Energy Research & Social Science 58 (2019) 101255

Contents lists available at ScienceDirect

Energy Research & Social Science

journal homepage: www.elsevier.com/locate/erss

Original research article

Letting the Gini out of the fuel poverty bottle? Correlating cold homes and income inequality in European Union countries

Ray Galvin

Behaviour and Building Performance Group, Department of Architecture, University of Cambridge, 1 Scroope Terrace, CB2 1PX, Cambridge, UK

ARTICLE INFO

Keywords:
Gini index
Cold homes
European Union
Economic inequality

ABSTRACT

The last 3–4 decades have seen large increases in economic inequality within high-income countries. A newly emerging strand of research is exploring how this impacts on energy consumption. However, as yet there are no macro-level studies on how economic inequality relates to energy poverty on the national level. This paper investigates whether income inequality in EU countries is correlated with energy poverty by one of its definitions: a household's inability to keep their home adequately warm. Using country-level data for all 28 EU countries for all years 2009–2017, I regress the percentage of households who cannot keep their home adequately warm, against a set of independent variables including: the Gini index as a measure of income inequality; GDP per capita; the number of heating degree days; and several other likely determinants of energy poverty. I find the Gini index, GDP per capita and the number of heating degree days to be persistently strong and statistically significant predictors of the percentage unable to heat their homes, while the effects of the Gini index increase disproportionately as this percentage increases. This suggests that macroeconomic policy needs to be addressed, along with energy efficiency policy, to relieve the discomfort of cold homes in European winters.

1. Introduction

In this paper I explore whether there is a correlation between the level of economic inequality in a country, and the percentage of households in that country who cannot afford to heat their homes adequately. I do this for all EU countries, including the UK (which was still an EU country at the time of writing), and I cover the years 2009–2017.

In exploring this question I seek to bridge a gap between two different strands of energy consumption research. One of these is a newly emerging literature on how economic inequality in high-income countries impacts on energy consumption, production and supply. Perhaps the most overt study within this literature is Galvin and Sunikka-Blank's 'Economic inequality and household energy consumption in high-income countries: A challenge for social science based energy research' [1]. I will refer to this study at some length in this paper, as well as others that cover similar and related themes. The theme of economic inequality in high-income countries is of great relevance today because levels of inequality have increased markedly over the past 3–4 decades in these countries [2–5]. As Galvin and Sunikka-Blank [1] point out, this major shift in society is not yet well reflected in social science based energy research.

The strand of literature at the other end of the bridge I wish to build is that on energy poverty. This very large and burgeoning literature puts forward a number of different definitions of energy poverty, as I explain in more detail in Section 2. I use one of these definitions in this paper, namely 'being unable to afford to heat one's home adequately' –. I make no claim that this definition is better than, more useful than or superior to any of the other definitions of energy poverty on offer. I will argue, however, that it is the best one to start with for a study of this kind. One reason a study such as this is important in the current climate of heightened economic inequality is that it is a *macro*-level analysis. That is, it is concerned with issues at country level, rather than at individual or household level. It is on the macro level that laws are made, tax rates are decided, and social welfare programs are adopted and administered, and it is on the macro-level that economic inequality is thereby maintained, increased or mitigated. It is abundantly clear that the steep increase in economic inequality of the past 3–4 decades is the result of macro-level policies, especially to do with taxation, social welfare payments, and monetary policy [3,5–7]. If we find, in a study such as this, that there is a clear correlation between the level of economic inequality and the percentage of households who cannot afford to heat their homes adequately, a major part of the solution to these households' problem might be for government to adjust these policies so as to bring about a reduction in the level of inequality. Hence the findings of a macro-level investigation can have direct implications for government at the national level.

E-mail address: rg445@cam.ac.uk.

https://doi.org/10.1016/j.erss.2019.101255
Received 30 January 2019; Received in revised form 5 August 2019; Accepted 9 August 2019
Available online 22 August 2019
2214-6296/ © 2019 Elsevier Ltd. All rights reserved.

The core of this paper is a set of multivariate statistical analyses. The dependent variable in these analyses is the percentage of households who cannot afford to heat their homes. One of the independent variables – the one of greatest interest in this paper – is the Gini index. The aim of the analyses is to find whether there is a strong correlation between the Gini index and the percentage unable to heat their homes, and if so, whether this holds up alongside other likely influences on the percentage unable to heat their homes. I therefore include further independent variables that seem appropriate, i.e. which seem likely to also influence the percentage of households who cannot heat their homes. These are: GDP per capita; the number of heating degree days; the percentage of households who are in arrears on their utility bills; and the percentage of inadequate homes, i.e. those with a leaking roof, damp walls, floors or foundation, or rot in window frames or floor. I explain below why these variables are deemed appropriate, and also why various others have not been included.

In order to improve the robustness of the analysis, I make use of a number of statistical features which – to spare readers who find statistical analyses challenging – I mostly present and discuss in the Appendix. These include various statistical tests; discussion of logarithmic transformations and re-transformations of variables; the combined use of both ordinary least squares (OLS) regressions and quantile regressions; and tests for best model specification.

I use the regression analyses to test a hypothesis, namely, that there is a robust correlation between the Gini index and the percentage of households who are unable to heat their homes, in the EU28; and that variations in this percentage are robustly associated with variations in the Gini index. If the Gini passes the tests of robustness, this will not *prove* the hypothesis is true, but it will show it is a strong hypothesis which has stood up to rigorous testing and should therefore be taken seriously by policymakers and energy researchers.

Section 2 of this paper presents brief literature reviews on the two strands of discussion this paper seeks to bridge. Section 3 describes the data and methods, and gives the descriptive statistics. Section 4 gives the results and their detailed interpretation. The results are discussed in Section 5. Section 6 concludes and offers recommendations for policy and research.

2. Literature review

As noted in Section 1, this paper attempts to forge a tentative bridge between two strands of energy-related literature: the newly emerging literature on economic inequality and energy in high-income countries; and literature on fuel poverty.

2.1. Literature on economic inequality and energy in high income countries

Many researchers have recently drawn attention to the large increases in economic inequality that have arisen in most high-income countries over the past 3–4 decades [3,5,6,8–12]. One might think this presents a fertile field for energy consumption research, since energy costs money. However, Galvin and Sunikka-Blank [1] point out that this issue has not yet penetrated energy consumption studies to a great extent, particularly in energy studies that are embedded in social science disciplines. Galvin and Sunikka-Blank suggest there are two possible reasons for this. Firstly, ever since the *Methodenstreit* (dispute over method) of the early-mid 20th century, economics and sociology have tended to go their separate ways, pursued as different disciplines, so that sociologists have left theorizing about money to economists [13]. Secondly, they argue that the social science frameworks prominently used in energy consumption studies today, such as practice theory [14–17], discourse analysis [18–20] sociotechnical systems theory [21,22] and actor-network theory [23], are rooted in social theories of the 1950s–1980s, a period which was characterised by historically low levels of economic inequality in high income countries. In other words, the social theory used in energy consumption studies today has simply not caught up with the huge increases in economic inequality that have made today's society very different from that of the latter decades of the 20th century.

Nevertheless Galvin and Sunikka-Blank point out that several authors have recently begun to draw attention to specific impacts of increased economic inequality on energy consumption in wider society. Firstly, a group associated with Andrew Jorgenson have offered a series of analyses on the correlation between countries' levels of economic inequality and their CO_2 emissions [24–26]. These scholars have shown that each 1% increase in wealth inequality is associated with a 0.8% increase in the level of CO_2 emissions from consumption, in high income countries. This finding is confirmed in further studies by Chancel and Piketty [27], who also calculate that CO_2 emissions from the consumption of the wealthiest strata of society may be 10–20 times higher than those of the average person.

Secondly, a number of scholars have begun to relate housing tenure and access to energy services such as home heating, to economic inequality in high income countries. In somewhat simplified form, their argument runs: (a) as economic inequality has increased, home ownership rates among young and middle-aged adults have decreased [28–32]; (b) this is also associated with reducing rates of social housing [33]; (c) these shifts cause more and more young and middle aged people to have to rent from private landlords [34]; and (d) private landlords tend to thermally upgrade their properties significantly less than social landlords and home owners [35–38]. This leads to the tentative conclusion that economic inequality in high income countries leads to a form of energy poverty: poorer households suffer the double disadvantage of being more likely to live in thermally inadequate houses, while having lower income to pay for heating.

Thirdly, Galvin and Sunikka-Blank present original work indicating that women in high income countries are more deeply disadvantaged than men, with respect to energy services, through increasing economic inequality. Their evidence includes a plot of the ratio between the percentages of single parent households who cannot heat their homes adequately, and all households who cannot heat their homes adequately, in EU countries. In 2016 the UK showed the highest ratio, at just below 3, while Italy showed the lowest, at just over 1. Galvin and Sunikka-Blank put this together with data that indicates that around 95% of single parent families are headed by a woman; that women have significantly lower average income and wealth than men [39,40], and that they are much less likely to own their own home mortgage-free [41].

Fourthly, these authors suggest that the incidence of energy poverty (which they call fuel poverty) is highly likely to increase with economic inequality, as inequality increases the number of poor households, and household poverty is widely seen as one of the drivers of energy poverty [42]. They further point to a recent EU-funded survey that found poorer households often cause higher CO_2 emissions than average income households because they cannot afford basic energy-saving technology like insulation and new appliances [43].

Galvin and Sunikka-Blank argue that social science-based energy research needs to take fuller account of the effects of economic inequality, not just because higher levels of economic inequality imply there are more poor households, but also because large inequalities of income and wealth can distort markets, including energy markets, making it harder for the poor to compete for basic energy services.

Regarding these emerging insights, no study has yet explored whether there are direct correlations between a measure of income inequality such as the Gini index, and an indicator of energy poverty on the macro level such as the percentage of households unable to keep their homes adequately warm.

2.2. Relationship to energy poverty studies

A comprehensive review of the vast literature on energy poverty would be way beyond the scope of this paper. Here I simply discuss

some of the strengths and weakness of using 'the percentage of households unable to heat their homes adequately' in a study such as this.

There are many insightful and informative studies on energy/fuel poverty in the EU using different definitions of energy poverty and from a range of different perspectives (see reviews in [44–48]). Most of these are in broad agreement that low income, expensive fuel and thermally inadequate dwellings are the three main contributors to fuel poverty [42,49,50]. Some, such as [51] and [52], add a fourth determinant, 'specific household energy needs'. Other studies note the diversity among EU countries, not only of the determinants of fuel poverty (e.g. [45]) but also its within-country geographical diversity (e.g. [53]).

A basic, recurring theme in this literature is the difficulty of defining energy (or fuel) poverty [46]. Castaño-Rosa et al. [44] provide perhaps the most recent survey and analysis of the more widely used approaches. Broadly speaking, the definitions in use can be divided into two types: objective and subjective [54]. Objective definitions consider measurable parameters to do with income, expenditure, household composition, buildings, etc. An example is the UK's relatively new metric, arising from the Hills Review [55,56], that a household is in fuel poverty if it has 'lower than average income and higher than average fuel costs' [42]. Other objective-type definitions among EU countries draw on parameters such as: the ages of household members; an objective measure of the thermal quality of dwellings; and whether or not a household is receiving welfare benefits [52,54,57,58]. What these measures have in common is that they seek to indicate whether a household *can be expected* to achieve an acceptable level of energy services, rather than a household's subjective judgments of relevant aspects of their situation, such as whether they find they are able to heat their home adequately.

These methods generally use a 'micro' or 'targeting' approach to identify the types and/or locations of households where energy poverty is most prevalent or households are most vulnerable (see summaries in [47,50]. They aim to provide policymakers with detailed information at fine levels of granularity, to enable interventions to be targeted economically efficiently toward vulnerable households (e.g. [59]). This is an important endeavour because resources for interventions are always limited and need to reach the people in need.

However, a disadvantage of these 'objective' approaches is that they are not able to take account of householders' actual, expressed needs in relation to energy use, which may be significantly different from the perceived needs embodied in the choice of parameters [60]. A further assumption of such approaches often appears to be that energy-poor households are passive receptors of targeted help, rather than actors who could be empowered to lift themselves out of fuel poverty if key elements of the macro-economic context were changed.

Subjective definitions, on the other hand, ask households one or more pertinent questions on whether they see their energy services as adequate. The best-known questions of this type in EU countries are those used in surveys for the Eurostat Statistics on Income and Living Conditions [61]. These cover:

- Whether the household is unable to heat their home to a satisfactory level;
- Whether they are in arrears on fuel bills; and
- Whether the building fabric of their dwelling is inadequate (leaking roof; damp walls, floors or foundation; or rot in window frames of floor).

A disadvantage of subjective measures is that households' responses may not be honest. Boardman ([62], cited in [52]) found people often deny being unable to heat their homes when questioned, often due to pride. Further, different groups or cultures may have different ideas of what thermal comfort is [63] or indeed whether there is any need to be comfortably warm [64]. The few studies which have used both objective and subjective approaches on the same households find overlaps between the two but also significant differences [57,60,65].

Nevertheless, an important advantage of a subjective approach is that it identifies households who feel cold because they *actually* cannot heat their homes, rather than those who *can merely be expected* to feel cold. These households experience the deprivations of energy poverty, at least from their own point of view, regardless of what their income and expenditure levels indicate and how sound their dwelling is [54].

Finally, an advantage of subjective approaches is that they lend themselves to studies on the macro-level, i.e. they yield whole-country results, enabling us to compare countries.

Bringing these two strands of literature together, I suggest it is worth investigating the relationship between the level of economic inequality in a country, and the level of energy poverty in that country. For the latter we need a country-wide indicator. One such indicator is the percentage of households who are unable to heat their home adequately. Because Eurostat surveys and data collection use consistent methods to give us both of these measures for all EU countries, along with other relevant indicators, this provides a reliable dataset for such a study to be performed.

3. Method and data

3.1. Which variables are relevant?

All the data used in this study are from Eurostat statistics (Eurostat, 2019). They cover all 28 EU countries in all years 2009–2017. The variable I use as the dependent variable in my analyses is the percentage of households in each of the 28 countries, in each of these years, who say they are unable to keep their homes adequately warm (abbreviation UH%, symbol ***U***). For reasons explained below, I transform this to its natural logarithm, Ln(UH%) for the regression analyses. I regress this against a number of independent variables which are also likely to influence UH%. I treat the dataset as pooled panel data, as explained below and in the Appendix. I perform all regressions and statistical tests using the software package Stata.

The independent variable which is the focus of this study is the level of income inequality (not wealth inequality) *after tax and welfare transfers*. It is therefore based on the *actual* income households can spend. It is represented by the Gini index for this definition of income, calculated by Eurostat and given as a percentage rather than a decimal. I give this the symbol ***G***.

The Gini index indicates the degree of discrepancy between people's income levels in relation to each other within a particular country, but it does not show how great their income is on average compared to those in other countries. Energy poverty literature consistently affirms that low income is one of the causes of energy poverty [48,66–68]. A variable that captures average household income reasonably well is real GDP/capita, i.e. GDP/capita adjusted for purchasing parity, as given in Eurostat statistics. I give this the symbol ***D***.

Ireland and Luxembourg have large flows of tax haven money, which is included in official figures for GDP/capita but does not translate into household income. Ireland's GDP/capita in 2016, of €65,207, is about 43% higher than it would be without these international banking flows [7], so I reduce Ireland's GDP/capita accordingly. Due to banking secrecy it is not possible to obtain figures for Luxembourg, but I reduce Luxembourg's GDP/capita by 30% as a rule of thumb.

Another determinant of households' ability or inability to heat their homes is the number of heating-degree-days in each country in each year (HDDs, symbol ***H***) as this is a measure of how cold, for how long, a country is on average during the winter [69]. There are two different ways this might correlate with UH%. On the one hand, for a particular country over a period of years, we could expect a positive correlation, controlling for other variables: in colder years more people would find it harder to heat their homes [70]. On the other hand, when differences between the 28 countries are considered, the correlation is likely to be

negative. This is because countries with colder climates, like Finland and Sweden, have far more stringent thermal building regulations and practices than warmer countries like Malta, Portugal and Cyprus [71]. Since the variations between the North Sea and the Mediterranean are far greater than the within-country variations over 9 years, the negative effect is likely to dominate.

Another likely influence on households' ability to heat their homes is their ability to pay their basic household bills [53]. This would not be captured by a debt-indicator variable such as the ratio of debt to income, because households in wealthy countries actually have far higher debt/income ratios than households in poorer countries. Denmark's was the highest in the EU in 2016, at 2.45, while Bulgaria's was the lowest, at 0.33 [61]. Wealthy households tend to use debt as a wealth accelerator, taking out loans to invest in shares, business ventures and real estate [72–75]. A more useful indicator of households' ability to pay their basic bills is the Eurostat data on percentages of households in arrears on their utility bills. Presumably, such households are forced to be frugal with heating. I use this as an independent variable (abbreviation 'arrears'; symbol ***A***). A caution is that it might correlate with the other independent variables in a similar way to the dependent variable, UH%. Hence I test for multicollinearity using variance inflation factor estimates, as outlined in the Appendix.

As noted above, the thermal quality of homes is regarded as one of the main determinants of energy poverty literature [42,50]. There is no macro-level variable that adequately captures this, but some of it would be captured by a variable that indicates the prevalence of substandard housing, particularly in relation to thermal quality. Eurostat's household surveys ask households whether their dwelling has a leaking roof, damp walls, floors or foundation, or rot in window frames or floor. The effects of this variable may not be very strong or consistent because different respondents may have different thresholds for what they call rot, damp or a leak, and it does not capture differences in thermal quality between dwellings that are *not* regarded as 'inadequate' [76]. Nevertheless, we would expect the percentage of households in substandard dwellings to be positively correlated with the percentage unable to heat their homes. I abbreviate this variable as 'bad homes %' and give it the symbol ***I*** (for 'inadequate').

There are several further variables which may be relevant to the analysis, but could or should not be included. One of these is energy price, since this is widely understood to be one of the three main determinants of energy poverty [42,50]. However, the types of fuel used for home heating vary enormously throughout the EU as well as within regions [77,78], and it was not possible to combine these in a rational, credible way on a country basis.

Secondly, household wealth is a likely determinant, since it influences the quality of dwelling a household can buy, or their ability to invest in energy efficient technology such as insulation, new windows or an efficient heating system [43]. However, statistics on average household wealth in EU countries showed it to be strongly correlated with GDP/capita (see test result in the Appendix). These two variables could not therefore be used in the same multivariate analysis. A similar issue applied with the variable "average level of household debt", which was mentioned above.

The unit of analysis in this paper is the 'country-year'. For example, Bulgaria in 2010, Bulgaria in 2017 and Germany in 2017 are three different country-years. Measures like the Gini index, GDP per capita and the percentage of households unable to heat their homes vary, not only from country to country, but also from year to year. This provides a much larger dataset to work with, than just 28 countries in one year. Since the analysis covers all 9 of the years 2009–2017 – This gives $28 \times 9 = 252$ country-years, i.e. 252 statistical observations. Note that each country therefore carries the same weight in the analysis, regardless of its population or wealth.

3.2. Characteristics of the variables

In an OLS regression analysis, the results are easiest to interpret if all the variables are normally distributed, or at least close to this. In the Appendix, I show that all but two of the variables used in this analysis fit those criteria. The other two, UH% and Arrears %, are strongly right-skewed, and the Appendix shows how the natural logarithm (log) transformations of these produce useable, normally distributed variables. Hence I use the log transformations of these variables in the analysis. This requires the results of the analysis to be re-transformed accordingly, as shown in Section 4.

In the Appendix I also give tests for skewness[1] and kurtosis[2] of the variables' distributions, showing that their deviations from normal and symmetrical are minimal. Nevertheless, because none of the variables are perfectly normally distributed, in the Appendix I give a further check on the robustness of results, using quantile regression [79]. I then use a quantile regression alongside OLS regression in the main analysis (Section 4) to compare results using the two methods.

Finally, because I am using a set of countries over a number of years, the data counts as "pooled panel data". In the Appendix I discuss the issues involved in this, and show why this approach is appropriate with this data set.

3.3. The regression equations

Given the above discussion and the tests in the Appendix, the basic regression equation is:

$$\mathrm{Ln}(U_{i,t}) = B_0 + B_1 G_{i,t} + B_2 D_{i,t} + B_3 H_{i,t} + B_4 \mathrm{Ln}(A_{i,t}) + B_5 I_{i,t} + er \quad (1)$$

where B_0 is the intercept; B_1 etc. are the correlation coefficients; i are the individual countries; t the years; and er is the error term.[3] As mentioned above, G is the Gini coefficient expressed as a percentage; D is GDP/capita in euros, adjusted to purchasing power parity; H is the number of heating degree days; A is the percentage of households in arrears on their energy bills; and I is the percentage of inadequate homes. The symbol "Ln" means "the natural logarithm of".

Because the Gini index is the focus of the study, separate regressions were also performed for each year, to check whether the Gini was a significant predictor of Ln(UH%) on a year-by-year basis. For each separate year the regression equation is simply:

$$\mathrm{Ln}(U_i) = B_0 + B_1 G_i + B_2 D_i + B_3 H_i + B_4 \mathrm{Ln}(A_i) + B_5 I_i + er \quad (2)$$

3.4. Descriptive statistics

Table 1 gives the mean, median, maximum and minimum values of the variables over the pooled panel data. For the sake of interest, both the log and non-log versions of UH% and Arrears % are given. To give the reader a feel for the scope of the variables and how they relate to specific countries, the full data for 2017 are given in Table 2.

[1] Skewness is a measure of how lopsided a statistical distribution is. For example, the distribution of wages in the UK is highly right-skewed, meaning that most people earn around the median wage, but some earn many times the median and some earn many thousands of times the median.

[2] Kurtosis is a measure of the 'peakiness' of a statistical distribution.

[3] The error term is the difference between the modelled values of $\mathrm{Ln}(U_{i,t})$ and the values of $\mathrm{Ln}(U_{i,t})$ in a theoretical larger population of country-years, from which the 249 observations are drawn. See comment in the Appendix about the difference between errors and residuals.

Table 1
Descriptive statistics for variables used in the regressions, EU27 + UK, 2009–2017.
Source: author's calculations based on raw data from Eurostat (2019).

Variable	Observations	Mean	Std. Dev.	Min	Max
UH%	250	11.404	11.1987	0.3	66.5
Ln(UH%)	250	1.9694	1.0193	−1.2039	4.1972
Gini	250	29.98	3.738	22.7	40.2
GDP/cap	252	24806	15836	5000	81000
HDDs	252	2853	1159	322	6190
Arrears %	249	11.624	8.9836	1.7	42.2
Ln(Arrears %)	249	2.1635	0.7770	0.5306	3.742
Bad Homes %	250	16.06	6.515381	4.2	34.7

4. Results and interpretation

4.1. Results of pooled panel data

Table 3 gives the results of the regressions that use pooled panel data. The first column of results (Model 1) is from OLS regression on the means; the second (Model 2) is from quantile regression on the median (performed at quantile 0.5 – see explanation in Appendix). The non-bracketed figures are the regression coefficients; the figures in brackets are the standard errors (see Appendix); and the asterisks represent levels of statistical significance (*** = p < 0.01, ** = p < 0.05, * = p < 0.1 – see note below on p-values). For both Model 1 and Model 2, all five independent variables are statistically significant, with the Gini index, GDP/capita, HDDs and Ln(Arrears%) significant at the 0.01 level and Bad homes % at the 0.05 level. The quantile regression (Model 2) is included merely to check that the coefficients are reasonably close to those of the OLS regression. Comparing the OLS and quantile results in Table 3 indicates that these are quite close, indicating that any deviations of variables from normal and symmetrical are not sufficient to call the OLS results in question. However, as noted in the Appendix, because of difficulties with quantile regression of pooled panel data alluded to above, only the results for OLS regression are used in the discussion of results.

The adjusted R-squared value for Model 1 (given at the bottom of Table 3) is moderately high, at 0.794, indicating that the model explains 79.4% of the variance in Log(UH%). The F-statistic, at 0.000, indicates that the model predicts the sign (+ or –) of the dependent variable extremely well. Hence in these respects the model gives a very good fit to the data.

Using the regression coefficients in Table 3, first column, the model of choice, called 'Model 1', is therefore:

$$\mathrm{Ln}(U) = 0.209 + 0.0840G - 2.62 \times 10^{-5}D - 2.86 \times 10^{-4}H + 0.240\mathrm{Ln}(A) + 0.0117I + er \quad (3)$$

Immediately we see, from the regression coefficients in both models, that the Gini index is positively correlated with Ln(UH%) and is strongly significant, with p-value lower than 0.01. In other words, *a country's Gini index (its level of economic inequality) is a strong predictor of its percentage of households who cannot heat their homes adequately, controlling for other relevant variables.*

We also see that GDP/capita is negatively correlated with Ln(UH%), as would be expected: the higher a country's average income, the more money households have to spend on heating and energy efficiency measures, so the lower the percentage who cannot heat their homes.

It is interesting that the number of heating degree days (HDDs) is also negatively correlated with Ln(UH%). This is most likely because, as explained in Section 3, countries with colder climates tend to have stricter thermal regulations.

The remaining two variables, the logarithm of the percentage in arrears on their utility bills and the percentage of households living in inadequate dwellings, are both positively correlated with Ln(UH), as would be expected (see Section 3).

I discuss these results in more detail in Section 4.3 below. Prior to this, however, two other issues need to be mentioned.

Table 2
Values of all variables for 2017. Note that all Ireland's figures except GDP/capita and HDDs are for 2016, and Ireland's and Luxembourg's GDP/capita figures are 0.77 of the official values, to take account of tax haven transfers that do not enrich households on average.
Source: author's calculations based on raw data from Eurostat (2019).

Country	UH%	Ln(UH%)	Gini	GDP/cap (PPP)	HDDs	Arrears%	Ln(Arrears %)	Bad Homes%
Belgium	5.7	1.74	26.0	35000	2580	4	1.39	18.5
Bulgaria	36.5	3.60	40.2	6300	2533	31.1	3.44	12.2
Czechia	3.1	1.13	24.5	17200	3310	2.1	0.74	8.0
Denmark	2.7	0.99	27.6	47100	3117	3.5	1.25	14.9
Germany	3.3	1.19	29.1	35500	2964	2.8	1.03	12.5
Estonia	2.9	1.06	31.6	14600	4208	6.3	1.84	13.9
Ireland	5.8	1.76	29.5	43428	2670	12.1	2.49	13.4
Greece	25.7	3.25	33.4	17400	1658	38.5	3.65	13.5
Spain	8.0	2.08	34.1	24500	1598	7.4	2.00	11.5
France	4.9	1.59	29.3	32300	2338	6.1	1.81	11.1
Croatia	7.4	2.00	29.9	11500	2331	21	3.04	11.4
Italy	15.2	2.72	32.7	26400	1878	4.8	1.57	16.1
Cyprus	22.9	3.13	30.8	22700	721	13.7	2.62	29.3
Latvia	9.7	2.27	34.5	11600	4016	11.9	2.48	22.8
Lithuania	28.9	3.36	37.6	12700	3830	7.9	2.07	15.7
Luxembourg	1.9	0.64	30.9	61831	2870	1.7	0.53	17.4
Hungary	6.8	1.92	28.1	11800	2742	13.9	2.63	24.8
Malta	6.6	1.89	28.3	20800	485	5.6	1.72	8.3
Netherlands	2.4	0.88	27.1	40700	2544	2.1	0.74	13.5
Austria	2.4	0.88	27.9	37200	3503	3.6	1.28	11.9
Poland	6.0	1.79	29.2	11800	3290	8.5	2.14	11.9
Portugal	20.4	3.02	33.5	17500	1055	5.6	1.72	25.5
Romania	11.3	2.42	33.1	8300	2916	15.9	2.77	11.1
Slovenia	3.9	1.36	23.7	19400	2833	14.3	2.66	22.0
Slovakia	4.3	1.46	23.2	15000	3280	5.5	1.70	6.7
Finland	2.0	0.69	25.3	35700	5524	7.8	2.05	4.2
Sweden	2.1	0.74	28.0	42800	5220	2.2	0.79	7.0
UK	5.9	1.77	33.1	32200	2865	5	1.61	17.0

Table 3
Results of regressions using pooled panel data of EU27 + UK, 2009–2017. First column is OLS regression about the means; second is quantile regression on the median. Coefficients are given with significance levels (*** p < 0.01; ** p < 0.05; * p < 0.1), and standard errors in parentheses.
Source: author's calculations based on raw data from Eurostat (2019).

Variables	Model 1: Standard OLS regression Ln(UH%)	Model 2: Quantile regression on median Ln(UH%)
Gini	0.0840***	0.0909***
	(0.00880)	(0.0108)
GDP/cap	−2.62e-05***	−2.49e-05***
	(2.47e-06)	(3.02e-06)
HDDs	−0.000286***	−0.000268***
	(2.67e-05)	(3.27e-05)
Ln(Arrears%)	0.240***	0.254***
	(0.0546)	(0.0668)
Bad Homes	0.0117**	0.0121**
	(0.00494)	(0.00604)
Constant	0.209	−0.0598
	(0.306)	(0.375)
Observations	249	249
R-squared	0.810	
Adjusted R-squared	0.794	
P > F	0.000	

4.2. P-values and robustness checks

Firstly, the meaning of 'statistical significance', indicated by p-values, needs clarification. The p-value of each independent variable gives the probability that the sign (+ or –) of the regression coefficient is not correctly predicted by the model, i.e. it is indeterminate[4], given the size of the sample. Clearly the sign needs to be determinate – either positive or negative – if the results are to have any meaning. If we are only interested in the existing data from these 28 countries over these 9 years, the p-values are of no consequence, because in such a case the sample itself is the statistical population, and the sign and value of the coefficients are simply those given in the results. However, if we want to use the results to make more general inferences about how the independent variables might correlate with the dependent variable in other years, we are dealing with a sample from a larger population, and p-values are therefore essential indicators of the significance of the results. Because I will make general inferences based on the results, I treat the p-values as important throughout.

Secondly, in order to be even more confident that the results are to be trusted, the Appendix gives further tests for the robustness of Model 1, namely tests for multicollinearity, correct specification of the model, and randomness of residuals. As shown there, there were no significant effects of multicollinearity; and the specification of the model is a good compromise between a serviceable model and a more perfect but highly complex model that would be difficult to interpret.

The tests for residuals, however, raise some concern. These test results suggested there may be a further variable, unaccounted for in the analysis, which drives Ln(UH%) higher at its top range, and lower at its bottom range, than is indicated by the independent variables used in the analysis. This effect was not sufficient to call the robustness of the model into question. However, it raised the interesting possibility that some other factor may be exacerbating householders' inability to keep their homes warm in countries where the percentage of such houses is already high, and doing the opposite in countries where it is low. Such a factor or set of factors may well show up in future research.

[4] Many popular texts state that the p-value gives the probability that the coefficient zero, but this is misleading.

4.3. Interpreting the regression coefficients

As I demonstrated above and in the Appendix, Model 1 gives a robust account of how the independent variables correlate with the logarithm of the percentage of households unable to heat their homes, in EU countries in the years 2009–2017. The regression coefficients of these variables therefore become the focus of interest. I now discuss these in detail.

To begin with, because the independent variable is in log form, the regression coefficients need to be interpreted carefully. The coefficient of the Gini index, at 0.0804, means that for each one percentage point increase in the Gini, the natural logarithm of the percentage unable to heat their homes (UH%) increases by 1. Because this is a logarithmic scale, the higher the value of UH%, the greater the rate of increase in UH% for a one percentage point increase in the Gini. Controlling for the other independent variables (i.e. holding them all constant), the modelling equation shown as Eq. (3) above becomes:

$$\ln(U) = 0.0804G + K \tag{4}$$

where K is the intercept in Eq. (3) plus the sum of the products of the coefficients of the other variables and their values for any particular country-year. It is now important to find how the percentage of households who cannot heat their homes varies with different values of the Gini coefficient. To do this we first make U the subject of Eq. (4) (by taking the exponential of both sides of the equation), then differentiate U with respect to G. The rate of change of U with G is thereby shown to be given by:

$$\frac{\partial U}{\partial G} = 0.0804e^{K} \bullet e^{0.0804G} \tag{5}$$

What does this mean in practice? Consider, for example, a country-year with a mid-range percentage unable to heat their homes, such as Italy in 2017, where the percentage unable to heat their homes (U) was 15.2. Substituting the values of the variables from Table 2 in Eq. (5) we get K = 0.0909. Substituting this in Eq. (4) gives $\frac{\partial U}{\partial G} = 1.220$. Hence for Italy in 2017, each one percentage point increase in the Gini is associated with an increase of 1.22 percentage points in the number of households who cannot heat their homes.

Now compare this to a country-year with high UH%, such as Lithuania in 2017, where $U = 28.9$. Here, following the same method as above, we find $K = 0.337$ and $\frac{\partial U}{\partial G} = 2.85$. Hence for Lithuania a one-percentage increase in the Gini is associated with more than twice the increase in the percentage unable to heat their homes, as in the case of a middle-range country like Italy. For Bulgaria in 2017 this is yet more extreme. Bulgaria in 2017 has the highest percentage unable to heat their homes and the highest Gini, giving it the highest $\frac{\partial U}{\partial G}$, at just under 3.0. This means that for Bulgaria, a small increase or reduction in the Gini coefficient is associated with a very large increase or reduction in the number unable to heat their homes.

For countries with low UH% the effect is smaller. For example, for Sweden in 2017, where UH% was 2.1%, we find K = -1.511, and $\frac{\partial U}{\partial G}$ is a mere 0.169. Values of $\frac{\partial U}{\partial G}$ for all countries in 2017, based on the coefficients for the pooled dataset, are displayed in Fig. 1.

The simple lesson is that the higher the percentage of households unable to heat their homes, the greater the effect on this of a change in the Gini index. Note that this is not due to any non-uniformity in the distributions of the variables used in the regressions – it is not a fault in the modelling – but to the characteristics of a logarithmic relationship between UH% and the Gini index.

The same dynamic occurs with the second independent variable, GDP/capita, though in the opposite direction. The coefficient, −3.10e-05 (i.e. -3.10×10^{-5}), means that each increase in GDP/capita of €1000 is associated with a decrease of 0.031 in Ln(UH%). Controlling for the other independent variables and considering units of €1000 rather than €1, the modelling equation is:

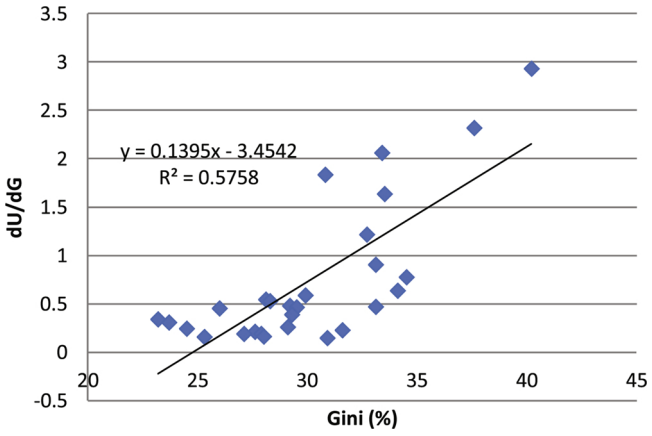

Fig. 1. Modelled effect of change in Gini on UH%, EU27 + UK, 2017, based on coefficients for 2009–2017.
Source: author's calculations based on raw data from Eurostat (2019).

$$\ln(U) = -0.0310D + K \tag{6}$$

Making U the subject of Eq. (6) and differentiating, we see that the rate of change of U with D is:

$$\frac{\partial U}{\partial D} = -0.0310e^{K} \bullet e^{-0.0310D} \tag{7}$$

Substituting as above, for the three countries Italy, Lithuania and Sweden in 2017, an *increase* of €1000 is associated with a *decrease* in the percentage unable to heat their homes, of 0.471, 0.893, and 0.065 respectively. Again because of the inverted log function, the effect of a change in GDP/capita is about twice as great for a high-UH% country as for a medium-UH% country, and very small for a low-UH% country. However, the pattern is by no means as orderly as for the Gini. There is a general trend, but with six outliers, Bulgaria, Lithuania, Greece Cyprus, Portugal and Italy, which are very heavily affected by change in GDP/capita. Fig. 2 displays the modelled effects of $\frac{\partial U}{\partial D}$ for all countries in 2017, based on the coefficients for the pooled dataset.

Apart from Ln(Arrears%), the effects of the other independent variables follow the same mathematical forms as the Gini and GDP/capita. The mathematical form for Ln(Arrears%) is more straightforward, since this has the same (logarithmic) form as the independent variable (see Appendix).

To provide an overview of the mean impacts of a change in the value of each independent variable on the dependent variable Ln(U), the t-statistics for Model 1 are displayed in Table 4 alongside coefficients and standard errors.[5] The t-statistic[6] gives an indication of the relative magnitude of each independent variable's impact on the dependent variable: the larger the t-statistic (i.e. the higher its absolute value), the larger the variations in the dependent variable that are associated with variations in the independent variable. Hence Table 4 shows that the largest impacts on the percentage unable to heat their homes come from HDDs, GDP/capita and the Gini index. It is important to note, however, that governments cannot change HDDs. Further, it is very difficult for governments to increase GDP/capita, and it takes great effort and expense to reduce the percentage of inadequate homes. The Gini index, however (after tax and transfers), can be changed simply by an Act of Parliament. Fig. 3 shows the ratio of pre- and post-tax and redistribution Gini indexes in EU countries in 2017. Note that for all countries the ratio is significantly lower than 1. Sweden, for example, reduces its Gini index to less than half its pre-tax value by using progressive taxation and redistribution of income.

Table 4
Coefficients of independent variables in Model 1, with standard errors and t-statistics.
Source: author's calculations based on raw data from Eurostat (2019).

Model 1				
Ln(UH%)	Coefficient.	Std Error	t-statistic	P
Gini	0.084004	0.008805	9.54	0.000
GDP/cap	−2.62E-05	2.47E-06	−10.6	0.000
HDDs	−0.000286	2.67E-05	−10.72	0.000
Ln(Arrears)	0.239789	0.054592	4.39	0.000
Bad Homes	0.011737	0.004943	2.37	0.018

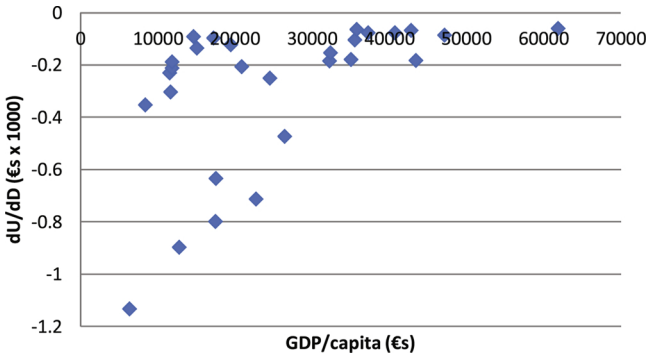

Fig. 2. Modelled effect of change in GDP/capita (€ x1000) on UH%, EU27 + UK, 2017, based on coefficients for 2009–2017.
Source: author's calculations based on raw data from Eurostat (2019).

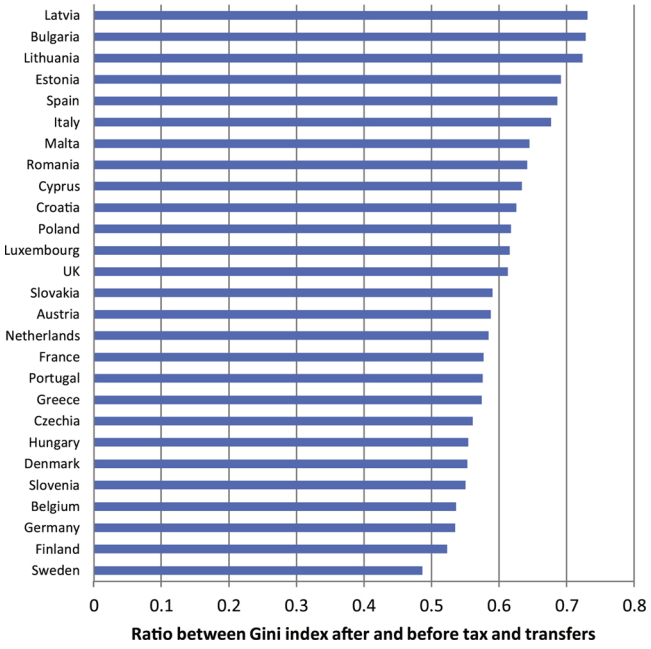

Fig. 3. Ratio between Gini index after and before tax and transfers (after/before), EU27 + UK 2017 (but no data for Ireland).
Source: author's calculations based on raw data from Eurostat (2019).

4.4. Year by year results

As a further check on the reliability of results, each year's data was regressed independently. The results are given in Tables 5 and 6. Points to note are, firstly, that the Gini index, GDP/capita and HDDs are statistically significant in all years ($p < 0.1$), while neither of the other

[5] The term "standard errors" is a measure of how spread out or scattered a variable's values are, divided by the square root of the number of observations in the sample. More formally, Standard error = standard deviation/sqrt(n).
[6] t = coefficient/standard error.

Table 5
Results of regressions for years 2017, 2016, 2015 and 2014.
Source: author's calculations based on raw data is from Eurostat (2019).

Variables	(2017) Ln(UH%)	(2016) Ln(UH%)	(2015) Ln(UH%)	(2014) Ln(UH%)
Gini	0.101***	0.100***	0.0648**	0.0656*
	(0.0204)	(0.0232)	(0.0264)	(0.0342)
GDP/cap	−1.89e-05**	−2.18e-05***	−2.22e-05**	−3.62e-05***
	(7.91e-06)	(7.08e-06)	(9.17e-06)	(1.16e-05)
HDDs	−0.000218***	−0.000214***	−0.000354***	−0.000397***
	(7.02e-05)	(7.38e-05)	(0.000103)	(0.000103)
Ln(Arrears%)	0.200	0.193	0.286	0.170
	(0.140)	(0.140)	(0.174)	(0.225)
Bad Homes	0.0147	0.0140	0.0112	0.00611
	(0.0136)	(0.0131)	(0.0166)	(0.0166)
Constant	−0.717	−0.597	0.720	1.404
	(0.756)	(0.795)	(1.005)	(1.223)
Observations	27	28	28	28
R-squared	0.851	0.853	0.816	0.795

Table 6
Results of regressions for years 2013, 2012, 2011, 2010 and 2009.
Source: author's calculations based on raw data from Eurostat (2019).

Variables	(2013) Ln(UH%)	(2012) Ln(UH%)	(2011) Ln(UH%)	(2010) Ln(UH%)	(2009) Ln(UH%)
Gini	0.0852**	0.0707**	0.0814**	0.0880***	0.0913**
	(0.0306)	(0.0310)	(0.0326)	(0.0289)	(0.0343)
GDP/cap	−3.18e-05***	−4.27e-05***	−3.34e-05***	−3.88e-05***	−4.41e-05***
	(9.88e-06)	(1.04e-05)	(1.14e-05)	(1.16e-05)	(1.36e-05)
HDDs	−0.000394***	−0.000301***	−0.000284***	−0.000261***	−0.000329***
	(9.00e-05)	(8.04e-05)	(9.47e-05)	(7.65e-05)	(0.000101)
Ln(Arrears%)	0.0328	0.140	0.260	0.161	0.315
	(0.188)	(0.190)	(0.216)	(0.217)	(0.256)
Bad Homes	0.0124	0.00911	0.0109	−0.00191	−0.00262
	(0.0147)	(0.0154)	(0.0167)	(0.0177)	(0.0223)
Constant	1.155	1.425	0.424	0.766	0.447
	(1.020)	(1.077)	(1.214)	(1.148)	(1.328)
Observations	28	28	28	28	26
R-squared	0.838	0.842	0.807	0.827	0.819

two variables is statistically significant in any year.[7] Secondly, the year-by-year coefficients of the Gini index vary, but in most years these are reasonably close to the value of 0.08 obtained for the pooled regression.

The effects of GDP/capita are interesting because its coefficient is steadily reducing over time, from larger than -4×10^{-5} in 2009 to smaller than -2×10^{-5} in 2017. It would be interesting to explore why this is. It could be because GDP/capita is increasing over time and, as shown in the above discussion, its impact on the percentage who cannot heat their homes tends to be smaller in countries where GDP/capita is larger. It could also be that a saturation point is reached, where increasing GDP/capita any further is unlikely to reduce the percentage unable to heat their homes, unless other factors such as the Gini index are reduced.

The coefficients of HDDs tend to vary from year to year with no particular pattern. This could be due to year-by-year weather variations, since the overriding impact of HDDs is via the long-term effect of building regulations. A particularly cold year may skew this relationship slightly, bringing colder weather where homes are not designed for it.

5. Discussion

The Gini index was a significant predictor of the natural logarithm of the percentage of households unable to heat their homes (Ln(UH%)). This held true in the regressions of pooled panel data, and for all individual years 2009–2017, with very low p-values, large t-statistics and large, positive regression coefficients. The size and sign of the coefficients indicated that changes in the Gini index are associated with changes in the percentage of households unable to heat their homes (UH%), and that the changes in this percentage are larger for higher initial values of both these variables: *the higher the Gini index and the percentage unable to heat their homes, the more difference it makes to this percentage if the Gini index changes*.

There are two main weaknesses in the modelling. Model 1 is somewhat simplified in that it is limited to linear and natural log representations of the data, which do not fully reflect the increases in the impact of the Gini and GDP/capita on the dependent variable as the Gini and GDP/capita increase. Model 3, which is developed in the Appendix, solves this problem, but at the expense of what would be a far more complicated process of interpretation of results. Nevertheless, this indicates that Model 1 downplays the effects of the Gini index slightly: the Gini's impact on the percentage unable to heat their homes is even more positively skewed at the high end, than Model 1 suggests.

The other weakness arises from the residual test shown in the Appendix. This indicates there is a degree of missing variable bias, suggesting there may be a further independent variable or variables which even further increase the percentage unable to heat their homes at the high end of the distribution. Again however, this does not negate the main thrust of the interpretation of coefficients for Model 1; it merely indicates there is another, unknown factor which makes the

[7] This is because their coefficients are not large enough, compared to their variances, to produce sufficiently small standard errors (see standard errors in brackets) so as to give high t-statistics and low p-values.

results less precise than the modelling equations indicate.

The results indicate a strong correlation between the Gini index and the logarithm of the percentage unable to heat their homes, controlling for other variables. As shown in Eqs. (4) and (5), this indicates that: (a) the Gini index is a strong predictor of the percentage unable to heat their homes; and (b) its impact increases as that percentage increases. The higher the percentage unable to heat their homes, the bigger the difference a change in the Gini index makes, at least in theory.

It must be emphasised, of course, that this does not prove causation, since a regression analysis can only prove correlations. It would always be possible, even in the best and most robust regression analysis, that some other factor may be influencing between-country variation in both the Gini index and Ln(UH%) in tandem.

Returning to our original hypothesis, the regressions have not been able to disprove the hypothesis that the Gini index is a predictor of the percentage of households unable to heat their homes, and in fact have bolstered it at every turn. There are also good intuitive reasons for believing this hypothesis. A high Gini index means there are a lot of very poor people, who are unlikely to have the means to heat their homes adequately or to renovate them to be more energy efficient (cf. [43]). In countries like Bulgaria, with a high Gini coefficient but high rates of owner-occupied homes, many homeowners may be stuck with thermally poor dwellings which they cannot afford to upgrade. A high Gini coefficient in a country also means there is a minority of extremely rich people in that country. In countries with modest to high rates of home ownership, real estate is a prime target of the very-rich [28]. This not only inflates its price, but also leads to a larger private rental sector, and there is abundant literature indicating that private landlords are often reluctant to upgrade their properties to high thermal standards [80].

6. Conclusions and implications for policy and research

This paper has shown there is a robust correlation between the Gini index of income inequality after tax and welfare transfers, and the percentage of households who cannot heat their homes adequately. This is indicated for the set of all 28 EU countries for all years 2009–2017. The tests carried out in this analysis controlled for the effects of other relevant independent variables, namely GDP/capita, the number of heating degree days, the percentage of households in arrears on their utility bills, and the percentage of inadequate homes. A range of tests for robustness of results showed there was no reason to doubt the clear correlation between the Gini and the percentage of households unable to heat their homes adequately.

As with all statistical studies, this one has its weaknesses. Because the Gini index is a countrywide indicator, a study such as this cannot indicate regional differences or differences of household type within countries. Further, the results indicated that a further variable or set of variables, which were not represented in the analysis, was or were also contributing to high percentages unable to heat their homes. Finally, such a study can only show correlations between variables, while other sources and methods need to be used to establish lines of causality.

Nevertheless, of all the independent variables in this study, the Gini index (after taxes and transfers) is by far the easiest for governments to influence. This can be done simply by Act of Parliament and associated administrative changes, for example by increasing marginal tax rates on the highest incomes while providing more adequate universal welfare. Referring back to Fig. 3, this chart shows, for example, that in 2017 the post-tax and welfare transfer Gini index in Sweden was 0.47 (47%) of its pre-tax and transfer level, indicating a radical redistribution of income. The country with the least radical redistribution was Latvia, at 0.73. The eight countries which redistribute income most radically (those at the bottom of the graph) include three from Eastern Europe (Slovenia, Hungary and Czechia), while the eight with the least radical redistribution (those at the top of the graph) include three from Western Europe and the Mediterranean (Spain, Italy and Malta). Hence it would be difficult to argue that (poorer) Eastern European countries find it inherently more difficult to redistribute income.

Further, tracking these ratios over time shows that governments are not only able to, but do in fact, alter their Gini coefficients by altering tax rates and welfare transfers. Denmark in 2009 had a ratio between pre- and post-tax Gini index of 0.43, but became less egalitarian by 2017 with a ratio of 0.55. Portugal became more egalitarian over that period, reducing its ratio from 0.70 in 2009 to 0.58 in 2017.

Countries need to consider, then, that having a high percentage of households unable to heat their homes may well be a symptom of poor social and economic policies, and could be significantly mitigated by adjusting these policies accordingly. Lowrey [2] has documented evidence from a number of countries that, when poor households in very unequal societies get more money, this gives them choices and empowerment to lift themselves out of deprivation. This was, of course, one of the guiding tenets of US President Franklin D Roosevelt's 'New Deal' in the 1930s, and was a mainstay of demand-side economics and welfare policies in most high-income countries in the 1940s–1970s [1,3]. With regard to energy use and climate change mitigation, it is also a basic assumption of the proposed 'Green New Deal' legislative programme being put forward for the US Congress by Representative Alexandra Ocasio-Cortez and Senator Ed Markley [81].

Of course, this does not imply that countries can fully solve the problem of households unable to heat their homes *solely* by macroeconomic and social welfare policy. Energy efficiency policy and interventions are also needed since, as discussed in Section 2, low thermal quality of homes is also a major driver of energy poverty. Reducing the Gini index would enable many households to replace inefficient boilers, insulate lofts and walls, replace the worst windows, afford to pay higher energy bills and compete on the real estate market for better homes. But energy efficiency policy interventions would also be needed, in order to address the deep, underlying issues of inefficient building stock in many European countries.

The findings of this study have important implications for social science based energy research. Firstly, most research on energy poverty is concerned with either micro-targeting households that can be expected to be in energy poverty, or exploring the consequences of energy poverty. Many of these studies note the fact that economic poverty and inequality are major drivers of energy deprivation, but they tend to treat poverty and inequality as a given, a fact of life, a black box which is left un-analysed. This study has indicated that economic inequality is closely associated with energy poverty (at least by one measure) and is not inevitable but can be changed, and changed quite simply, by legislative shifts in fiscal and welfare policy. It would be interesting to see energy poverty studies paying more attention to the economic poverty inherent in energy poverty and its causes and cures more directly.

Secondly, this study represents a small step in bridging of the decades-old gap between sociology and macroeconomics, which has been evident since the *Methodenstreit* (dispute over method) of the 1930s [13,82]. As a macroeconomic indicator, the Gini index is well discussed in current economics research. I have attempted to show here, however, that our understanding of an issue which is firmly in the domain of social science research, namely energy poverty, can be better understood by reaching across to this macroeconomic domain. Social science based energy researchers need to include macroeconomics in their toolkit, since a country's macroeconomic regime, which has profound effects on people's lives, is a decisive factor in how people can consume energy and obtain energy services.

Declaration of Competing Interest

No funding sources were used in the research or writing of this article. The work is entirely the original work of the author and has not been submitted for publication elsewhere.

Appendix A

A1 Distribution of variables

All the variables except UH% and Arrears% are more or less normally distributed. The distributions of these two variables is given in Figs. A1 and A2 respectively, showing both are strongly right-skewed.

One option for obtaining more reliable results in such cases is to regress about the median rather than the mean, using quantile regression [79]. Another is to see whether the data for the problematic variables can be transformed into a form with near-normal, near-symmetrical distributions. If so, the transformations of those variables can then be used in the same regressions as the non-transformed data of the non-problematic variables. The results of the regression must then be re-transformed for interpretation. For both UH% and Arrears%, a natural logarithmic transformation of the data produced near-normal, near-symmetrical distributions, as displayed in Figs. A3 and A4.

All the variables were then tested for skewness and kurtosis to check the extent of their deviations from normal and symmetrical. All had skewness and kurtosis close to zero (see Table A1) except that Ln(Arrears%) had skewness of 0.88 and Bad Homes % had kurtosis of 0.93. Neither of these is above the widely accepted rule-of-thumb problematic level of 1.0. Hence, using the log transformations of Ln(UH) and Arrears %, it was then possible to proceed with an OLS regression.

A further issue is the possible unevenness of the correlation of the dependent variables on Ln(UH%), across its entire distribution. Unless all the variables are perfectly normally and symmetrically distributed, it is possible that some independent variables correlate better with the dependent variable in its low value range than in its high or mid-value range, or vice versa. Quantile regression is often used to address this issue, as it enables us to regress about different quantiles, e.g. quantile 0.25 or 0.80, etc., rather than regressing about the mean, as in OLS regression. By performing a quantile regression about the median, i.e. quantile 0.50, we can compare the coefficients and standard errors with those of an OLS regression about the mean [83]. If these do not differ greatly between the two, it is reasonable to conclude that deviations from normality and symmetry are not biasing the results significantly. The results of the quantile regression (Model 2 in the main text) are close to those of the OLS regression (Model 1), indicating there are no significant problems with deviations from normality of the distributions.

However, pooled panel data (see below) presents serious challenges for quantile regression [84,85]. The results of the quantile 0.50 regression are therefore only used for comparison with those of the OLS regression and not for further analysis.

A2 Issues with pooled panel data

Our data is panel data, as it utilizes data for a number of entities (in this case 28 countries) over a number of time periods (in this case 9 years). By pooling this data we obtain a composite dataset with $28 \times 9 = 252$ observations, which records variations within countries over time, and between

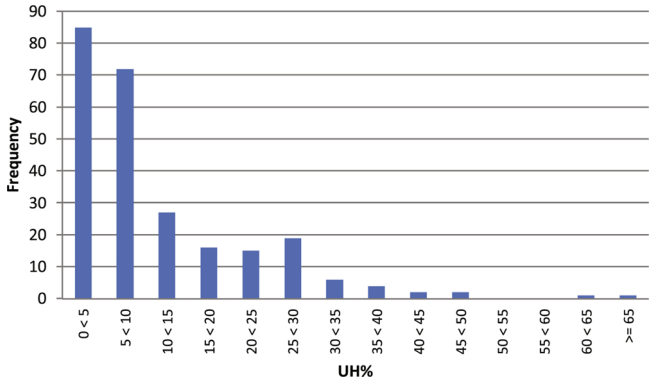

Fig. A1. Distribution of UH%, EU27 + UK, 2009–2017.
Source: author's calculations based on raw data from Eurostat (2019).

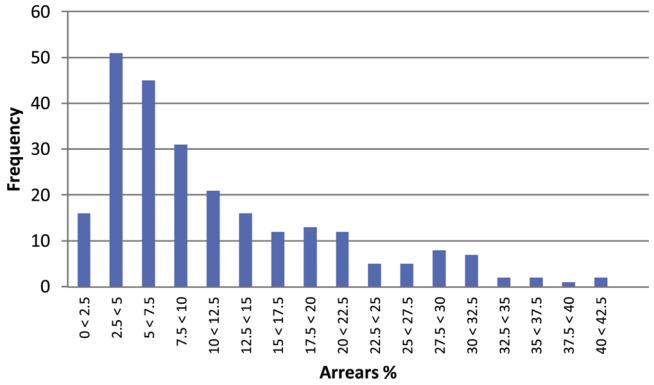

Fig. A2. Distribution of Arrears %, EU28 + UK 2009–2017.
Source: author's calculations based on raw data from Eurostat (2019).

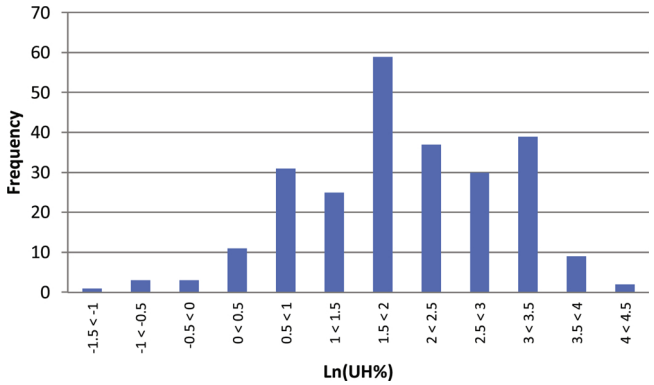

Fig. A3. Distribution of Ln(UH%), EU28 + UK 2009–2017.
Source: author's calculations based on raw data ifrom Eurostat (2019).

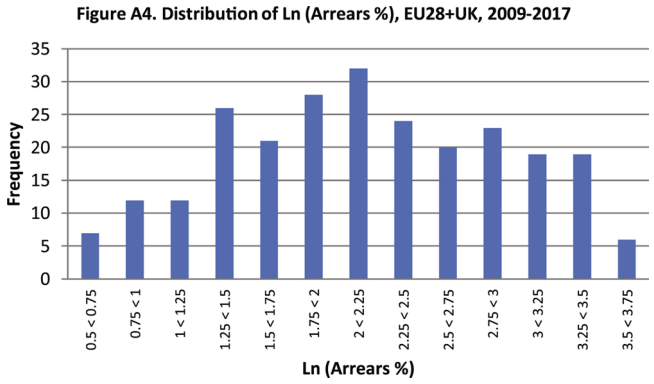

Fig. A4. Distribution of Ln (Arrears %), EU28 + UK 2009–2017.
Source: author's calculations based on raw data from Eurostat (2019).

countries in each year (in fact there were only 249 observations, since two data items for Croatia and one for Ireland were missing from the dataset). A danger in using pooled panel data is that if one or more of the variables does not vary much within one or more countries over time, we can end up effectively using the same information several times over. This will exaggerate the statistical significance of some of the results, because the divisor n (number of observations) will be larger than the number of pieces of information, thereby reducing the standard errors and p-values artificially.

The pooled dataset was therefore checked for within-country fixed and random effects. The fixed effects were non-negligible but there was still sufficient within-country variation to justify pooling the data.

A3 Further statistical tests

The OLS model, Model 1, was tested for multicollinearity, correct specification of the model, and randomness of residuals.

Multicollinearity was tested using a variance inflation factor (VIF) test, where each independent variable is regressed against all the others. The strengths of correlations between variables are given by R-squared values, which are used in the VIF formula.[8] As a rule of thumb, a VIF score under 5 is non-problematic, over 10 is highly problematic, and between 5 and 10 indicates a need for caution. The highest VIF was for Ln(Arrears%), at 2.47, followed by GDP/capita at 2.19; while the others ranged from 1.19 to 1.35. Hence there was no problem of independent variables correlating with each other such as to give misleading correlations with the dependent variable. However, the variable 'Average net household wealth' gave a VIF score of 12.1, indicating high correlation with other independent variables. It was, therefore, not used in the regressions.

A Ramsey Regression equation specification error test (RESET) was used to check whether the model was correctly specified, i.e. whether the independent variables best predict the dependent variable in the forms in which they are used, or whether there are omitted variables in the form of squares, cubes, etc., of the independent variables. The results indicated that adding further variables such as squares, cubes and square roots of GDP/capita and of Arrears %, and the square of the Gini index, produced models with a slightly better fit, and increased the adjusted R-squared value from 0.729 to as high as 0.824. However, such an approach brings complications in interpreting such models, since the transformations of variables have to be translated back into their plain form to make sense of the results. The simplest of these enhanced models (which I call Model 3) is given in equation (A1), shown here out of interest only. It was derived by regressing Ln(U) against all five independent variables and their squares and cubes, then using stepwise regression[9] to find the best fit:

[8] VIF $= 1/(1 - R_i^2)$, where R_i^2 is the correlation coefficient of a variable i against the other variables.

[9] Stepwise regression starts with a full set of independent variables (in this case including their squares, cubes, etc.) then iteratively performs a series of regressions, eliminating the worst fitting variable each time, until there remain only those variables with significance levels within the range specified, in this case $p < 0.1$.

Table A1
Skewness and kurtosis rest results for variables used in the regression.
Source: author's calculations based on raw data from Eurostat (2019).

Variable	Observations	Skewness	Kurtosis	Adjusted $chi^2(2)$	Prob > chi^2
Ln(UH%)	250	0.151	0.191	3.800	0.150
Gini	250	0.269	0.000	27.910	0.000
GDPcap	252	0.000	0.000	54.940	0.000
HDDs	252	0.253	0.220	2.830	0.243
Ln(Arrears%)	249	0.878	0.000	23.060	0.000
BadHomes	250	0.000	0.927	12.990	0.002

Table A2
Coefficients, standard errors, t-statistics and p-values for Model 3.
Source: author's calculations based on raw data from Eurostat (2019).

Model 3				
Ln(UH%)	Coefficient	Std Error	t-statistic	p
Gini^2	0.001459	0.000145	10.07	0.000
(GDP/1000)^2	−0.000524	4.53E-05	−11.58	0.000
HDDs	−0.000268	2.67E-05	−10.04	0.000
(LnArreas%)^2	0.050982	0.011468	4.45	0.000
BadHomes	0.012421	0.004917	2.53	0.012

$$\mathrm{Ln}(U) = 0.00146G^2 - 0.000534\left(\frac{D}{1000}\right)^2 - 0.00268H + 0.0510(\mathrm{Ln}(A))^2 + 0.0124I + er \quad \text{(A1)}$$

All these independent variables are significant at the 0.001 level except *I*, which is significant at the 0.05 level. However, I use the more simple form given in Model 1, as this provides for more straightforward interpretation of results, for the trade-off of slightly less precise coefficients.

The results of Model 1 were tested for randomness of residuals. The residuals are the differences between the modelled values of the dependent variable and its actual measured values in the dataset – unlike the errors, which are the differences between the modelled values and the (unknown) values in the population of which the dataset is a sample. In an ideal case, the residuals are truly random with no correlation between residuals and measured values. In Model 1, however, there was a degree of correlation (R-squared = 0.202), as the residuals increased in magnitude as Ln(*U*) increased in magnitude. Part of this was due to a small degree of misspecification of the model, since in Model 3 the correlation reduced considerably (R-squared = 0.168). No alternative specification of the model reduced it further, indicating that there was a missing independent variable (or variables) whose effects have not been taken into account. The absence of this variable produces results for Ln(*U*) that are low at the high end and a bit high at the low end. Hence, the five independent variables used here do not fully account for the extremities of the very high percentages of households unable to heat their homes. Model 1 is therefore not ideal, but is adequate for providing a credible ball-park view of correlations between Ln(*U*) and the five independent variables.

It is also important to note that the t-statistics for Models 1 and 3 are very similar. The relative impact on Ln(*U*) of each independent variable is about the same in both cases. The t-statistics for Model 3 are given in Table A2.

A4 Modelling equation for Ln(Arrears%)

The modelling equation for Ln(Arrears%) is:

$$\ln(U) = 0.216 \bullet \mathrm{Ln}(A) + K \quad (10)$$

Making *U* the subject and differentiating gives the rate of change of *U* with *A*:

$$U = 0.216e^K \bullet A^{-0.784} \quad (11)$$

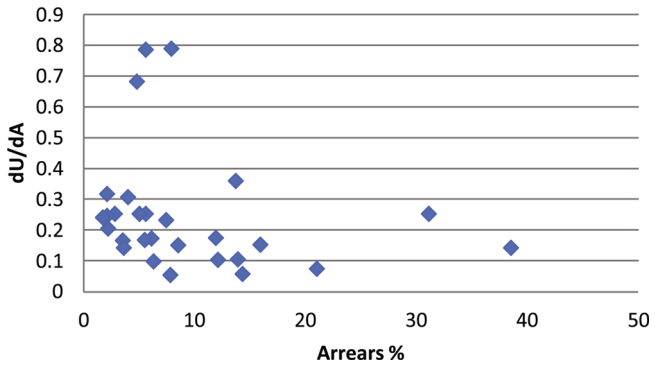

Fig. A5. Modelled effect of change in Arrears% on UH%, EU27 + UK, 2017, based on coefficients for 2009–2017.
Source: author's calculations based on raw data from Eurostat (2019).

The modelled results for this are displayed in Fig. A5. The effect on U of a change in A is very small, generally less than a 0.4 percentage point increase in U for each one percent increase in A. The exceptions are Portugal, Lithuania and Italy, which have very low A and moderate to high U.

References

[1] R. Galvin, M. Sunikka-Blank, Economic inequality and household energy consumption in high-income: countries: a challenge for social science based energy research, Ecol. Econ. 153 (2018) 78–88.
[2] A. Lowrey, Give People Money: The Simple Idea to Solve Inequality and Revolutionise Our Lives, Penguin Random House, London, 2018.
[3] T. Piketty, Capital in the Twenty-First Century (Translated From the French by Arthur Goldhammer), Belknapp-Harvard University Press, Cambridge (Mass), 2014.
[4] K. Raworth, Doughnut Economics: Seven Ways to Think Like a 21st-Century Economist, Penguin Random House, London, 2017.
[5] J. Stiglitz, The Price of Inequality, Penguin, London, 2013.
[6] J. Winters, Wealth defense and the limits of liberal democracy, Paper for Annual Conference of the American Society of Political and Legal Philosophy, Washington, D.C., 2014 August 28–31, 2014. Revised 29 April 2015.
[7] G. Zucman, Taxing across borders: tracking personal wealth and corporate profits, J. Econ. Perspect. 28 (4) (2014) 121–148.
[8] A. Atkinson, Top Incomes in the United Kingdom over the Twentieth Century, (2003) http://www.nuffield.ox.ac.uk/users/atkinson/TopIncomes2003.pdf.
[9] T. Piketty, E. Saez, Income inequality in the United States, 1913–1998, Q. J. Econ. 118 (1) (2003) 1–41.
[10] A. Atkinson, T. Piketty (Eds.), Top Incomes over the Twentieth Century: A Contrast between Continental European and English-Speaking Countries, Oxford University Press, Oxford and New York, 2007.
[11] M. Savage, Social Class in the 20th Century, Pelican, London, 2015.
[12] D. Dorling, Do We Need Economic Inequality? Polity Press, Cambridge, 2018.
[13] G. Ingham, On the underdevelopment of the 'Sociology of Money', Acta Sociol. 41 (1998) 3–18.
[14] T. Schatzki, Social Practices: A Wittgensteinian Approach to Human Activity and the Social, Cambridge University Press, New York, 1996.
[15] T. Schatzki, Practices and actions: a Wittgensteinian critique of Bourdieu and Giddens, Philos. Soc. Sci. 27 (3) (1997) 293–308.
[16] A. Reckwitz, Toward a theory of social practices: a development in culturalist theorizing, Eur. J. Soc. Theory 5 (2002) 245–265.
[17] E. Shove, Beyond the ABC: climate change policy and theories of social change, Environ. Plan. A 42 (6) (2010) 1273–1285.
[18] M. Foucault, C. Gordon (Ed.), Power/Knowledge: Selected Interviews & Other Writings 1972–1977, Pantheon Books, New York, 1980.
[19] Wodak, Meyer (Eds.), Methods of Critical Discourse Analysis, Sage, London, 2001.
[20] Ü Berger, T. Luckmann, The Social Construction of Reality: A Treatise in the Sociology of Knowledge, Anchor Books, New York, 1966.
[21] W. Bijker, T. Hughes, T. Pinch (Eds.), The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology, MIT Press, Cambridge (MASS) and London (UK), 1987.
[22] H. Lovell, Exploring the role of materials in policy change: innovation in low-energy housing in the UK, Environ. Plann. A (39) (2007) 2500–2517.
[23] B. Latour, We Have Never Been Modern. Translated by Catherine Porter, Harvard University Press, Cambridge (MAS), 1993.
[24] A. Jorgenson, J. Schor, X. Huang, J. Fitzgerald, Income inequality and residential carbon emissions in the United States: a preliminary analysis, Hum. Ecol. Rev. 22 (1) (2015) 93–105.
[25] A. Jorgenson, J. Schor, K. Knight, X. Huang, Domestic inequality and carbon emissions in comparative perspective, Sociol. Forum 31 (S1) (2016) 770–786.
[26] K. Knight, J. Schor, A. Jorgenson, Wealth inequality and carbon emissions in high income countries, Soc. Curr. (2017) 1–10.
[27] L. Chancel, T. Piketty, Carbon and Inequality: From Kyoto to Paris. Trends in the Global Inequality of Carbon Emissions (1998–2013) & Prospects for an Equitable Adaptation Fund, Paris School of Economics, 2015.
[28] R. Fernandez, R. Hofman, M. Aalbers, London and New York as a safe deposit box for the transnational wealth elite, Environ. Plan. A (2016) 1–19 (in press).
[29] S. Bouyon, Recent Trends in EU Home Ownership, European Credit Research Institute, ECRI Commentary, 2015.
[30] A. Hay (Ed.), The Wealth Report: The Global Perspective on Prime Property and Investment, 10th edition, Knight-Frank, London, 2016.
[31] Hometrack, Our Housing Market, December 2015. Housing Market Bulletin, edition 28, (2016) Cambridge Region.
[32] L. Schembri, The Long-Term Evolution of House Prices: An International Perspective. Remarks by Lawrence Schembri, Deputy Governor of the Bank of Canada, Canadian Association for Business Economics Kingston, Ontario, 25 August 2015, Bank of Canada, 2015.
[33] S. Clarke, A. Corlett, L. Judge, The Housing Headwind: The Impact of Rising Housing Costs on UK Living Standard, Resolution Foundation, London, 2016.
[34] A. Clarke, S. Morris, P. Williams, How Do Landlords Address Poverty? Joseph Rowntree Foundation, 2015, https://www.jrf.org.uk/report/how-do-landlords-address-poverty.
[35] L. Maruejols, D. Young, Split incentives and energy efficiency in Canadian multi-family dwellings, Energy Policy 39 (2011) 3655–3668.
[36] G. Wood, R. Ong, C. McMurray, Housing tenure, energy consumption and the split-incentive issue in Australia, Int. J. Hous. Policy 12 (4) (2012) 439–469.
[37] K. Gillingham, M. Harding, D. Rapson, Split incentives in residential energy consumption, Energy J. 33 (2) (2012) 37–62.
[38] A. Hope, A. Booth, Attitudes and behaviours of private sector landlords towards the energy efficiency of tenanted homes, Energy Policy 75 (2014) 369–378.
[39] F. Botti, M. Corsi, C. D'Ippoliti, The gendered nature of multidimensional poverty in the European Union, CEB Working Paper N° 12/026, Centre Emile Bernheim Research Institute in Management Studies, Brussels, 2012.
[40] Resolution Foundation, Home Ownership in the UK, Available at (2017) https://www.resolutionfoundation.org/data/housing/.
[41] B. Tunstall, Housing and Gender, WGB, London, 2017.
[42] L. Middlemiss, A critical analysis of the new politics of fuel poverty in England, Crit. Soc. Policy 37 (3) (2017) 425–443.
[43] BRISKEE, Behavioural Response to Investment Risks in Energy Efficiency: A Horizon2020 Research Project, (2018) http://www.briskee-cheetah.eu/briskee/.
[44] R. Castaño-Rosa, J. Solís-Guzmán, C. Rubio-Bellido, M. Marrero, Towards a multiple-indicator approach to energy poverty in the European Union: a review, Energy Build. 193 (2019) 36–48.
[45] U. Dubois, H. Meier, Energy affordability and energy inequality in Europe: implications for policymaking, Energy Res. Soc. Sci. 18 (2016) 21–35.
[46] S. Moore, Definitions of fuel poverty: implications for policy, Energy Policy 49 (2012) 19–26.
[47] H. Thomson, C. Snell, C. Liddell, Fuel Poverty in the European Union: A Concept in Need of Definition? People, Place & Policy Online, University of York, 2016, pp. 5–24, https://doi.org/10.3351/ppp.0010.0001.0002.
[48] G. Walker, R. Day, Fuel poverty as injustice: integrating distribution, recognition and procedure in the struggle for affordable warmth, Energy Policy 49 (2012) 69–75.
[49] EC, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions—A Framework Strategy for a Resilient Energy Union with a Forward-looking Climate Change Policy: COM(2015) 80 Final, Brussels, 2015. EU-SILC Survey (2017) European Union Statistics on Income and Living Conditions, Available free online: (2015) http://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions.
[50] S. Pye, A. Dobbins, C. Baffert, J. Brajković, P. Deane, R. De Miglio, Energy poverty across the EU: analysis of policies and measures, Europe's Energy Transition - Insights for Policy Making: Findings Informing the European Commission, (2017), pp. 261–280.
[51] S. Bouzarovski, Energy poverty in the European Union: landscapes of vulnerability, WIREs Energy Environ. 3 (2014) 276–289.
[52] H. Thomson, S. Bouzarovski, C. Snell, Rethinking the measurement of energy poverty in Europe: a critical analysis of indicators and data, Indoor Built Environ. 26 (7) (2017) 879–901.
[53] S. Bouzarovski, S. Herrero, Geographies of injustice: the socio-spatial determinants of energy poverty in Poland, the Czech Republic and Hungary, Post- Communist Econ. 29 (1) (2017) 27–50.
[54] C. Price, K. Brazier, W. Wand, Objective and subjective measures of fuel poverty, Energy Policy 49 (2012) 33–39.
[55] J. Hills, Fuel Poverty: The Problem and Its Measurement. Interim Report of the Fuel Poverty Review, Centre for Analysis of Social Exclusion, LSE, London, 2011.
[56] J. Hills, Getting the Measure of Fuel Poverty. Final Report of the Fuel Poverty Review, Centre for Analysis of Social Exclusion, LSE, London, 2012.
[57] Mirasgedis S. Atsalis, C. Tourkolias, D. Diakoulaki, Fuel poverty in Greece: quantitative analysis and implications for policy, Energy Build. 131 (2016) 87–98.
[58] J. Romero, P. Linares, X. López, The policy implications of energy policy indicators, Energy Policy 115 (2018) 98–108.
[59] C. Liddell, S. Langdon, Tackling Fuel Poverty in Northern Ireland-An Area-Based Approach to Finding Households Most in Need, (2013) http://www.ofmdfmni.gov.uk/de/tackling-fuel-poverty-in-ni-liddell-lagdon.pdf.
[60] G. Palmer, T. MacInnes, P. Kenway, Cold and Poor: An Analysis of the Link Between Fuel Poverty and Low Income, New Policy Institute, London, 2008.
[61] Eurostat, Your Key to European Statistics, (2019) http://ec.europa.eu/eurostat.
[62] B. Boardman, Participant benefits and quality of life: the challenge of hard to measure benefits, IEA Workshop Evaluating the Co-Benefits of Low-Income Weatherisation Programmes, Dublin, 27–28 January 2011, 2011.
[63] H. Chappells, E. Shove, An Annotated Bibliography of Comfort Research, Available online: (2003) http://www.lancaster.ac.uk/fass/projects/futcom/comfort_biblio.pdf.
[64] J. Cupples, V. Guyatt, J. Pearce, Put on a Jacket, You Wuss": cultural identities, home heating, and air pollution in Christchurch, New Zealand, Environ. Plann. A: Econ. Space 39 (12) (2007) 2883–2898.
[65] L. Papada, D. Kaliampakos, Measuring energy poverty in Greece, Energy Policy 94 (2016) 157–165.
[66] S. Bouzarovski, N. Simcock, Spatializing energy justice, Energy Policy 107 (2017) 640–648.
[67] R. Day, G. Walker, H. Simcock, Conceptualising energy use and energy poverty using a capabilities framework, Energy Policy 93 (2016) 255–264.
[68] B. Sovacool, Fuel poverty, affordability, and energy justice in England: policy insights from the warm front program, Energy 93 (2015) 361–371.
[69] N. Katsoulakos, Combating energy poverty in mountainous areas through energy-saving interventions, Res. Dev. 31 (4) (2011) 284–292.

[70] B. Lloyd, Fuel poverty in New Zealand, Social Policy J. N. Z. 27 (2006) 142–155.
[71] R. Briggs, R. Lucas, G. Taylor, Z. Todd, Climate classification for building energy codes and standards: part 1-development process, ASHRAE Trans. 109 (2003) 4610–4619.
[72] J. Aron, J. Duca, J. Muellbauer, K. Murata, A. Murphy, Credit, housing collateral and consumption: evidence from the UK, Japan and the US, Discussion Paper Series, No. 487, University of Oxford, Department of Economics, 2010.
[73] M. Lazzarato, The Making of the Indebted Man: An Essay on the Neoliberal Condition, MIT Press, Cambridge MAS, 2012.
[74] M. Lazzarato, Governing by Debt, Semiotext(e), South Pasadena, CA, 2015.
[75] S. Soederberg, Debtfare States and the Poverty Industry: Money, Discipline and the Surplus Population, Routledge, London, 2014.
[76] M. Sunnika-Blank, R. Galvin, Introducing the prebound effect: the gap between performance and actual energy consumption, Build. Res. Inf. 40 (3) (2012) 260–273.
[77] H. Schweiger, B. Schlammadinger, The potential of fuelwood to reduce greenhouse gas emissions in Europe, Biomass Bioenergy 15 (4–5) (1998) 369–377.
[78] D. Styles, M. Jones, Current and future financial competitiveness of electricity and heat from energy crops: a case study from Ireland, Energy Policy 35 (8) (2007) 4355–4367.
[79] N. Kaza, Understanding the spectrum of residential energy consumption: a quantile regression approach, Energy Policy 38 (11) (2010) 6574–6585.
[80] B. Ástmarsson, P. Jensen, E. Maslesa, Sustainable renovation of residential buildings and the landlord/tenant dilemma, Energy Policy 63 (2013) 355–363.
[81] Congress, RESOLUTION recognizing the duty of the Federal government to create a Green New Deal. 116TH CONGRESS, 1ST SESSION H. RES. (2019).
[82] M. Mizruchi, L. Sterns, Money, Banking and financial markets, in: N. Smelser, R. Swedberg (Eds.), The Handbook of Economic Sociology, Princeton University Press, Princeton, 1995, pp. 313–341.
[83] R. Koenker, G.J. Bassett, Regression quantiles, Econometrica 46 (1) (1978) 33–50.
[84] C. Hsaio, Analsyis of Panel Data, second edition, Cambridge University Press, Cambridge, 2003.
[85] M. Harding, C. Lamarche, A quantile regression approach for estimating panel data models using instrumental variables, Econ. Lett. 104 (2009) 133–135.