations of (*Y*, *X*, *Z*). Either we have data on EPCs and dwelling characteristics, but not the demographic characteristics of their inhabitants, or we have information on the characteristics of the dwellings and their inhabitants, but not on the EPC.

We define a binary variable *E* which equals 1 if the EPC is available, and zero otherwise. Our first dataset corresponds to the case where *E*=1 and (*Y*, *Z*) are observable, and the second one to the case where *E*=0 and (*X*, *Z*) are observable. We employ a *nearest neighbour propensity score matching procedure* (Caliendo and Kopeinig, 2008) to link every 'recipient' observation in our dataset where *E*=1 (i.e., EPC administrative archive) with a 'donor' observation in the dataset where *E*=0 (i.e., census records).

Specifically, we firstly pool the two datasets and estimate the probability of each observation being included in the EPC archive as a function of the variables in common, that is, the propensity score

[15] See https://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions.
[16] See http://dati.istat.it/?lang=en for Italian national and regional income and poverty statistics based on EU-SILC.

$\Pr\left(E=1|Z\right)$. Then, for each observation in both datasets, the estimated probability of being in the EPC archive is computed as $\Pr_{E=1}\left(E=1|Z\right)$ for the observations in the EPC archive, and $\Pr_{E=0}\left(E=1|Z\right)$ for the census records. Finally, each observation in the EPC archive is associated to an observation of the census records where $\Pr_{E=0}\left(E=1|Z\right)$ is as close as possible to $\Pr_{E=1}\left(E=1|Z\right)$. If more than one census record has the same $\Pr_{E=0}\left(E=1|Z\right)$, one of them is drawn at random. The selected census record 'donates' its information on inhabitants (say X_d) to the record in the EPC archive. This means that the conditional distribution $\left(Y,X\right)|Z,E=1$ is approximated by $\left(Y,X_d\right)|Z,E=1$. Note that this procedure relies on two assumptions (i.e. the conditional independence and the common support assumptions). The conditional independence assumption requires that - conditional on the characteristics of the dwelling (*Z*) - the joint distribution of energy needs and household characteristics (*Y*, *X*) is independent of whether or not the accommodation has a EPC, that is, $\left(Y,X\right)\perp E|Z$. The common support assumption requires a sufficient overlap of the characteristics of the dwellings in the EPC records with those of donors from the census records. While the conditional independence assumption is statistically untestable, the common support assumption can be assessed to document the reliability of the matching procedure.

As for data on income (*Inc*) to be used in computing the fuel poverty indicator, we need a third dataset. The EU-SILC database includes information on households (*X*), their accommodation (*Z*) and income; it is representative at the regional level and provides no other geographical information apart from the dimension of the municipality. We use EU-SILC data for the Veneto region as follows. First, we estimate a (log)linear regression model of income (ln(*Inc*)) on some variables in *X* and *Z* that are in common between the EPC archive and the census records. Then, to each observation in the EPC archive we impute the income Inc based on (X_d, *Z*) and on a random component drawn from $N\left(0,s_{\hat{u}}^2\right)$, where $s_{\hat{u}}^2$ is the sample variance of the residuals of the previous regression.

We now have $\left(Y, X_d, Z, Inc\right)|\, E=1$ as an approximation of $\left(Y, X, Z, Inc\right)|\, E=1$: this allows us to study the incidence of energy poverty for certified dwellings only. To extend the results of our analysis to the entire population, we use post-stratification weights defined by combining dwellings and household characteristics.[17]

4 - Results

4.1 - The matching procedure

The first step in the procedure described above is estimation of the propensity score $\Pr\left(E=1|\, Z\right)$. We exploit information on dwelling technological endowments and location, available from both the EPC archive and the census dataset. We work at the municipality level and distinguish dwellings whose main energy source for heating is natural gas from those that use other sources of energy. This ensures that dwellings with an EPC will be matched only to dwellings (and inhabitants) in the same municipality with the same type of heating.

After some re-coding of the original variables,[18] we define the variables included in Z as follows: a set of dummies for the construction period; the census area; presence of a centralized heating system (serving the entire building) and/or sources of renewable energy; type of domestic hot water system; and surface area of the accommodation (sqm). For dwellings whose heating systems are not based on natural gas, we include a set of dummy variables for fuel type.

In terms of our specification and estimation strategy, to obtain propensity scores, we use 190 standard logit models (95 municipalities × 2 with/without natural gas = 190), estimated via maximum likelihood. Overall, we have 20,278 EPCs and 279,964 census records containing the required information; the recipient/donor ratio for the 190 cases ranges from 1.5% to 13%, with higher ratios

[17] For an introduction to post-stratification weighting issues see Holt and Smith (1979).
[18] The two data sources sometime code similar information in different ways.

for natural gas. Having estimated the 190 propensity score functions, we impose the common support restriction, that is, we match only EPCs with an estimated propensity score $\Pr_{E=1}\left(E=1|Z\right)$ lower than the maximum propensity score for the census records $\Pr_{E=0}\left(E=1|Z\right)$. We find about 11% of the EPCs that do not satisfy this requirement, which leaves 18,094 EPCs matched to their 'nearest' dwelling (and its inhabitants) from the census records.

Table 2 provides useful statistics to assess how the characteristics *Z* of the dwellings differ between the donor and the recipient datasets (Census and EPC columns, respectively), before and after the matching procedure (unmatched and matched rows) and conditioned on use of natural gas as the main heating fuel, or not. For each variable, we show the sample means of the two datasets, the standardized percentage bias,[19] and the p-values for the t-tests for equality of means in the two samples. The statistics for the matched dwellings are computed only for donors and recipients, are weighted using the estimated propensity score. Table 2 shows that dwellings with EPC are systematically different from those registered by the census; nevertheless, the matching procedure is effective for selecting donors that are similar to recipients. For instance, the first row (Construction period "Before 1960" and "Unmatched") tells us that, according to the census data, 20.92% of all dwellings were built before 1960, whereas only 12.89% of EPCs refer to these old buildings. The difference between these two percentages is statistically relevant (the p-value for the t-statistic is less than 0.001), and the standardized percentage bias is 21.5%. Focusing on donors and recipients (matched), the means are more similar (13.54% for EPC recipients and 13.98% for census donors), their difference is not statistically appreciable (p-value 0.256) and the standardized bias is negligible (-1.2%). In some cases, the difference between donors and recipients (in terms of means) is statistically sizeable, but the matching procedure is always able to reduce the standardized bias hugely, and the (average) differences are often practically irrelevant. For instance, in the case of the surface of dwellings, the

[19] The relative bias is computed as the difference among the sample means of the EPC and census datasets as a percentage of the square root of the average of the sample variances in the two datasets (see Rosenbaum and Rubin, 1985).

average dwelling size reported in the census data is much higher than the average dwelling size shown in the EPC (114.47 sqm vs 100.02 sqm, respectively). For the matched observations, when the main heating fuel is natural gas, the means are still statistically different (the *p-value* for the test is less than 0.001), but the difference is reduced to only 2 sqm (96.76 sqm for census data vs 94.37 for EPCs).

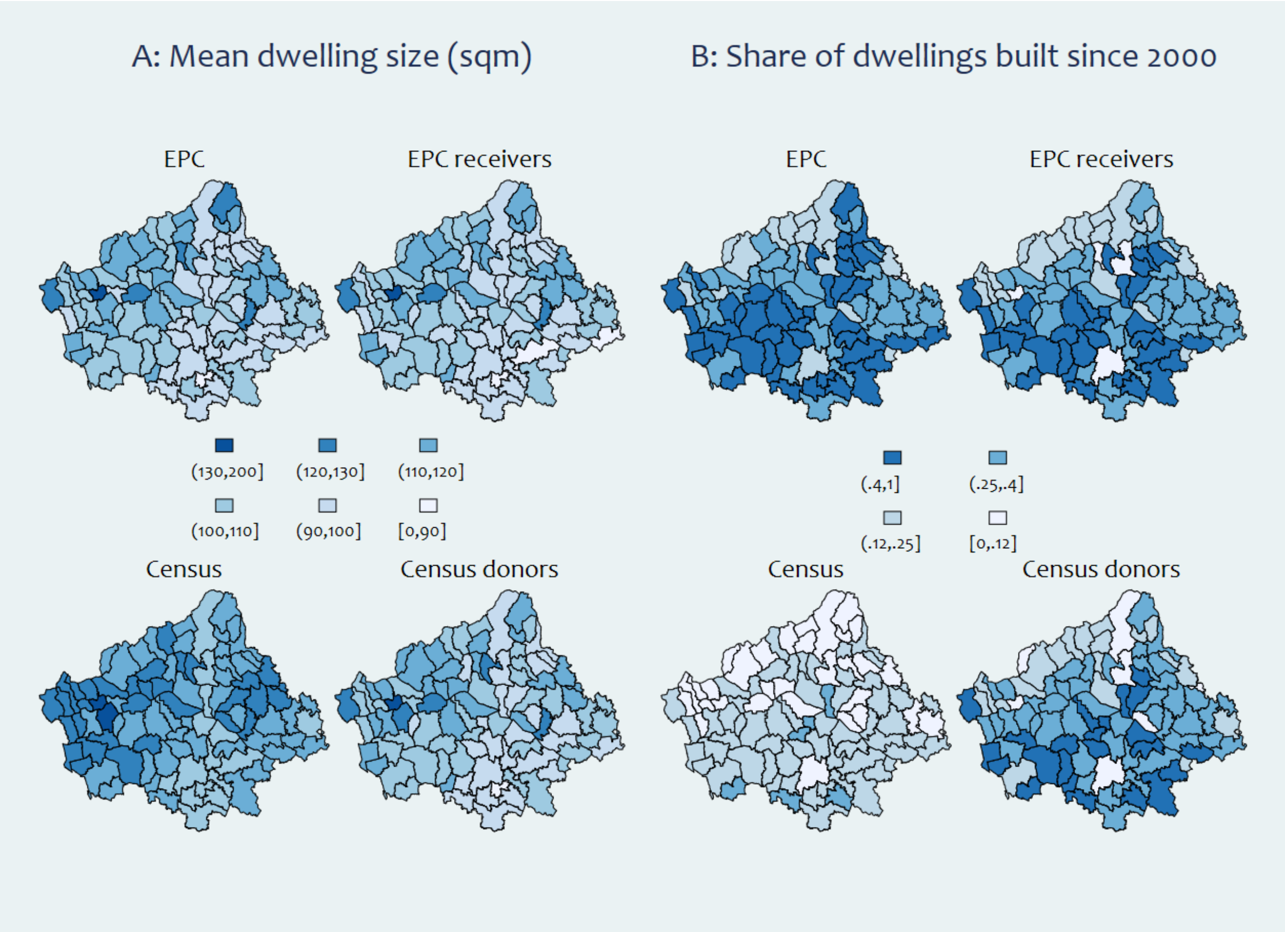
Table 2: Comparing EPC and Census datasets, pre and post matching, by main heating fuel.

		Main heating fuel: any				Main heating fuel: no natural gas				Main heating fuel: natural gas			
		EPC	Census	%bias	pvalue	EPC	Census	%bias	pvalue	EPC	Census	%bias	pvalue
Construction period (0/1 variables)													
Before 1960	Unmatch.	0.1289	0.2092	-21.5	<0.001	0.1333	0.2896	-39	<0.001	0.1280	0.1797	-14.3	<0.001
	Matched	0.1354	0.1398	-1.2	0.256	0.1645	0.1823	-4.4	0.131	0.1309	0.1334	-0.7	0.534
1960-1969	Unmatch.	0.1443	0.1728	-7.8	<0.001	0.1258	0.1993	-20	<0.001	0.1479	0.1631	-4.2	<0.001
	Matched	0.1492	0.1462	0.8	0.453	0.1509	0.1646	-3.7	0.228	0.1489	0.1434	1.5	0.2
1970-1979	Unmatch.	0.1356	0.1929	-15.5	<0.001	0.1518	0.2311	-20.3	<0.001	0.1324	0.1788	-12.8	<0.001
	Matched	0.1442	0.1479	-1	0.35	0.1914	0.1951	-0.9	0.765	0.1369	0.1409	-1.1	0.346
1980-1989	Unmatch.	0.1006	0.1330	-10.1	<0.001	0.0849	0.1256	-13.3	<0.001	0.1037	0.1356	-9.9	<0.001
	Matched	0.1062	0.1142	-2.5	0.023	0.1075	0.1197	-4	0.217	0.1060	0.1134	-2.3	0.051
1990-1999	Unmatch.	0.1477	0.1270	6	<0.001	0.0777	0.0824	-1.7	0.335	0.1615	0.1435	5	<0.001
	Matched	0.1527	0.1564	-1.1	0.366	0.0959	0.1016	-2.1	0.543	0.1614	0.1646	-0.9	0.483
From 2000	Unmatch.	0.3429	0.1652	41.7	<0.001	0.4265	0.0721	89.8	<0.001	0.3264	0.1994	29.2	<0.001
	Matched	0.3125	0.2956	4	0.001	0.2898	0.2368	13.4	<0.001	0.3160	0.3044	2.7	0.04
Surface (sqm)	Unmatch.	100.02	114.47	-28	<0.001	131.54	127.53	7.1	<0.001	93.787	109.67	-32.5	<0.001
	Matched	99.229	101.29	-4	<0.001	130.71	131.55	-1.5	0.645	94.372	96.762	-4.9	<0.001
Heating system (0/1 variables)													
Individual apparels	Unmatch.	0.0449	0.0226	12.4	<0.001	0.0768	0.0698	2.7	0.119	0.0386	0.0052	23	<0.001
	Matched	0.0336	0.0243	5.1	<0.001	0.0695	0.0659	1.3	0.655	0.0280	0.0181	6.8	<0.001
Autonomous	Unmatch.	0.8289	0.8911	-18	<0.001	0.7992	0.8264	-7	<0.001	0.8348	0.9148	-24.4	<0.001
	Matched	0.8463	0.8794	-9.6	<0.001	0.8181	0.8382	-5.2	0.088	0.8507	0.8855	-10.6	<0.001
Central	Unmatch.	0.1262	0.0864	12.9	<0.001	0.1240	0.1038	6.4	<0.001	0.1266	0.0800	15.4	<0.001
	Matched	0.1202	0.0963	7.8	<0.001	0.1124	0.0958	5.2	0.083	0.1213	0.0964	8.2	<0.001
Main heating fuel (0/1 variables)													
Other	Unmatch.	0.0630	0.0112	27.6	<0.001	0.3817	0.0418	91.5	<0.001				
	Matched	0.0318	0.0229	4.8	<0.001	0.2381	0.1757	16.8	<0.001				
Natural gas	Unmatch.	0.8350	0.7314	25.3	<0.001								
	Matched	0.8663	0.8698	-0.9	0.364								
LPG	Unmatch.	0.0467	0.0708	-10.3	<0.001	0.2830	0.2636	4.4	0.012				
	Matched	0.0452	0.0466	-0.6	0.549	0.3382	0.3581	-4.5	0.18				
Solid (wood)	Unmatch.	0.0106	0.0613	-27.5	<0.001	0.0643	0.2284	-47.7	<0.001				
	Matched	0.0101	0.0098	0.1	0.825	0.0752	0.0753	<0.1	0.99				
Heating oil	Unmatch.	0.0447	0.1252	-29.2	<0.001	0.2711	0.4663	-41.3	<0.001				
	Matched	0.0466	0.0509	-1.6	0.076	0.3485	0.3909	-9	0.005				
Main source for hot sanitary water (0/1 variables)													
Other	Unmatch.	0.0615	0.2058	-43.4	<0.001	0.3249	0.6563	-70.3	<0.001	0.0095	0.0404	-19.9	<0.001
	Matched	0.0629	0.0714	-2.6	0.003	0.4055	0.4772	-15.2	<0.001	0.0100	0.0106	-0.4	0.612
Natural gas	Unmatch.	0.8059	0.7070	23.2	<0.001	0.0983	0.0443	21.1	<0.001	0.9457	0.9504	-2.1	0.007
	Matched	0.8325	0.8403	-1.8	0.065	0.0785	0.0286	19.5	<0.001	0.9489	0.9618	-5.8	<0.001
Electricity	Unmatch.	0.0862	0.0256	26.6	<0.001	0.2971	0.0701	61.3	<0.001	0.0445	0.0092	22	<0.001
	Matched	0.0605	0.0432	7.6	<0.001	0.1877	0.1477	10.8	0.001	0.0408	0.0276	8.2	<0.001
LPG	Unmatch.	0.0464	0.0616	-6.7	<0.001	0.2797	0.2293	11.6	<0.001	0.0003	0	2.4	<0.001
	Matched	0.0441	0.0451	-0.5	0.662	0.3282	0.3465	-4.2	0.214	0.0003	0	2.1	0.081
Renewable sources (0/1)	Unmatch.	0.0577	0.0788	-8.4	<0.001	0.2534	0.1163	35.9	<0.001	0.0190	0.0650	-23.1	<0.001
	Matched	0.0408	0.0372	1.4	0.105	0.1757	0.1533	5.9	0.054	0.0200	0.0199	0.1	0.944

Note: For each variable, we report the means for the two data sources and assess their difference in terms of standardized percentage bias (% bias) and p-value of the zero equality t-tests. We compare the two raw datasets (unmatched), and only donors and recipients identified by the statistical matching procedure (matched).

Figure 2 presents some additional differences between census and EPC raw data, on the one hand, and between census and matched data, on the other. Panel A shows the mean dwelling size and Panel B the share of accommodation built since 2000 for each municipality in the Treviso province. Each panel presents the statistics for the entire EPCs archive (labelled *EPC* in the figure), the census records (*Census*), the matched EPCs (*EPC recipients*), and the set of potential donors from the census records (*Census donors*). In this last case, the statistics are weighted using the estimated propensity score.

Figure 2: Comparing EPC and Census datasets, pre and post matching, by the municipality.



*Note: The two panels show the mean dwelling size (in sqm) and the share of accommodation built since 2000 for each municipality in the Treviso province, for the entire EPCs archive (*EPC*) and the census dataset (*Census*), and for the matched EPCs (*EPC recipients*) and the set of potential donors from the census records (*Census donors*). In this last case, the averages are weighted using the estimated propensity score.*

Figure 2 confirms that the dwellings with an EPC are smaller and newer than the other dwellings included in the database. It shows, also, that: i) there is substantial spatial heterogeneity among housing stock; and ii) the matching procedure produces a spatial distribution of the donor characteristics (*Census donors* maps) that is reasonably close to the one of the matched EPCs.

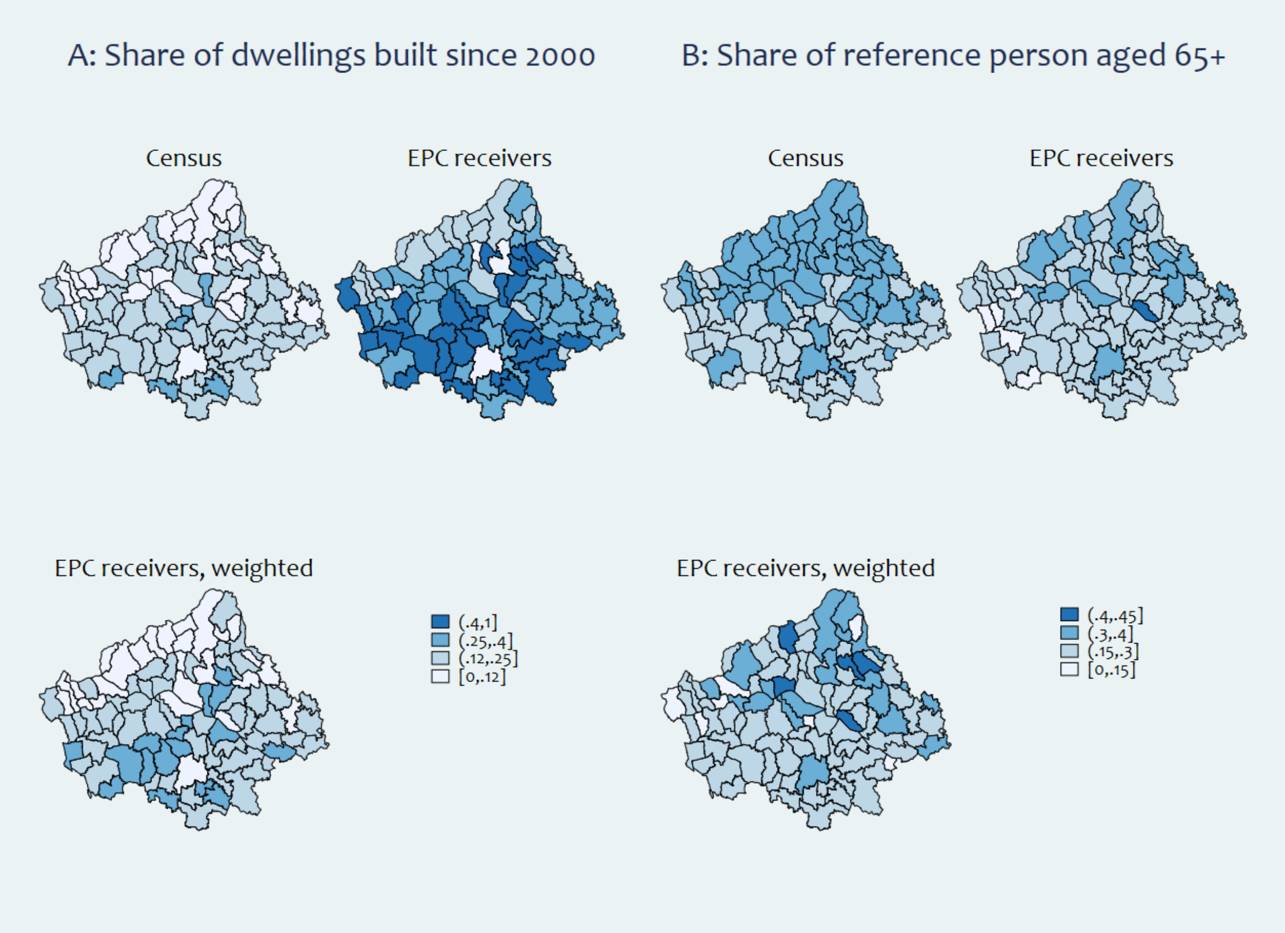
The evidences in Table 2 and Figure 2 lead us to conclude that, at the provincial level, the matching procedure provides an appropriate balancing of donor characteristics (from the census records) with certified dwellings of the EPC archive. We would not claim that the balancing is as good at the municipal level; in particular, it is difficult to match certified dwellings in small municipalities, especially if they do not use natural gas as their main heating fuel.

Since the maps of the entire population (*Census* in Figure 2) are remarkably different from the map of the EPC recipients, it is clear that inference analysis for the entire provincial population on the matched EPC data will require post-stratification weights. We compute the weights defining strata, based on dwelling size, date of construction, degree of urbanization (below/above 500 inhabitants/km^2), main heating fuel (natural gas or not), and family type (singles, couples, couples with children, single parents, other). Post-stratification weights should reduce the 'distance' between the enriched EPC dataset and the reference population: Figure 3 shows that they are effective for some of the relevant drivers (e.g., the construction period, Panel A), but less so for others (e.g., age of reference person, Panel B). In the following regression analyses, we use the post-stratification weights to obtain results reliable at the provincial level.

4.2 - Housing and socio-demographic characteristics, energy efficiency and standard costs of heating

We can now study the relationship between dwellings and their residents' characteristics, energy efficiency and standard costs of heating, using the enriched dataset. The multiple regression exercises investigate which (if any) of the dwelling and household features recorded by censuses and surveys are associated with the standardized heating costs obtained from the EPCs. Were the standardized heating cost predictable on the basis of the information gathered by a general household survey, such as the EU-SILC, we could use EU-SILC data to compute a fuel poverty index based on the standardized energy needs of the household and its ability to pay. In the absence of information on dwellings' efficiency, it results difficult to predict standardized heating costs, as highlighted by our findings presented in what follows.

Figure 3: Comparing matched EPC and Census datasets, with and without post-stratification weighting, by the municipality



Panel A shows the share of accommodation built since 2000, panel B shows the share of reference persons aged 65 or more, for each municipality of the Treviso province, for the entire census dataset (Census), the matched EPCs (EPC recipients) and the weighted matched EPCs (EPC recipients, weighted).

Table 3 presents the weighted maximum likelihood estimates of our order logit model for the probability that a dwelling is categorized as A-B, C-E or F-G energy efficient, as a function of the dwelling's and the household's characteristics assigned by the matching procedure. To compute the standard errors, we consider the presence of the covariates generated (namely, household characteristics, X_i), whose values depend on estimation of the propensity score function and a random drawn (from the set of the census records with an estimated propensity score similar to that of the receiving EPC). Therefore, we compute the standard errors by bootstrapping the matching procedure and, consequently, computing the weights and the weighted maximum likelihood estimates, one hundred times. We apply the same procedure to all the estimates presented in the paper.

Table 3: Maximum likelihood estimates of an order logit model for the probability of living in a dwelling of energy class A-B, C-E, or F-G.

	Coef.	Std. Err.	Z	p-value	Average marginal effects on the probability to fall in class: A-B	C-E	F-G
Construction period (Ref: pre-1960)							
1960-1969	-0.0016	0.0940	-0.02	0.986	0.00004	0.0003	-0.0003
1970-1979	-0.1833	0.0944	-1.94	0.052	0.0050	0.0334	-0.0384
1980-1989	-0.8566	0.1015	-8.44	<0.001	0.0309	0.1618	-0.1927
1990-1999	-1.7895	0.0821	-21.8	<0.001	0.0958	0.3016	-0.3974
From 2000	-3.1211	0.0865	-36.1	<0.001	0.2876	0.2885	-0.5761
Surface (up to 60 sqm)							
60-80 sqm	-0.1021	0.0814	-1.25	0.21	0.0056	0.0131	-0.0186
80-100 sqm	-0.2201	0.0809	-2.72	0.006	0.0125	0.0281	-0.0405
100-120 sqm	-0.3810	0.0873	-4.37	<0.001	0.0227	0.0482	-0.0709
120-140 sqm	-0.4429	0.1010	-4.38	<0.001	0.0269	0.0559	-0.0828
140+ sqm	-0.7343	0.0924	-7.95	<0.001	0.0486	0.0903	-0.1389
Primary heating fuel: natural gas	-0.2589	0.0744	-3.48	0.001	0.0160	0.0331	-0.0491
No central heating	0.1393	0.0751	1.86	0.063	-0.0093	-0.0172	0.0265
Renewable resources	-2.8160	0.2161	-13.03	<0.001	0.3403	0.0907	-0.4310
Number of dwellings in the building (Ref: One)							
2	-0.0561	0.0722	-0.78	0.437	0.0037	0.0070	-0.0107
3-4	-0.0534	0.0892	-0.6	0.549	0.0035	0.0067	-0.0102
5-8	0.1036	0.0825	1.26	0.209	-0.0065	-0.0131	0.0195
9+	0.0113	0.0708	0.16	0.874	-0.0007	-0.0014	0.0021
More than 500 inhab/km^2	-0.0006	0.0530	-0.01	0.991	0.00004	0.00007	-0.0001
Plain area	-0.0054	0.0539	-0.1	0.92	0.0004	0.0007	-0.0010
Owner occupied	-0.0100	0.0595	-0.17	0.867	0.0006	0.0013	-0.0019
Family type (Ref: single)							
Couple	0.0023	0.0885	0.03	0.979	-0.0001	-0.0003	0.0004
Couple with children	-0.0164	0.1278	-0.13	0.898	0.0011	0.0021	-0.0031
Single parent	-0.0382	0.1227	-0.31	0.756	0.0025	0.0048	-0.0072
Other	-0.1119	0.2004	-0.56	0.577	0.0074	0.0139	-0.0213
At least high school	0.1061	0.0558	1.9	0.057	-0.0068	-0.0133	0.0201
Age class (Ref: at most 40)							
41-65	0.1572	0.0563	2.79	0.005	-0.0104	-0.0197	0.0301
66+	0.1972	0.1049	1.88	0.06	-0.0129	-0.0248	0.0376
Female	0.0119	0.0642	0.18	0.853	-0.0008	-0.0015	0.0023
Household size	0.0097	0.0430	0.22	0.822	-0.0006	-0.0012	0.0018
Immigrants	0.0071	0.0794	0.09	0.929	-0.0005	-0.0009	0.0013
Occupational status (ref: employed)							
Retired from work	0.0758	0.0822	0.92	0.357	-0.0048	-0.0096	0.0144
Other not employed	0.1731	0.0891	1.94	0.052	-0.0107	-0.0219	0.0326
Cutoff 1 (A-B vs C-E)	-4.2498	0.1627	-26.12	<0.001			
Cutoff 2 (C-E vs F-G)	-1.0450	0.1547	-6.75	<0.001			

Note: Estimates use post-stratification weights; standard errors obtained by bootstrapping the entire matching procedure (and the computation of the post-stratification weights) 100 times. A variable with a positive estimated coefficient increases the probability of the accommodation of being categorized as less energy efficient. Age, education, gender and occupational status refer to the household's reference person. The last three columns show the weighted average marginal effects on the predicted probability to be categorized in the three classes. The actual (average predicted) probability of categorization as A-B is 9.29% (9.26%), as C-E is 42.39% (41.9%) and as F-G is 48.32% (48.84%). Number of observations: 18,094.

Table 3 shows that the older the dwelling, the more likely it is energy inefficient. In contrast, dwellings that rely on natural gas as the primary energy source for heating, dwellings with plants for energy production from renewable sources and large houses are rarely categorized as F or G.

After controlling for dwelling characteristics, we find that very few household characteristics are associated to energy efficiency: *ceteris paribus*, families whose reference person is aged over 40 years old, families whose reference persons is better educated, and households headed by someone who is not employed or has retired from work, are more likely to live in less efficient accommodations. At this stage, we have no information on income; thus, given that income and education are positively correlated, the negative association between education and dwelling energy efficiency may be induced by the ambiguous relation between household income and building efficiency. On the one hand, wealthier families can pay for more efficient accommodation, but on the other hand, they also can afford higher heating costs. Which of the two effects prevails is an empirical matter. Moreover, our results do not support the view that, *ceteris paribus*, owner-occupied dwellings are of better quality, or that the number of units in the building matters, or that immigrants live in worse housing conditions (at least in terms of energy efficiency). As for predictive performance, we can classify dwellings according to the energy class with the highest predicted probability. That is, dwelling *i* is predicted to fall into classes C-E if:

$$\Pr\left(Class_i \in \{C,...,E\} \mid Z_i, X_i\right) > \Pr\left(Class_i \in \{A,B\} \mid Z_i, X_i\right), \Pr\left(Class_i \in \{F,G\} \mid Z_i, X_i\right).$$

The percentage of correctly classified dwellings is 68%, with a tendency to misclassify dwellings in energy classes A-B and C-E (about 80% and 30% of misclassified cases, respectively). This shows the difficulty involved in reliably predicting the energy efficiency of homes occupied by families, based on SILC and census information.

The energy efficiency class of the dwellings is important for predicting the standardized heating costs CS_i, defined in equation (1). Table 4 shows the ordinary least squares estimates of a log-linear model. We regress $\ln\left(CS_i / Surface_i\right)$ on accommodation (Z_i) and household (X_i) characteristics. We estimate the model including and excluding the energy class of the dwelling in Z_i. Information on CS_i is available for 17,405 out of 18,094 matched EPCs. We drop from the estimation sample the observations with standard heating costs/sqm below (above) the first (99th) percentile of its distribution, conditional on energy class, dwelling size and natural gas as the primary heating fuel. This reduces the number of usable observations to 17,023.

The estimated parameters associated to classes C-E and F-G show that, *ceteris paribus*, the standardized heating costs/sqm of accommodations in classes C to E is 79.4%, higher than for similar dwellings in classes A and B, and the difference rises to 131.8% for accommodations classed as F or G. The energy efficiency classification is a crucial, but not unique, determinant of the standardized heating costs: construction period, size and technological endowment of the dwelling are all important. Standard heating costs are higher for owner-occupied homes (+4.25%/sqm), and lower if the reference person has retired (-6.28%) or has at least a high school degree (-3.96%).

The model explains 52.4% of the overall variance when we consider energy class. Notice that energy classes are not available in census or survey data. Consequently, we could not use these estimated parameters to impute a measure of standard heating costs in a survey. If we exclude the energy classes from the conditioning set (i.e., ‘Without energy classes’ in Table 4), the percentage of explained variance drops to 34.9%. Unsurprisingly, omitting the energy class variables from the conditioning set affects the estimates of the parameters of the construction period dummies, exacerbating the gradient related to the age of the building. The other parameters change very little.

Table 4: Ordinary least squares estimates of a log-linear model for the standard heating costs per square metre.

	With energy classes				Without energy classes			
	Coef.	Std.err.	z	p-value	Coef.	Std.err.	Z	p-value
EPC class (ref: A-B)								
C-E	0.7940	0.0358	22.18	<0.001				
F-G	1.3182	0.0361	36.47	<0.001				
Construction period (Ref: pre-1960)								
1960-1969	-0.1221	0.0269	-4.54	<0.001	-0.1254	0.0314	-3.99	<0.001
1970-1979	-0.1823	0.0267	-6.83	<0.001	-0.2022	0.0324	-6.24	<0.001
1980-1989	-0.2156	0.0271	-7.97	<0.001	-0.3194	0.0319	-10	<0.001
1990-1999	-0.1919	0.0294	-6.53	<0.001	-0.4306	0.0297	-14.48	<0.001
From 2000	-0.2116	0.0292	-7.26	<0.001	-0.7243	0.0266	-27.19	<0.001
Surface (up to 60 sqm)								
60-80 sqm	0.0073	0.0280	0.26	0.795	-0.0261	0.0310	-0.84	0.399
80-100 sqm	0.0493	0.0280	1.76	0.078	-0.0047	0.0298	-0.16	0.873
100-120 sqm	0.0958	0.0321	2.99	0.003	0.0149	0.0345	0.43	0.666
120-140 sqm	0.1282	0.0333	3.85	<0.001	0.0336	0.0366	0.92	0.358
140+ sqm	0.0833	0.0331	2.52	0.012	-0.0696	0.0354	-1.97	0.049
Primary fuel: natural gas	-0.7612	0.0319	-23.85	<0.001	-0.7901	0.0344	-22.95	<0.001
No central heating	0.1279	0.0240	5.34	<0.001	0.1482	0.0260	5.71	<0.001
Renewable resources	-0.3189	0.0650	-4.9	<0.001	-0.8113	0.0733	-11.08	<0.001
Number of dwellings in the building (Ref: One)								
2	-0.0066	0.0272	-0.24	0.809	-0.0207	0.0295	-0.7	0.483
3-4	-0.0320	0.0271	-1.18	0.239	-0.0491	0.0306	-1.61	0.108
5-8	-0.0011	0.0228	-0.05	0.961	0.0116	0.0268	0.43	0.665
9+	-0.0316	0.0221	-1.43	0.152	-0.0292	0.0229	-1.28	0.201
More than 500 inhab/km^2	-0.0215	0.0200	-1.07	0.284	-0.0191	0.0226	-0.85	0.397
Plain area	-0.0948	0.0177	-5.36	0	-0.0919	0.0190	-4.84	0
Owner occupied	0.0425	0.0192	2.21	0.027	0.0407	0.0206	1.98	0.048
Family type (Ref: single)								
Couple	0.0163	0.0249	0.65	0.513	0.0146	0.0294	0.5	0.62
Couple with children	0.0001	0.0393	<0.01	0.998	-0.0019	0.0434	-0.04	0.965
Single parent	0.0145	0.0344	0.42	0.674	0.0110	0.0418	0.26	0.793
Other	0.0886	0.0586	1.51	0.131	0.0771	0.0660	1.17	0.243
At least high school	-0.0396	0.0159	-2.49	0.013	-0.0244	0.0169	-1.44	0.15
Age class (Ref: at most 40)								
41-65	0.0122	0.0199	0.61	0.54	0.0298	0.0213	1.4	0.162
66+	0.0318	0.0279	1.14	0.254	0.0599	0.0347	1.73	0.084
Female	0.0017	0.0193	0.09	0.931	0.0010	0.0217	0.05	0.964
Household size	-0.0029	0.0132	-0.22	0.824	-0.0020	0.0155	-0.13	0.9
Immigrants	0.0255	0.0213	1.2	0.232	0.0211	0.0260	0.81	0.416
Occupational status (ref: employed)								
Retired from work	-0.0628	0.0274	-2.29	0.022	-0.0580	0.0296	-1.96	0.05
Other not employed	0.0266	0.0274	0.97	0.333	0.0483	0.0312	1.55	0.122
Constant	2.0620	0.0715	28.83	<0.001	3.2483	0.0560	57.97	<0.001

Note: Estimates use post-stratification weights, standard errors obtained by bootstrapping the entire matching procedure (and the computation of the post-stratification weights) 100 times. Age, education, gender and occupational status refer to the reference person of the family. Number of observations: 17,023.

4.3 - Determinants of fuel poverty

To investigate the incidence of fuel poverty in Treviso province, we need to define an indicator of fuel poverty and to complement our dataset with information on household disposable income. As for the fuel poverty index, we mimic the LIHC index suggested by Hills (2012) and described in detail in the recent BEIS Department handbook (2019). The basic idea of the LIHC index is to classify as fuel poor households in relative poverty whose standardized heating costs are above the median standardized heating cost. The indicator does not refer to households' actual expenditure on heating. This might be higher or lower than the standardized costs, either because the family tends to overheat its home or because, faced with the '*eat or to heat'* dilemma, the family lives in an uncomfortably cold environment. Operationally, we classify a household as being in relative poverty if its imputed net disposable income falls below the national relative poverty line, which is provided by the Italian Statistical Office (ISTAT). For fuel costs, we consider the costs, inferred from the EPC, of keeping the dwelling at a constant 20C temperature for 14 hours a day during winter (see Section 2). We depart from Hills (2012), in two ways. First, we do not subtract housing costs (rent or mortgage repayments) from the household disposable income to be consistent with the poverty line we refer to. Second, we consider that the entire accommodation needs to be heated. We show how the dimension of the house relative to household size is a key determinant of fuel poverty (see Table 6).

We impute household disposable income based on the 1,454 observations in the 2015 EU SILC for the Veneto region. We use a log-linear specification for the expected value of total per capita household disposable income as a function of education attainment, occupational status, housing conditions and household demographic characteristics. We estimate the parameters via weighted ordinary least squares using EU-SILC data, and relying on them we impute the income information missing from the enriched EPC dataset. When imputing, we add to the deterministic component a random draw from the normal distribution $N\left(0, s_{\hat{u}}^2\right)$, where $s_{\hat{u}}^2$ is the sample variance of the residuals of the log-linear regression. At the end of the process, we have an integrated dataset of 17,405 EPCs complemented by socio-demographic information on the households (likely) living in those dwellings, and their imputed income. Combining information on disposable income and standardized heating costs to define the LIHC fuel poverty index, we obtain the results showed in Table 5. Our results highlight that 7.87% of households in the Treviso province live in fuel poverty, that is, an estimated total of 27,374 households.

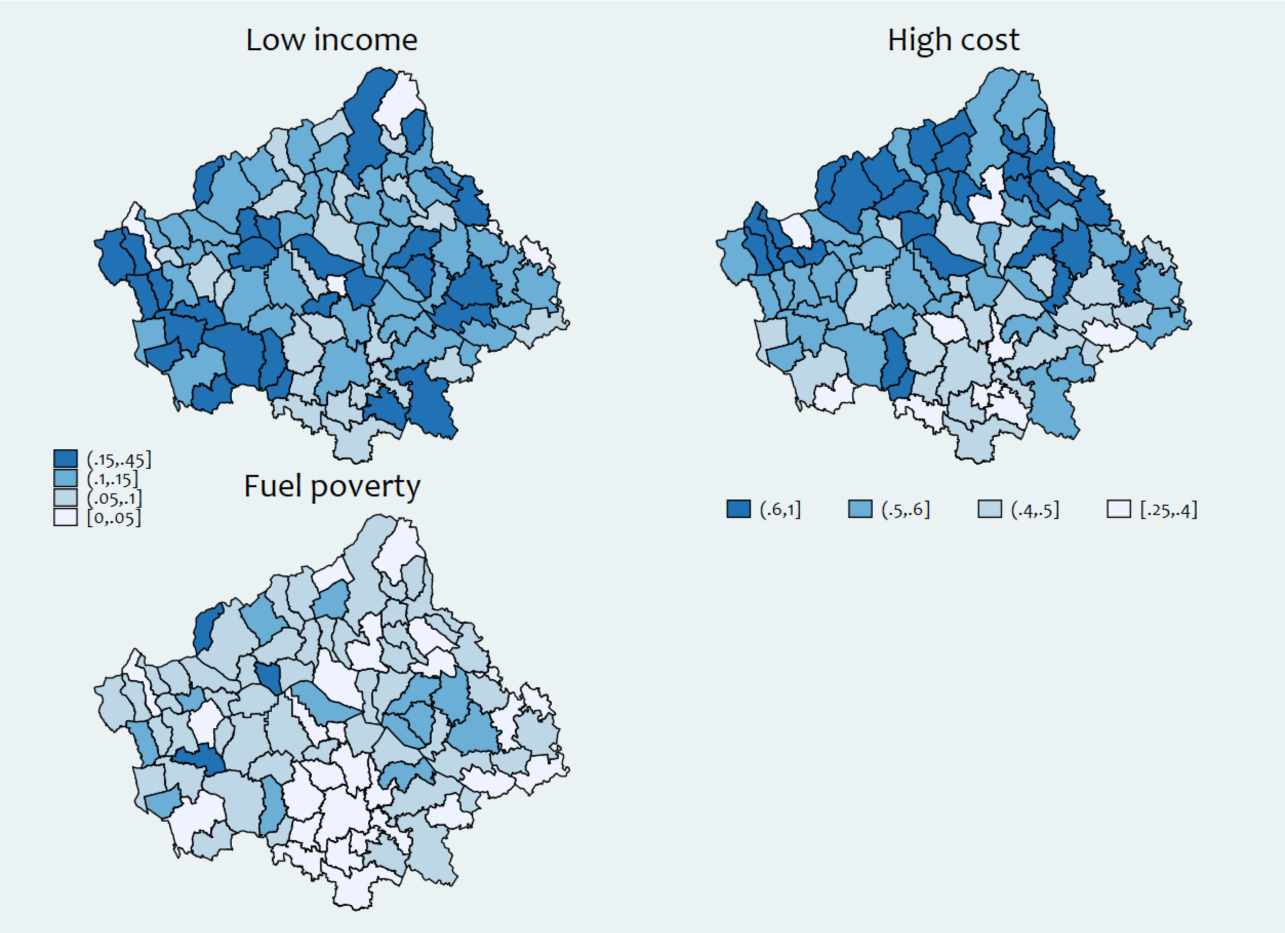
Table 5: Fuel poverty in Treviso province.

	Low heating costs	High heating costs	Total
High income	41.71	42.12	83.83
Low income	8.30	7.87	16.17
Total	50.01	49.99	100

Note: Percentage of households with standardized heating costs below or above the median (low/high heating cost) and equivalent income above or below the relative poverty line (high/low income). According to the LIHC indicator, a family is fuel poor if its income is low and its fuel costs are high, which applies to 7.87% of cases. Statistics use post-stratification weights.

We can tentatively reconstruct the spatial distribution of poverty in the province by exploiting information on the exact location of the EPCs in the dataset. Figure 4 includes three maps showing, for each municipality, the estimated fractions of households with, respectively, low income, high heating costs and that live in fuel poverty according to the LIHC indicator. It is worth stressing that while the map for the high heating costs does not depend on the matching procedure adopted, those showing the shares of households in relative and fuel poverty do depend on that matching procedure and their spatial distribution should be treated with caution. Nevertheless, we think these maps help to identify the areas at highest risk, that is, those areas where heating costs are highest while incomes are relatively low. In fact, according to our results, not all the lowest income municipalities (e.g., many municipalities in the south-west of the province) or those with the highest heating costs (e.g., many municipalities in the northeast of the province) are in or at risk of fuel poverty. Rather, it is the combination of the two (i.e., low income and high heating costs) which leads to a high risk of being in fuel poverty.

Figure 4: Fuel poverty map



Note: Estimated percentage of households with equivalent income below the national relative poverty line (Low income); with standardized heating costs above the provincial median standardized heating cost (High cost), and in fuel poverty, i.e., with low income and high heating costs. Statistics use post-stratification weights.

Table 6 reports the weighted maximum likelihood estimates of a logit model for the probability of a household of being in fuel poverty according to the LIHC indicator and the corresponding average marginal effects.

Table 6: Maximum likelihood estimates of a logit model for the probability of being in fuel poverty according to the LIHC indicator.

					Average marginal effects (AME) on the probability to be in fuel poverty			
	Coef.	std.err.	z	p-value	AME	std.err.	z	p-value
EPC class (ref: A-B)								
C-E	1.2841	0.5328	2.4100	0.0160	0.0236	0.0079	2.9800	0.0030
F-G	2.2160	0.5241	4.2300	<0.001	0.0508	0.0077	6.5800	<0.001
ln(Per capita income)	-6.5585	0.2736	-23.9700	<0.001	-0.1881	0.0070	-26.8300	<0.001
Construction period (ref: pre-1960)								
1960-1969	-0.2123	0.1956	-1.0900	0.2780	-0.0069	0.0064	-1.0800	0.2790
1970-1979	-0.4796	0.2056	-2.3300	0.0200	-0.0148	0.0064	-2.2900	0.0220
1980-1989	-0.5188	0.2431	-2.1300	0.0330	-0.0159	0.0073	-2.1800	0.0290
1990-1999	-0.6163	0.3029	-2.0300	0.0420	-0.0185	0.0088	-2.1100	0.0350
From 2000	-1.4244	0.2949	-4.8300	<0.001	-0.0367	0.0071	-5.1800	<0.001
Surface/Household members (up to 20 sqm)								
21-40	1.3124	0.2967	4.4200	<0.001	0.0248	0.0047	5.2500	<0.001
41-60	2.1397	0.3572	5.9900	<0.001	0.0482	0.0072	6.7100	<0.001
61-80	2.7911	0.3906	7.1500	<0.001	0.0717	0.0101	7.1200	<0.001
81-100	3.4921	0.4546	7.6800	<0.001	0.1023	0.0147	6.9800	<0.001
121-140	3.1613	0.4670	6.7700	<0.001	0.0872	0.0154	5.6400	<0.001
>140	4.4210	0.4534	9.7500	<0.001	0.1519	0.0222	6.8500	<0.001
Primary fuel: natural gas	-0.9900	0.1668	-5.9400	<0.001	-0.0305	0.0057	-5.3400	<0.001
No central heating	0.4445	0.2406	1.8500	0.0650	0.0120	0.0060	2.0000	0.0450
Renewable resources	-0.0714	0.4874	-0.1500	0.8840	-0.0020	0.0136	-0.1500	0.8820
Number of dwellings in the building (Ref: One)								
2	-0.3946	0.1939	-2.0400	0.0420	-0.0112	0.0055	-2.0500	0.0400
3-4	-0.2780	0.2169	-1.2800	0.2000	-0.0081	0.0062	-1.3000	0.1920
5-8	-0.2254	0.2520	-0.8900	0.3710	-0.0066	0.0073	-0.9100	0.3630
9+	-0.0896	0.2042	-0.4400	0.6610	-0.0027	0.0061	-0.4400	0.6590
More than 500 inhab/km^2	-0.2860	0.1821	-1.5700	0.1160	-0.0080	0.0050	-1.6000	0.1100
Plain area	-0.0756	0.1700	-0.4400	0.6560	-0.0022	0.0049	-0.4400	0.6570
Owner occupied	0.4764	0.1948	2.4400	0.0140	0.0131	0.0051	2.5700	0.0100
Family type (ref: single)								
Couple	0.1088	0.3831	0.2800	0.7760	0.0033	0.0115	0.2900	0.7730
Couple with children	-0.1726	0.5647	-0.3100	0.7600	-0.0050	0.0167	-0.3000	0.7630
Single parent	-0.2438	0.4400	-0.5500	0.5790	-0.0070	0.0130	-0.5400	0.5900
Other	-0.5252	0.5787	-0.9100	0.3640	-0.0143	0.0161	-0.8900	0.3730
At least high school	0.0493	0.1818	0.2700	0.7860	0.0014	0.0053	0.2700	0.7870
Age class (ref: at most 40)								
41-65	-0.0316	0.1688	-0.1900	0.8520	-0.0009	0.0047	-0.1900	0.8520
66+	0.3680	0.2822	1.3000	0.1920	0.0110	0.0086	1.2800	0.2000
Female	-0.0501	0.2100	-0.2400	0.8110	-0.0014	0.0060	-0.2400	0.8100
Household size	0.7072	0.4095	1.7300	0.0840	0.0006	0.0046	0.1200	0.9040
Household size^2	-0.1103	0.0463	-2.3800	0.0170				
Immigrants	0.0132	0.2467	0.0500	0.9570	0.0004	0.0071	0.0500	0.9570
Occupational status (ref: employed)								
Retired from work	-0.3643	0.2696	-1.3500	0.1770	-0.0102	0.0074	-1.3800	0.1690
Other not employed	0.1003	0.2671	0.3800	0.7070	0.0029	0.0079	0.3700	0.7120
Constant	50.7757	2.5554	19.8700	<0.001				

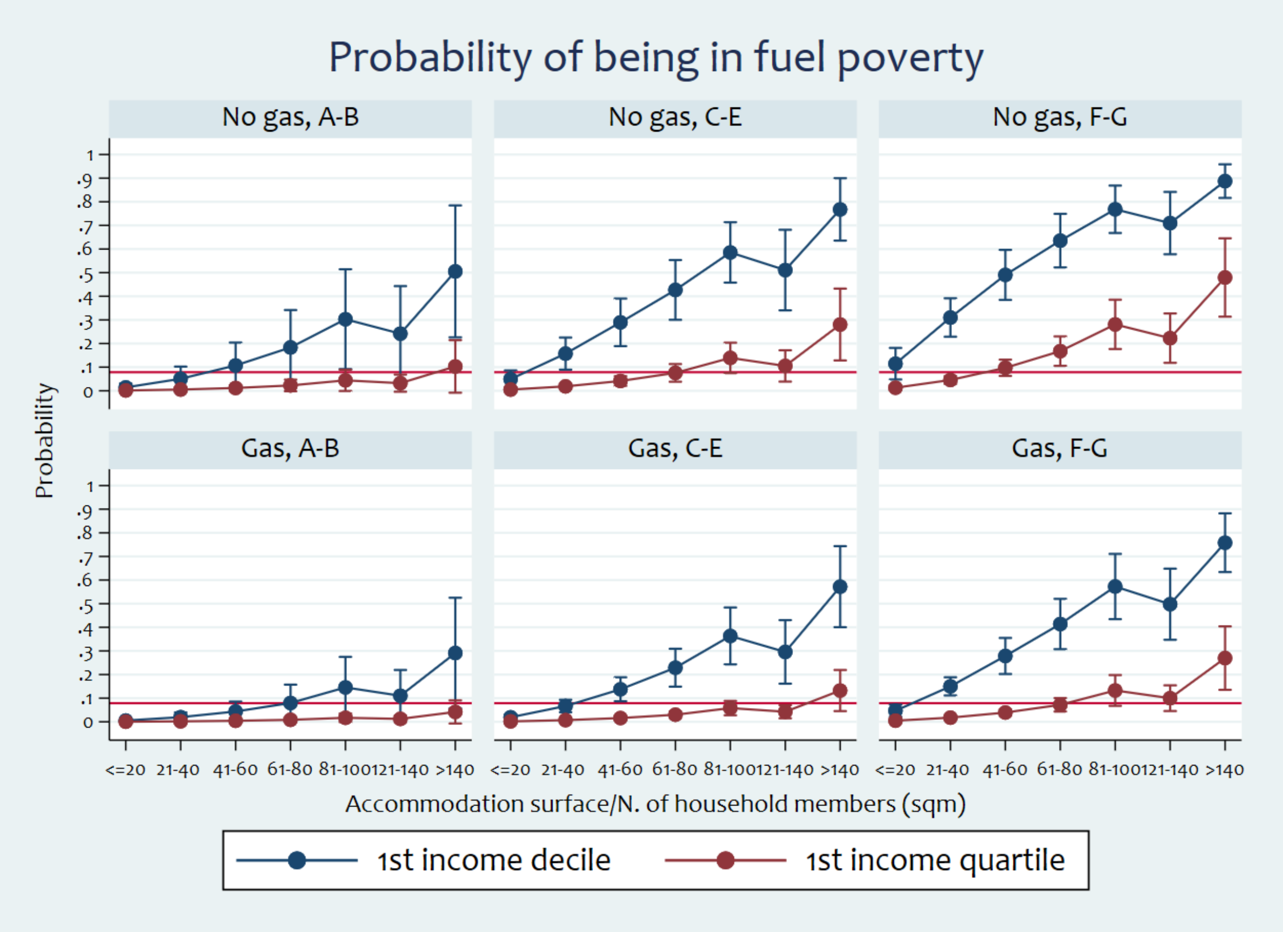
Note: Estimates use post-stratification weights; standard errors obtained by bootstrapping the entire matching procedure, the computation of the post-stratification weights and income imputation, 100 times. A variable with a positive estimated coefficient increases the probability of being in fuel poverty. Age, education, gender and occupational status refer to the household reference person. The last four columns show the weighted average marginal effects on the predicted probability to be in fuel poverty. Number of observations: 17,405.

The estimates and the average marginal effects show that fuel poverty is due to a combination of low income and housing conditions. *Ceteris paribus*, households living in dwellings classed F or G have a 5.08 percentage points (pp) higher fuel poverty risk than households living in the most efficient houses; keeping housing conditions constant, a 10% per capita income increase reduces this risk by 1.9 pp. Even after controlling for income and energy class, construction date and dwelling size relative to the size of the household are crucial factors: on average, the newer the building and the smaller the surface per capita, the lower the fuel poverty risk. Having natural gas as the main heating fuel reduces the risk of fuel poverty by 3 pp on average, whereas homeowners, *ceteris paribus*, have an augmented risk (+1.3 pp). Other household and dwelling characteristics are not significant, on average.

Figure 5 depicts the main results of our analysis. It plots the predicted average probabilities (and confidence intervals) of being in fuel poverty as a function of the surface/occupants ratio. We consider separately six clusters of households, based on use of natural gas (Gas/No gas) and the dwelling energy class (A or B; C to E, and F or G). All the variables are considered at the observed values, except the per capita income, for which we contrast the predicted values when income is at the 10th and 25th percentiles.

Figure 5 highlights the scope of policy interventions for low-income households. In the case of income at the 1st quartile (i.e., the 25th percentile), the risk of fuel poverty for households that use natural gas for their heating is almost always at or below the mean (the horizontal line at 7.87%). Improvements to energy efficiency or a shift to use of natural gas, would reduce the fuel poverty risk appreciably only for those households without natural gas and living in dwellings with more than 60 sqm per capita (9.45% of the total). Overall, Figure 5 suggests that, apart from households living in very large and inefficient dwellings, there is no scope for policy intervention to help households whose income is at around the 25th percentile of the income distribution.

Figure 5: Predicted average probabilities (and 95% confidence intervals) of being in fuel poverty, based on sq metres per capita (horizontal axis) for six clusters of households based on their primary heating fuel (Gas/No gas) and the dwelling's energy class (A or B; C to E, and F or G).



Note: Most of the variables are considered at the observed values with the exception of per capita income; in that case, we contrast predicted values for income at the 10th (1st income decile) and 25th percentiles (1st income quartile). Standard errors obtained by bootstrapping the entire matching procedure, computation of the post-stratification weights and income imputation, 100 times. Number of observations: 17,405.

For incomes in the 10th percentile the picture changes: if the energy class is above B, the risk of fuel poverty is usually higher than average, regardless of dwelling size or main energy source. In this case, the reduced fuel poverty risk associated to a reduction in the (relative) size of the accommodation could be substantial and similar to that achieved by improving building efficiency. Keeping other aspects constant, switching to natural gas could have equally large effects. For instance, a household living in class F accommodation providing 50 sqm per capita, without natural gas and in the 10th percentile of the per capita income distribution, faces a 50% risk of fuel poverty. Moving to a smaller accommodation - say 30 sqm per capita - would reduce the risk to 30%; but this improvement could be obtained - alternatively - by switching to natural gas or improving the dwelling's energy efficiency to class D.

5 - Conclusions

In this paper, we propose a novel methodological approach to investigate the energy efficiency of dwellings and the socio-demographic and economic characteristics of the households residing in them. By exploiting a *nearest neighbour statistical matching procedure*, we construct an integrated dataset which combine the information included in the EPCs with that in the census data. Specifically, we link the publicly registered EPCs to the census records for accommodations in the same municipality, with the same heating systems and characteristics most likely to match the certified dwellings. Finally, we enrich these data with income information imputed from SILC, which resulted in 17,405 records for the Treviso province, a small and densely populated county in the north-east of Italy. Thus, each record in our dataset contains the standardized heating cost of the dwelling as well as the socio-demographic condition and the income of the household most likely to live there.

This dataset is used to estimate the households' risk of fuel poverty at the municipal level. We follow the LIHC index (Hills, 2012), which classifies as fuel poor households in relative poverty whose standardized heating costs are above the median standardized heating cost. Operationally, we classify a household as being in relative poverty if its imputed net disposable income falls below the national relative poverty line published by ISTAT, the national statistical office. In relation to fuel costs, we used the information on the EPCs to quantify the amount of fuel necessary to maintain a standard indoor temperature, given the observed efficiency of the dwelling.

Our findings confirm that measuring fuel poverty is a complex task, and that low income and inefficient housing conditions are the main drivers of fuel poverty. In this perspective, policies aimed at reducing fuel poverty need to consider not only the household income, but also the type of the dwelling's heating system and the efficiency and size of the accommodation relative to the household size. Our results suggest also that such policies should differentiate among municipalities in the same local area (province).

Finally, our analysis highlights the utility of the EPC database to investigate energy poverty in EU member states and design policies to fight it. Reducing fuel poverty in Europe is key for at least two reasons: by increasing buildings' efficiency, on the one hand, it would improve vulnerable households' living conditions; on the other hand, it would contribute to decreasing carbon emissions since energy poor households tend to live in high energy cost dwellings.

References

Caliendo, Marco, and Sabine Kopeinig, 2008. Some practical guidance for the implementation of propensity score matching, *Journal of Economic Surveys*. 22(1): 31-72.

Department for Business, Energy & Industrial Strategy, 2019. Fuel Poverty Methodology Handbook, Accessed 31/12/2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/829010/Fuel_Poverty_Methodology_Handbook_2019.pdf

European Energy Network, 2019. *Position paper on energy poverty in the European Union*, http://enr-network.org/wp-content/uploads/ENERGYPOVERTY-EnRPositionPaper-Energypoverty-Jan-2019.pdf

Fabbri, K., 2015. Building and fuel poverty, an index to measure fuel poverty: an Italian case study, *Energy*. 89: 244–258.

Fuerst, F., McAllister, P., 2011. The impact of Energy Performance Certificates on the rental and capital values of commercial property assets, *Energy Policy*. 39: 6608–6614.

Fuerst, F., McAllister, P., Nanda, A., Wyatt, P., 2016. Energy performance ratings and house prices in Wales: an empirical study, *Energy Policy*. 92: 20–33.

Harsman, B., Daghbashyan, Z., Chaudhary, P., 2016. On the quality and impact of residential energy performance certificates, *Energy Build*. 133: 711–723.

Hills, J., 2012. Getting the measure of fuel poverty. Final Report of the Fuel Poverty Review, CASE report 72. Accessed 31/12/2019. https://www.gov.uk/government/publications/final-report-of-the-fuel-poverty-review

Holt, D. and T. M. F. Smith, 1979. Post Stratification, *Journal of the Royal Statistical Society. Series A (General)*, 142(1): 33-46.

IEA, 2010. Energy Performance Certification of Buildings, www.iea.org.

Las-Heras-Casas, J., Lopez-Ochoa, L.M., Lopez-Gonzalez, L.M., Paredes-Sanchez, J.P. 2018. A tool for verifying energy performance certificates and improving the knowledge of the residential sector: a case study of the Autonomous Community of Aragon (Spain), *Sustain. Cities Soc*. 41: 62–72. https://doi.org/10.1016/j.scs.2018.05.016.

Liddell C, Morris C, McKenzie SJP and Gordon R. 2012. Measuring and monitoring fuel poverty in the UK: national and regional perspectives, *Energy Policy*. 49: 27–32.

Miniaci, R., Scarpa, C. and P. Valbonesi, 2008. Measuring the affordability of basic public utility services in Italy, *Giornale degli Economisti e Annali di Economia*, 121-67/2: 185-230.

Jenkins, D., Simpson, S., Peacock, A., 2017. Investigating the consistency and quality of EPC ratings and assessments, *Energy*. 138: 480–489.

Moore R., 2012. Definitions of fuel poverty: implications for policy, *Energy Policy*. 49: 19–26.

Pasichnyi, O., Wallin, J., Levihn, F., Shahrokni, H. and Kordas, O., 2019. Energy performance certificates — New opportunities for data-enabled urban energy policy instruments? *Energy Policy*. 127: 486–499.

Perez-Lombard, L., Ortiz, J., Gonzalez, R., Maestre, I.R., 2009. A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes, *Energy and Building*. 41: 272–278.

Rosenbaum, P.R. and Rubin, D.B., 1985. Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score, *The American Statistician*. 39(1): 33-38.

Tirrado Herrero, S., 2017. Energy poverty indicators: a critical review of methods, *Indoor and Built Environment*, 26(7): 1018–1031.

Thomson, H., Bouzarovski, S., Snell, C., 2017. Rethinking the measurement of energy poverty in Europe: A critical analysis of indicators and data, *Indoor and Built Environment*, 26(7): 879–901.

Thomson, H., Simcock, N., Bouzarovski, S. and Petrova, S., 2019. Energy poverty and indoor cooling: An overlooked issue in Europe, *Energy and Buildings*, 196: 21-29.

Veneto Region, 2017. Piano Energetico Regionale Fonti Rinnovabili, Risparmio Energetico ed Efficienza Energetica. Venezia: Regione Veneto.