Distributional and climate implications of policy responses to energy price shocks

Thiemo Fetzer Ludovica Gazze Menna Bishop *

first version: February 27, 2023
this version: July 12, 2023

Abstract

Which households are most affected by energy price shocks? What can we learn about the distributional implications of carbon taxes? How do interventions in energy markets affect these patterns? This paper introduces a measurement framework that leverages granular property-level data representing more than 50% of the English and Welsh housing stock. We use this ex-ante measurement framework to investigate these questions and set out an empirical evaluation framework to study the causal effects of the energy crisis more broadly. We find that the energy price shock has a more pronounced effect on relatively more affluent areas highlighting the likely progressive impact of carbon taxation. We document that commonly used untargeted interventions in energy markets significantly weaken market price signals for able-to-pay households. Alternative, more targeted policies are cheaper, easily implementable, and could better align energy saving incentives.

Keywords: Energy crisis, Carbon taxation, Climate change, Energy efficiency gap
JEL Classification: Q48; C55

---

*All three authors are based at the University of Warwick. We would like to thank Arun Advani, Jane Snape, Stephanie Seavers, Sheila Kiggins, Ben Moll, Carlo Perroni, Ben Lockwood, Mirko Draca, Jakob Schneebacher, Pedro CL Souza, Lily Shevchenko, Carlo Schwarz, Mike Waterson, Lucie Gadenne, and many others for helpful comments. Reuben Collins, Leon Beaufils, Natalia Tronina, Suraj Rajesh, and Tymoteusz Gozdzik provided excellent research assistance. We acknowledge funding from the 2022 Leverhulme Trust Prize in Economics and the Department of Economics and CAGE research centre at the University of Warwick.

1 Introduction

Carbon taxes, which are key to curbing climate change, will increase the price of energy derived from fossil fuels. But the distributional consequences of carbon taxation are still unclear. Propelled by the post-pandemic economic recovery first, and Russia's invasion of Ukraine later (Ari et al., 2022), wholesale energy prices in Europe more than tripled in the first quarter of 2022 relative to the first quarter of 2021. Thus, this energy crisis provides a unique window into studying the economic effects of higher energy prices, to mimic similar increases induced by carbon taxation. In the absence of government intervention, shocks to wholesale energy prices pass through to household energy bills, meaning price hikes can create significant welfare costs. Higher energy prices likely have heterogeneous effects across socioeconomic groups and geographies, due, at least in part, to the energy efficiency and the fuel mix of the residential building stock, as well as to households' ability to invest in energy efficiency to shield themselves from price shocks. Thus, wholesale price shocks have the potential to create differential incentives to invest in insulation and energy efficiency measures (Houde and Myers, 2021). As such, the current crisis can inform the debate on potential implications of a carbon tax in the residential building sector, which accounts for 40% of energy consumption and 36% of energy-related greenhouse gas emissions in Europe (European Union, 2021).

The United Kingdom represents an interesting context in which to study the distributional consequences of the energy crisis and the policies deployed to counter it for three reasons. First, for residential use, the UK relies disproportionally on natural gas (63%, second only to the Netherlands in Europe), while renewables and biofuels play a limited role (Figure 1, Panel A). Partially due to this reliance on natural gas, UK energy prices were projected to grow even more than in other European countries, by over 600% between 2021 and 2023 (IEA, 2022). Second, the UK is among the lowest ranking European countries in terms of the energy efficiency of its residential building stock across several measures. For example, UK homes lose heat faster than in most other European countries, according to data from the company tado°.[1] Moreover, its fuel poverty rates are among the highest in Europe, standing at 12% in Wales, 13% in England, 18% in Northern Ireland and

[1]Source: `https://www.tado.com/t/en/uk-homes-losing-heat-up-to-three-times-faster-than-european-neighbours/` accessed on January 30, 2023.

25% in Scotland (Hinson and Bolton, 2022; Guertler et al., 2015). Third, the UK – just as many other countries around the globe – has a wealth of underutilized data that enable a distributional analysis of different energy pricing policies. As such, quantifying the impact of the current energy crisis on UK households can shed light on some of the worst-case scenarios that other countries and regions might face in the transition to the decarbonisation of the residential building sector.

For this purpose, we develop a measure of energy consumption for properties in England and Wales that allows us to assess the distributional impacts of different energy price scenarios or policy interventions. We harness the Energy Performance Certificate (EPC) database, which includes over 22 million certificates detailing estimates of energy expenditure along with other granular public data on energy consumption. The underlying set of unique properties – around 15 million – represents a large share – over 50% – of the English and Welsh residential building stock. Each EPC includes model-based energy consumption estimates for space heating, hot water generation, and electrical light consumption based on the physical characteristics of a particular building, a thermodynamic modelling approach, and *assumptions on occupancy*. We anchor the derived energy consumption measure with anonymized individual-level meter reading data along with granular spatial energy consumption data. This moment-matching rescaling approach allows us to capture local demographics and socio-economic characteristics that may affect energy consumption *over and above* what the model-based consumption figures predict.[2]

We use these energy consumption estimates to project energy bills under different price policies. For example, we model how energy bills change for different households as a result of changes in the UK's uniform energy price cap.[3] Ours is a counterfactual quantification exercise which abstracts away from variation in energy expenditure driven by differences in behaviour across households. We carry out two sets of interconnected descriptive analyses at the Middle Layer Super Output Area level, of which there are 6,791 in England, to enable matching with socioeconomic

[2]This approach recognises that energy consumption in a given area is a product of the area's housing stock and the behavior of the area's residents. See Lyubich (2022) for a discussion of the role of people vs. places in determining carbon emissions.

[3]The energy cap sets the maximum price that energy suppliers are allowed to charge customers, and is chosen by regulator Ofgem.

data.

First, we characterise which areas are, on average, more exposed to the energy price shock using a best-subset selection approach. In absolute terms, more affluent regions tend to be more exposed to the shock, likely because well-off households tend to live in bigger, older, and more energy inefficient properties. Consistent with this hypothesis, we find that MSOAs with higher shares of underoccupied homes, and higher shares of elderly residents are more exposed to the energy price shock. We speculate that low energy prices in the past few decades translated into low projected annual savings from energy efficiency investments, lowering adoption of these investments for able-to-pay households (Adam et al., 2022).

Second, we evaluate the energy price shock under different pricing policies: we consider both a highly subsidized uniform unit-price cap, which is the policy implemented by the UK government in October 2022 under a scheme known as the Energy Price Guarantee (EPG), as well as an alternative two-tier energy tariff, where an initial quantity of energy consumption is charged at a subsidised rate but consumption beyond this threshold is priced at market rates. Notably, 86% of the UK government budget earmarked to aid households weather the energy crisis involved untargeted subsidies (Figure 1, Panel B). Because the energy price shock is more pronounced in affluent areas, the current uniform price cap disproportionally benefits these areas, with regressive implications considering the difference between the cap and wholesale prices has to be funded, as highlighted in Fetzer (2022). We also consider prices absent intervention during the energy crisis, that is at a higher uniform cap intended to allow moderate profits for energy suppliers. We refer to this as a market price scenario. Importantly, our methodology accounts for variation in heating system and energy efficiency performance at the property level, as well as area-specific demographic characteristics. Because untargeted subsidies lower the benefits of energy efficiency investments in terms of savings under higher prices, we speculate that these policies reduce investments among able-to-pay households, those that otherwise would have been most exposed to the energy price shock. Figure 1, Panel B shows that most European countries used primarily untargeted subsidies to aid households during the energy crisis, suggesting that those countries might likewise see decreases in energy efficiency investments among able-to-pay households.

Perhaps most importantly, this paper introduces an *ex-ante impact tracking and measurement framework* to evaluate the wide-reaching causal impacts of the energy crisis on society. As such, it may serve as an example of empirical evaluations that can enable evidence-based policy making as a lived practice.[4] Most policy evaluation frameworks rely on inducing experimental variation, are planned well ahead of time, and typically focus on small scale pilot interventions. Yet, in many scenarios, these robust evaluation frameworks are not feasible for logistical, political, or ethical reasons. The pandemic has given rise to many of such natural experiments that may or may not have been carefully evaluated alongside (see Fetzer, 2021a,b; Fetzer and Graeber, 2021 for some examples). As a result, economists typically evaluate policy responses *ex-post* giving rise to a broad literature documenting the unintended consequences of policy choices. These studies, however, may be subject to the vagaries and incentive problems around the research publication process (Brodeur et al., 2020; Blanco-Perez and Brodeur, 2021; Brodeur et al., 2016). By contrast, our measurement framework provides an *ex-ante* shock intensity measure that can be leveraged empirically in a broad range of difference-in-difference style estimation frameworks to study the causal impact of the energy price shocks on a range of socio-economic outcomes. Further, it can be used to comment on the role that policy choices had in shaping the impacts of the shock. For example, it will be used as an input to study the impact and the mechanisms through which the energy crisis and its policy response affected crime, health, deprivation, investment, adaptation, and financial stability outcomes more broadly.[5] Fetzer (2023a) provides an illustration of one such research paper documenting the impact that the energy crisis had on crime.

This work contributes to several strands of the literature. First, it furthers our understanding of how interventions in energy markets affect the distributional impact of the energy crisis (see e.g. Harari et al., 2022; Bhattacharjee et al., 2022; Bachmann et al., 2022; Fetzer, 2022; Ruhnau et al., 2022).[6] Our results that wealthier

[4]For the wider public, the underlying data was made available as an interactive lookup via the *Financial Times* via `https://ig.ft.com/uk-energy-efficiency-gap/`.

[5]A public register of data and some first output is presented on `https://osf.io/vhnjz/`. Subsequent output will be made available on `http://www.trfetzer.com/climate-crisis-research/`.

[6]Importantly, we do not discuss efficiency implications of alternative policies. See for example Levinson and Silva (2022) for a discussion of efficiency-equity tradeoffs in the United States.

areas, where energy consumption is highest, are hit the most by the energy price shock *in absolute terms* and therefore benefit the most from untargeted subsidies funded through general taxes complements existing literature that emphasizes how disadvantaged households are less able to weather energy price shocks in general (Cong et al., 2022; Doremus et al., 2022). A key unknown is the extent to which households can adjust their energy consumption. Labandeira et al. (2017) carry out a meta-analysis, finding a short-term elasticity of -0.21 and a long term elasticity of -0.61, with additional heterogeneity by fuel type.

Second, we contribute to research investigating the existence of an energy efficiency gap, its determinants, and implications for the targeting of policies to increase take-up of energy efficiency investments (Allcott and Greenstone, 2012; Gerarden et al., 2017; Christensen et al., 2021). Regulatory barriers and lack of awareness of own energy use and of the returns of different investments appear to play a role (see e.g. Fetzer, 2023b; Attari et al., 2010), although homebuyers appear to be attentive to changes in fuel prices Myers, 2019. Moreover, there is mixed evidence on the existence of a 'green premium', i.e. the capitalization of energy efficiency into higher prices and rents (Dalton and Fuerst, 2018; Myers, 2020; Myers et al., 2022; Guin et al., 2022; Ghosh et al., 2022; Cassidy, 2023), with evidence showing that some properties become "stranded assets" following the introduction of minimum efficiency standards Ferentinos et al., 2023. In addition, socioeconomic factors appear strongly correlated to energy efficiency Zhang et al., 2012; Ahlrichs et al., 2022; Gregório and Seixas, 2017. Our work shows that different energy pricing policies affect incentives to invest in energy efficiency for households with different socioeconomic characteristics.

In the following section, we describe how we arrive at a measure of the vulnerability to energy price shocks in England and Wales.

2 Measuring exposure to energy price shocks

To model the likely exposure of a household to the energy price shock, we need to gain an understanding of baseline energy consumption. Energy consumption of household i in property p is driven by at least three factors:

$$E_{i,p} = f(\text{What}_p, \text{Who}_{i,p}, \text{How}_{i,p})$$

The What_p captures the type of property or building in which energy is consumed. The predominant sources of domestic energy use are space heating, hot water generation, room lighting, and appliances. Certain properties, all else equal, consume more energy across these uses because of their physical characteristics, such as insulation levels. The second factor, $\text{Who}_{i,p}$, captures residents' characteristics, for example household size and composition which may imply different levels of energy demand. The third factor, $\text{How}_{i,p}$, represents people's preferences. For example, people have different perceptions as to what constitutes a comfortable indoor temperature. Importantly, these factors may interact nonlinearly: energy demand may be structurally higher in a poorly insulated property, but even more so if its residents prefer a relatively high indoor temperature.

We develop our vulnerability measure starting from an estimate of energy consumption based on the What_p, i.e. the underlying characteristics of a property. We augment this exogenous measure with anonymised data on actual energy consumption at the individual property level, along with energy consumption aggregates at spatially granular levels using a moment-matching approach. In doing so, we are also able to incorporate the $\text{Who}_{i,p}$ and the $\text{How}_{i,p}$ into our measure of energy consumption, i.e. the patterns of energy consumption behaviour that exist in reality across households. This rescaling ensures that we are more likely to achieve a good simulated actual exposure measure to the energy price shock. The data generation sequence is visually described in Appendix Figure A1.

In the next sections, we describe the underlying data and the generation of energy consumption estimates.

2.1 Deriving proxy measures for energy consumption

The first step in our data construction involves deriving energy consumption measures from energy performance certificate (EPC) data. EPCs provide buyers and tenants with information on the energy efficiency rating of residential properties as well as estimates of likely energy costs. EPCs also contain recommendations of measures to improve the properties' energy efficiency, including estimates of the costs and impact of these measures on energy consumption. These recommended improvements are tailored to each property, including whether it has double glazing and which type of insulation its walls permit.

The requirement for properties to have an EPC was introduced in 2007 following the EU Directive on the energy performance of buildings (Department for Levelling Up, Housing & Communities, 2017). This requirement was initially applied just to homes for sale, but has since been extended to all domestic and commercial properties being sold, constructed, or rented (Department for Levelling Up, Housing & Communities, 2021). EPCs for domestic and commercial buildings are available to download online from the national database of all registered EPCs.[7] In total, the database includes 22,179,913 current certificates for 15,621,668 unique properties across England and Wales. While we derive energy consumption measures for all certificates and the underlying properties, we focus in most exercises on slightly smaller subsets of the data that include only properties that use electricity and/or gas for space-heating and hot water generation. This amounts to 13,462,394 properties or around 51% of the English and Welsh residential building stock, as council tax data estimates the total number of residential properties at 26,328,530.

A limitation of the EPC data is that certificates are valid for 10 years, meaning properties may have undergone changes, for example via the addition of an extension or insulation, that are not reflected in their most recent certificate. A second potential concern is that the EPC data may not be representative of the entire building stock, because buildings without EPCs might differ from those with EPCs. A comparison by the ONS of the EPC data vis-a-vis the population of properties from the Valuation Office Agency (VOA) data, built for council tax purposes, suggests that the properties are very similar on observables.[8][9] In terms of the potential energy savings, there are good reasons to believe that the properties that do not have an EPC rating may have, on average an even worse energy efficiency.[10]

[7]Data are available here `https://epc.opendatacommunities.org/`.

[8]See Office of National Statistics, Energy efficiency of housing in England and Wales: 2021, `https://www.ons.gov.uk/peoplepopulationandcommunity/housing/articles/energyefficiencyofhousinginenglandandwales/2021`.

[9]Still, Department for Business, Energy & Industrial Strategy (2020) suggests that that the EPC database under-represents medium-sized properties and bungalows and over-represents smaller properties and flats.

[10]This assertion is based on data suggesting that known energy efficiency measures produce larger energy savings among properties without an EPC certificate. See BEIS National Energy Efficiency

The estimates of annual energy costs included in the public EPC data are expressed in terms of GBP and not in energy units (kWh). They are provided separately for space heating, water heating, and lighting. The Standard Assessment Procedure (SAP) sets out the methodology used to produce these estimates (BRE, 2014). We combine these estimates with price data to back out estimated energy consumption in kWh for space heating, water heating, and lighting, reverse-engineering the SAP calculations.[11]

This process yields a vector of energy consumption proxies measured in kWh for each property p. We detail the technical approach in Appendix A and refine these measures further in the next section.

$$E_p^{EPC} = \{S_p, W_p, V_p\}$$

The three main energy use functions that are modelled are for space heating (S), hot water generation (W), and electricity use for lighting (V). These breakdowns allow us to model energy bills as a function of the fuel used for each energy use type. Note that these forms of energy use exclude the running of appliances like TVs, computers, cookers, washing machines, or dishwashers. The predominant driver of combined modelled energy consumption is space-heating.

A natural mismatch between energy consumption across properties and the EPC-derived measures of consumption can arise because properties are not inhabited by the number of people that are assumed in the model used to produce the EPC data. We next describe how we refine and rescale the E_p^{EPC} measure to match with other observed data on energy consumption.

2.2 Percentile-based rescaling

We refine the EPC-derived measures using two percentile matching-based rescaling approaches. We leverage two sources of energy consumption data derived from

Data-Framework (NEED): impact of measures data tables 2021, https://www.gov.uk/government/statistics/national-energy-efficiency-data-framework-need-impact-of-measures-data-tables-2021.

[11]The Department for Business, Energy and Industrial Strategy (BEIS) publishes data on average gas and electricity prices for 2010-2021. Data are available here https://www.gov.uk/government/statistical-data-sets/annual-domestic-energy-price-statistics.

meter readings. By doing so, we are able to anchor E_p^{EPC} in data reflecting *who* lives in property p and *how* they live.

Anonomized individual property level consumption data. The first approach leverages anonymized energy data collected through the UK's National Energy Efficiency Data Framework (NEED). This dataset includes gas and electricity meter reading data for 4 million properties. The sample is designed to be representative of domestic properties in England and Wales.[12] The NEED data also include a range of property and area-level characteristics, such as property age and region, which can also be found in the EPC data, allowing for matching.

We rescale consumption estimates in a given percentile of EPC-derived energy consumption using the consumption estimates for properties in the same percentile of NEED-derived energy consumption. We do this separately for properties with different characteristics. While this first rescaling allows us to account for variation in consumption driven by property characteristics, it may still exclude variation driven by local demographics, as the NEED data do not include granular geographic identifiers. To incorporate local variation, we employ a second rescaling method.

Local area consumption data. BEIS publishes energy consumption data down to the postcode level, excluding only postcodes that include fewer than five readings. The data include both mean and median consumption for electricity and gas.[13] We repeat the moment-matching approach described above, rescaling both the EPC and EPC-NEED augmented measures using the mean and median energy consumption values that correspond to a property's postcode.

Appendix B describes this process in more detail, which yields four measures proxying actual energy consumption at the individual property p level:

$$\mathbf{E_p} = (E_p^{EPC}, E_p^{NEED}, E_p^{Local}, E_p^{EPC,NEED,Local})$$

[12]The data are available on `https://www.gov.uk/government/statistics/national-energy-efficiency-data-framework-need-anonymised-data-2021`.

[13]The data are available for electricity at `https://www.gov.uk/government/collections/sub-national-electricity-consumption-data` and for natural gas at `https://www.gov.uk/government/collections/sub-national-gas-consumption-data`.

We break down each measure into space heating, water heating, and lighting. For most of the analysis, we leverage a simple *ensemble* average measure, $E_p^{ensemble}$, which is the unweighted average of the four measures. In Figure 2 we plot the *ensemble* EPC-derived median energy consumption measure $E_p^{ensemble}$ against MSOA-level medians. While the fit is very good, our EPC-derived estimates systematically underestimate total energy consumption. This underestimation can be explained by the fact that the EPC data covers only around 50-60% of properties.[14]

We next describe how we use the energy consumption estimates to arrive at estimates of energy bills under different policy and price scenarios.

2.3 Estimating energy bills

With the above vectors of energy consumption proxies broken down by respective energy use functions, along with information on which fuels are used to heat properties and the appropriate energy tariff, we can derive estimates of household energy bills under the following price and policy scenarios.

1. Market price. In January 2019, the UK regulator, Ofgem, adopted a uniform energy cap, that is a maximum price that energy suppliers are allowed to charge customers for gas and electricity. This cap reflects the costs of supplying energy and allows modest profits (Ofgem, 2022a). The cap has been updated every 6 months until October 2022, when it started to be updated on a quarterly basis. The price cap was originally conceived to protect inattentive consumers from being charged unfair rates. In its early years, some energy contracts on the market were cheaper than the cap, but since the summer of 2021, the cap has been the cheapest rate available. This phenomenon is due to price increases between the time at which the price cap is set and the time at which it comes into effect (Ofgem, 2022b).[15] As such, the cap has been a more accurate reflection of the prices faced by households in 2022 than in previous years. Our study incorporates price cap values from October 2021 and

[14] Appendix Section E provides further validation for our methodology. Appendix Figure A3 plots the R^2 of a set of regressions using MSOA-level energy consumption data: mean, median and total energy consumption against corresponding moments from our *ensemble* consumption measure $E_p^{ensemble}$. The fit improves rapidly as the data gets more representative.

[15] As of October 2022, this gap has been shortened from two months to 25 working days.

October 2022.

2. Energy Price Guarantee (EPG). In September 2022, the UK government announced the Energy Price Guarantee (EPG) programme as a response to the ongoing energy crisis. Another form of uniform price cap, the EPG reduces the maximum per unit rate below the level of the October 2022 price cap in an attempt to limit the average household energy bill to around £2,500. As discussed in Fetzer (2022), the standing charge is maintained at the level of the October 2022 price cap.

3. Two-tier tariff. As an alternative policy proposal to the EPG, discussed in more detail in Fetzer (2022), we consider a two-tier tariff such that the standing charge is fixed at the level of the October 2021 price cap, as are unit prices for the first 9,500 kWh of natural gas consumption and the first 2,500 kWh of electricity consumption. As 50% of UK households consume less than 12,100 kWh of natural gas and 2,900 kWh of electricity, this would drastically limit energy price increases for the bulk of households.[16] We consider a second tier unit price of 20 pence per kWh for natural gas and 60 pence per kWh for electricity, which, together with the first tier described above, would have a similar cost to the government as the EPG. This tariff would offer much more targeted support without undermining the incentive to save energy created by higher unit prices.

We are therefore able to produce vectors of spending estimates for our property-level energy consumption estimates $\mathbf{E_p}$. For example, for the preferred ensemble average energy consumption estimate $E_p^{ensemble}$, we produce the following four spending estimates that use the October 2021, October 2022, EPG and Two-tier tariff scenarios, respectively:

$$\mathbf{C}_p^{ensemble} = (C_{p,21}^{ensemble}, C_{p,22}^{ensemble}, C_{p,EPG}^{ensemble}, C_{p,Two-tier}^{ensemble})$$

These estimates allow us to measure changes in energy bills under different price scenarios and policy interventions at the individual property-level. We next

[16]See `https://www.gov.uk/government/statistics/national-energy-efficiency-data-framework-need-consumption-data-tables-2021`.

carry out a distributional analysis under different price scenarios using data aggregated to the MSOA-level to characterise how these measures affect households with different characteristics.

3 Explaining variation in exposure to energy price shocks

We begin by statistically characterising which areas in England would be hardest hit by hikes to energy prices. To do so, we focus on simulated energy bills assuming the UK government had not intervened in energy markets. This exercise is relevant beyond the energy crisis due to the necessary roll out of carbon taxation to tackle climate crisis and energy security. As such, this correlational analysis can cast some light on the likely distributional effects of carbon taxation to come.

3.1 Empirical approach

For each mid-layer super output area (MSOA) we compute the percent increase in the median household's expected annual energy bills between the October 2021 and the October 2022 price cap using *actual* estimated consumption as:

$$\Delta C_m^{22-21} = (C_m^{22} - C_m^{21})/C_m^{21} \tag{1}$$

We use 2011 census boundaries and consider a vector of k socio-economic variables $\mathbf{X}_m$ at the MSOA level.

Our main focus is to statistically characterise which areas are more exposed to the energy price shock to understand which observable features of an area's resident population best explain the variation in the modelled exposure of a representative consumer across MSOAs. That is, we are interested in the extent to which linear combinations of features $x_{m,j}$ are producing fitted values $\Delta \widehat{C_m^{22-21}}$ that best capture the variation in our modelled exposure measure. This analysis will help us formulate a broad set of hypotheses about the likely implications that the energy price shock and any future carbon taxation may have across a broad range of socio-economic divides.

To do so, we perform a best subset selection (BSS) procedure. Best subset selection is a machine learning method used to perform "feature selection" in settings where the aim is to reduce dimensionality of a feature space (Guyon and Elisseeff, 2003). The idea of best subset selection is to estimate all possible regressions including all combinations of control variables and return the statistically optimal model, which minimizes an information criterion. The BSS algorithm we employ finds the solution to the following non-convex combinatorial optimization problem:

$$min_{\beta} \underbrace{\sum_{c=1}^{C} (\Delta C_m^{22-21} - \beta_0 - \sum_{j=1}^{k} x_{mj}\beta_j)^2}_{\text{Residual sum of squares}} \text{ subject to } \sum_{j=1}^{k} \mathbf{I}(\beta_j \neq 0) \leq s, \tag{2}$$

where k is the set of regressors of which a subset s is chosen to maximize overall model fit. The result is a sequence of models $\mathcal{M}_1, ..., \mathcal{M}_s, .., \mathcal{M}_k$, where the overall optimal model $\mathcal{M}_{s^*}$ is chosen by using either cross validation or some degree-of-freedom-adjusted measure of goodness of fit such as the Akaike information criterion (AIC). Throughout, we use the AIC, which measures the quality of a model by weighing up its goodness-of-fit against its simplicity, i.e. the number of features that are included in the statistical model.[17]

We build features in x_m from public data sources described in more detail in Appendix G. We consider variables characterising deprivation, housing and labor markets, households' economic and demographic features, and residential properties in each MSOA and group them under these headings. It is important to highlight that the best subset selection approach may yield models of different complexity that are non-nested. For each group, we present the sequence of "best" models for each number of regressors s and explore how the inclusion of more covariates expands the goodness of fit. One caveat with this approach is that certain variables may be dropped in case they are highly correlated with each other. That is, even if a predictor x^i contains a distinct signal conditional on x^j, it may be dropped from the analysis as the signal contained is not sufficiently strong.

[17]We note that BSS is computationally complex, leading us to pre-select variables for their theoretical relevance. Suppose there are k potential regressors. Best subset selection proceeds as follows: the first model estimates – using OLS – all $\binom{k}{1} = k$ different models containing a single regressor and chooses as optimal the model that results in the largest reduction in the residual sum of squares. The second model estimates all possible $\binom{k}{2}$ models containing exactly two regressors, and so on. In total, $\sum_{j=1}^{k} \binom{k}{j} = 2^k$ models are estimated. With $k = 30$ this amounts to estimating just over one billion regressions.

3.2 Results

We present the results in Tables 1-6, which look at different variable groups separately, first.[18] In these tables, the BSS algorithm introduces more variables in each column. The order in which these are added reflects the signal carried by each variable.

Occupation Table 1 looks at the make-up of the resident population based on their occupation. These data measure the share of residents that are classed to work in elementary or basic routine occupations at the lower end, and in managerial- or directorship roles at the upper end. The median household's exposure to the energy price shock appears to be most noticeably negatively correlated with the population share employed in elementary occupations and in associate professional, technical and service occupations. These two variables together explain around 18.7% of the variation in the dependent variable, and are included in all models in a nested fashion, suggesting a robust correlation structure. These patterns suggest that the energy price shock disproportionally affects areas where residents are more likely to be employed in higher status occupations, possibly because occupation status is correlated with property size. It is worth noting that the model fit improves little when adding more features beyond the two described above.

Deprivation Table 2 focuses on proxies for deprivation of the resident population in each MSOA, primarily the English Index of Multiple Deprivation (IMD). This index measures an area's relative deprivation drawing on 39 separate indicators organised across seven distinct domains, which we consider here. We also consider measures of employment participation and fuel poverty, that is the share of households whose home has a Fuel Poverty Energy Efficiency Rating (FPEER) of band D or below and whose income is below the poverty line after subtracting their modelled energy costs and housing costs.

The best model among all models with a single variable includes the share of the resident population that is unemployed, as per the 2021 census. The higher the share of unemployed population in an area, the lower is the estimated exposure of

[18] Appendix Tables A1-A6 repeat the same exercise looking at changes in average prices, with similar results.

the median household to the energy price shock. This measure by itself explains more than 16% of the variation in the median household's exposure to the energy price shock. This result is notable given that the best model with a single feature capturing an area's occupational profile captures only around 8% of the variation in the exposure measure. Once we include two features, the share unemployed drops out of the model and the share inactive and the income deprivation score are included. The combined R2 increases by around 4 percentage points which suggests that these two variables together contain similar amount of statistical signal compared to the measure of the share of the unemployed.

The share of households that are classified as fuel poor is only included in the best model with four features. The sign is as expected: areas with more resident classed as fuel poor experience higher exposure. The sign is also positive for the share of residents that are economically inactive. By contrast, all deprivation scores are negatively correlated with exposure to the energy price shock of the median household, again suggesting that the energy price shock disproportionally affects areas where households are relatively better off. The observation that areas with relatively higher levels of deprivation appear *less exposed* to the energy price shock is an indication that the bulk of the energy efficiency gains are unlikely to be found in the most deprived areas.

Housing market Table 3 focuses on characteristics of the housing market at the MSOA level. Both the share of underoccupied properties and the share of overcrowded properties are selected early on by the model, with a positive sign, and never dropped.[19] Moreover, these measures have high quantitative importance: the share of underoccupied properties has the highest explanatory power among all univariate features that we consider with an R2 of 30%, indicating that the variation in this measure captures almost a third of the variation in the median household's

[19]The measures of underoccupation and overcrowding are constructed based on the Bedroom Standard, which judges whether a household's accomodation has sufficient bedrooms given the makeup of the household. According to the Bedroom Standard, the following should have their own bedroom: each adult couple, any remaining adults over 21 years, two individuals of the same sex aged 10-20 years, an individual aged 10-20 years and another individual aged 9 years or under of the same sex (if there are an odd number of indivduals of this sex aged 10-20), an individual aged 10-20 (if there are no individuals of the same sex aged 0-9 to pair with them), two children aged 9 years or under, any remaining child aged 9 years or under.

exposure to the energy price shock.

Across MSOAs, nearly 70% of properties are underoccupied in England, while around 5% of properties are over-occupied. This pattern points to a stark issue in terms of the allocative inefficiency of the housing market that may coincide with demographic change and changing patterns of mating and cohabitation. The unconditional correlation between the share of residents that are aged above 65 and the under-occupation measure is 77%. Given that areas with a high share of underoccupied properties face higher exposure to the energy price shock, significant reallocation gains may be realized if those elderly residents downsize in the wake of the financial pressure that the energy price shock entails. By contrast, younger families in overoccupied homes may face notable financial barriers to make their homes more energy efficient. We posit, owing to the demographic similarities across much of Western Europe that a similar pattern arises in terms of energy efficiency and underoccupation across most of Western Europe.

Encouraging downsizing of homes may thus achieve multiple societal objectives, including reducing the need to build more homes, which typically comes with a high carbon footprint. Yet, past policy making in this domain has been inconsistent and potentially motivated more by political concerns rather than concerns to achieve a more efficient allocation of the housing stock. Among the measures that aimed to encourage downsizing was the so-called "bedroom tax" which implied a penalty reduction in housing benefits if low-income households lived in an underoccupied property. This policy directly increased support for populist policies and Brexit (Fetzer, 2019). One the other hand, policies such as the expansion of the state pension through the "so-called triple lock" reduced economic incentives to downsize. The triple lock was introduced in 2010 and effectively implied above wage- and inflation growth increases in the generosity of the state pension, thus reducing the need to cash out of large homes.

Table 3 also shows that areas with a high share of households living in rental accommodations – either private or social rented – see, on average, a lower exposure to the energy price shock. This is not surprising given that rental properties and in particular social housing typically include flats or apartments in recently-built multi-unit housing units which are, on average, more energy efficient (see also Ahlrichs et al., 2022; ONS, 2021). Further, rental homes were the focus of the Decent Homes Programme, which sought to bring properties to minimum efficiency standards by 2010 (Leicester and Stoye, 2017). Nevertheless, there may be significant variation across rental homes in terms of their relative attainable energy efficiency. Indeed, Petrov and Ryan (2021) and ONS (2021) show that rented homes are generally less energy efficient than owner-occupied homes in England and Ireland (although these properties differ also on other dimensions). Finally, policies such as "right-to-buy" may have resulted in a fragmented ownership structure, in particular in multi-unit blocks, which may impose higher coordination costs to implement energy efficiency investments.

Income Because MSOA-level administrative data on incomes and wealth are not available, in Table 4 we focus on proxies of wealth, such as housing prices per square meter, and a survey-based, statistically-fitted income variable. We find that MSOAs with higher income are more exposed to the energy price shock. This result is in line with findings in Tables 2 and 3. The measures of property prices per square meter do not add much additional explanatory variation and appear with negative signs. This pattern masks a strong positive correlation between the annual income of households and the underlying property prices: the unconditional correlation between the property price measures and the income variable stands at around 77%.

Demographics Next we focus on a range of measures pertaining to the demographic makeup of each MSOA's resident population. Our findings can speak to the likely distributional impacts of carbon taxation, for example across generations and communities. Table 5 presents these results. Among these variables, the single most important correlate of the median household's exposure to the energy price shock is the age make-up of an MSOA: areas with a high share of older residents see a disproportionate higher exposure. This finding is in line with our findings on underoccupation, which is more prevalent in older demographics. We further note that areas in which there is a higher share of larger households, typically households with children, are significantly more exposed, consistent with our findings on overoccupation. The point estimate suggests that a one standard deviation higher share of households with three or more members exhibits a 4.1-5.5 percentage point higher exposure to the energy price shock.

Consistent with the earlier findings on occupation, deprivation, and income, the exposure of the median household to the energy price shock appears to be also correlated with education. Both areas with a high share of residents without formal qualification and areas with a high share of relatively advanced educational attainment (professional degrees or university degrees) appear to be disproportionally exposed. Finally, both areas with a high share of EU migrants and those with a high share of UK-born residents see a higher exposure. By contrast, areas with a high share of residents born in non-EU countries see a lower exposure.

Property characteristics Table 5 investigates the correlation between property characteristics and changes in energy bills due to the energy crisis. Not surprising areas with newer homes and apartment buildings are less affected by the energy crisis, as these are more likely to be energy efficient, while MSOAs with older properties and terraced, semi-detached, and detached houses are more affected. Importantly, MSOAs with older properties might also be more likely to have conservation areas that restrict households' ability to invest in energy efficiency (Fetzer, 2023b).

Combined best model Finally, Appendix Table A7 considers all these variables together.[20] Column 1 performs a BSS on the subset of variables that for each group are included in the best model (marked by an "X" in Tables 2-6), while Column 2 includes all of them in a linear regression. Insights from the group-by-group results carry through. Overall, these variables explain almost 70% of the variation in the change in energy bills brought about by the energy crisis. Our analysis thus underlines that targeting of interventions in energy markets will be key to ensuring these interventions reach the right set of households (Allcott and Greenstone, 2017; Knittel and Stolper, 2019). In the next section, we document that the existing policy framework appears untargeted.

[20]Appendix Table A8 reports results for the combined model looking at the bill shock for the average household in an MSOA.

4 (How) are policy interventions affecting exposure?

The previous analysis highlighted what MSOA characteristics we expect to correlate with the energy price shock and potential future carbon taxation. We characterized correlations with exposure to the energy price shock in the *absence of a policy intervention*. We next consider how this picture changes when we study patterns under the implemented policy.

4.1 Best subset selection result with energy market intervention

We perform the same analysis with a different dependent variable: the change in the median household's bills under the untargeted energy support policy (EPG) introduced by the UK government (Fetzer, 2022). The dependent variable is constructed as follows:

$$\Delta C_m^{EPG-21} = (C_m^{EPG} - C_m^{21})/C_m^{21} \tag{3}$$

The EPG introduces a wedge between consumer-facing prices and the regulated market prices. This wedge implies that the underlying variation in the energy bills that is driven by differences in energy consumption is suppressed.[21] Relative to the market prices set by the energy regulator Ofgem in October 2022, the EPG lowered the unit price of gas by nearly 50%. As a result, the variation in the left-hand side should also be compressed by a similar amount and we would expect the coefficient sizes and the goodness of fit to shrink, mechanically by 50%. However, what we observe is that the reductions in the goodness of fit and the point estimates differ across variable and regressor groups. This result points to the distributional impacts of the policy response, relative to what would have happened without intervention, with some consumer groups receiving more insurance and others less.

[21] For ease of illustration: consider the fact that a household i's energy bill basically is $bill_i = p \times q_i(\text{what, who, how})$. The price shock implies a much larger variance in estimated bills; the EPG intervention, by lowering prices, reduces this drastic increase in variation.

4.2 Best subset selection comparison

First, we compare the adjusted R2 goodness of fit measures for different variable groups for models looking at predicted changes energy bills for the median household in an MSOA, without government intervention, ΔC_m^{22-21}, (left) and under the EPG scenario, ΔC_m^{EPG-21}, (right) in Figure 4. Not surprisingly, the goodness of fit decreases across the board under the EPG as the EPG compresses variation in estimated energy bills by lowering the unit price of energy. Indeed, the overall adjusted R2 decreases by 30%, from 0.691 to 0.481.[22].

We also observe different reductions in the goodness of fit attributable to different regressor groups. The largest decreases in explanatory power appear among the occupation (47%), income (46%), and housing (45%) categories, while the reduction in the adjusted R2 is only 36% for deprivation measures. This suggests that the energy price guarantee benefits relatively well-off households compared to a scenario in which market prices would have prevailed.

4.3 Comparison of individual regressors

Second, we examine why better off household appear to benefit from the energy price guarantee. We make reference to the best subset selection results in Appendix Tables A9 - A14 that repeat the analysis in the previous section, which focused on the unmitigated energy price shock in Tables 1 - 6, but looking at bill shocks under the EPG. We focus on a subset of most relevant regressors, summarized in Figure 5 which plots selected coefficients from the best subset selection results in the respective tables for bill shocks under both the market prices and the EPG. Specifically, we plot the point estimate pertaining to a variable taken from the "best model" for that particular variable group.

The chart highlights how MSOA-level characteristics within different regressor groups are correlated with the median household's exposure to the energy price shock under market prices and the energy price guarantee, respectively. Mechanically, because the energy price guarantee limits energy price increases, coefficients should shrink towards zero. The extent of shrinkage, however, is heterogeneous,

[22]Appendix Figure A2 shows similar patterns when looking at variation in average energy bills

implying a lack of uniformity in the insurance provided by the energy price guarantee

We start with the occupation regressors. We note that there is a positive correlation between the share of managers, directors and senior officials and bill shocks under market prices. This correlation suggests that in areas with a higher share of managers, the average household – without the energy price guarantee – would have experienced a higher exposure to the energy crisis. By contrast, under the energy price guarantee, this feature is dropped from the regression. Symmetrically, we would predict that the median household is less exposed to the energy price shock in areas with a high share of individuals in elementary occupations under market prices. The energy price guarantee amplifies that negative coefficient, more than doubling its relative size. This result points to the insurance component of the EPG.

We next look at the deprivation measures. We saw that areas with a relatively high share of fuel poor would have been more exposed to the energy crisis on average. The energy price guarantee reduces this exposure with a near 50% shrinkage, which again points to the insurance component of the EPG. Turning to the income deprivation score measure, however, we see that the energy price guarantee increases the correlation between the income deprivation measure score and the median households exposure. This result suggests that, with the energy price guarantee in place, areas with a high degree of income deprivation may find themselves – relatively speaking – more exposed to the energy price shock.

Among demographic characteristics, we find a higher shrinkage for the coefficient on share of the population born in UK than for foreign-born populations. This pattern suggests that the EPG might provide more respite for UK-born households, likely due to tenure patterns such that foreign-born households live in rental accommodations. Moreover, the EPG does not seem to reduce the exposure households with three or more members (typically families) much. This result suggests that the incidence of the shock is mostly on the working age adult population with families – the middle of society, as was the case with housing benefit cuts (see e.g. Fetzer et al., 2022b) and other welfare reforms (see e.g. Fetzer, 2019).[23]

[23]This is not surprising because much of the toolbox used for ex-ante economic impact assessments uses micro-simulation based on survey data – not administrative data – implying that any

Overall, we note that the energy price guarantee appears to significantly weaken the "progressive" nature of the energy price shock, while offering insurance against it. Alternative, better-targeted policy interventions are possible. We illustrate one such policy in the next section, focusing on the median households annual income as explanatory variable to describe incidence.

4.4 Focus on income variation

With the EPG, the increase in bills that would have arisen if energy prices had been set as per the Ofgem price cap announced in October 2022 relative to the October 2021, ΔC_p^{22-21}, can be decomposed into two components. For each property p, the first component represents the increase from October 2021 bills to the energy bills faced by *consumers* under the EPG:

$$\Delta C_p^j = C_p^{EPG} - C_p^{21} \tag{4}$$

The second component represents the implicit subsidy that the government pays, that is the wedge between the Ofgem price cap and the EPG price.

$$\Delta C_p^{22-EPG} = C_p^{22} - C_p^{EPG} \tag{5}$$

Rather than performing a broad set of regressions we now perform a simple univariate regression focusing on the annual household income as main explanatory variable. While in the previous analysis we had no additional fixed effects as control variables, here we partial out district fixed effects $\alpha_{d(m)}$ to account for notable level differences across local authorities in the income measure:

$$\Delta C_m^{22-21} = \alpha_{d(m)} + x_m \times \beta + \epsilon_d$$

Figure 6 Panel A plots a binned scatterplot with the linear regression fit of the change in bills from October 2021 under different price scenarios against MSOA-level average household income. We consider three price scenarios: 2022 values without intervention, C_p^{22} (navy circles), 2022 values under the EPG, C_p^{EPG} (maroon diamonds), and 2022 values under the two-tier tariff, $C_p^{\text{Two-tier}}$ (gray squares). In

policy output is designed for the median or average household (representative agent).

the absence of government support, bills would have increased drastically between 2021 and 2022 and more so in high-income MSOAs. This is not surprising as the energy price shock has a greater effect on households that consume a lot of energy – who tend to be better off. We speculate that the energy price shock has a similar effect to carbon taxation.[24]

The energy price guarantee shifts consumer-facing prices downwards, thereby providing relief for all households based on levels of consumption. However, owing to the fact that wealthier households consume more energy and tend to live in particularly energy-inefficient properties, the EPG disproportionately benefits better-off areas (the maroon line is flatter than the blue one). In contrast, a two-tier tariff increases the income and energy bill gradient (gray line). In other words, a two-tier tariff resembles a lump sum transfer to households that have relatively low energy consumption. The marginal price signal is steepened maintaining energy saving incentives. This pattern arises from the fact that the two-tier tariff is much more targeted than the EPG implicit subsidy. Importantly, this exercise is for illustration purposes only: the tiers can be adjusted to provide more support to lower-income households, and more tiers can be introduced.

Figure 6 Panel B decomposes the effects of the EPG and the two-tier tariff into the consumer-facing component and the government subsidy. The attenuation in the income gradient of the energy shock under the EPG (maroon diamonds) is due, not surprisingly, to an attenuation of this gradient for the consumer-facing price shock, ΔC_p^{EPG-21}. Moreover, the government subsidy, ΔC_p^{EPG-22} also appears regressive in absolute terms: the government supports households that live in areas that are economically better off. By contrast, the income gradient of the consumer-facing energy price shock under the two-tier tariff is similar to that under market prices. Finally, the government subsidy is uncorrelated to area income under the two-tier tariff due to the better individual targeting properties of this price scheme.

[24] The evidence on whether a carbon tax is progressive or regressive is mixed and based largely on theoretical models. A notable empirical exception is Andersson and Atkinson (2020).

5 Conclusion

This paper develops a measurement framework that enables us to model the likely impact of the energy price shock across England and Wales and analyse the likely incidence of actual and counterfactual policy interventions. First, we find that, absent government intervention, the energy price shock has a more pronounced effect on relatively more affluent areas, characterised by a higher fraction of high-income, high-status occupation, high-education, and elderly individuals, as well as underoccupied properties. These patterns highlight the likely progressive impact of carbon taxation that would similarly increase energy prices. Second, we document that that commonly used untargeted interventions in energy markets are quite incoherent, including present UK policy path. Indeed, the energy price guarantee disproportionately benefits well-off households because 1) the reduction in the unit rate relative to market prices disproportionately benefits households with high energy consumption and 2) energy consumption increases with income. Alternative, more targeted policies are cheaper, easily implementable, and could better align incentives.

Notably, the EPG lowers the prices that consumers face. As a result, energy efficiency upgrade investments under the EPG appear less economical. Indeed, in a companion paper we estimate that the EPG weakens incentives to invest in energy efficiency upgrades by around 30% (Fetzer et al., 2022a).

Crucially, the UK government has much of the data needed to ensure timely, targeted, and cost-effective interventions at its disposal. With these data, a broad set of alternative policies could be considered. For example, instead of a uniform price cap, the government could propose a two-tier tariff providing more generous targeted support without eroding energy savings incentives. Alternatives that provide even more targeted support with better incentive preservation may also be implementable (see Bhattacharjee et al., 2022; Bachmann et al., 2022). The two tier-tariff could be designed to have a similar costing as the government's uniform price cap but could be even more targeted.

References

Adam, Stuart, Isaac Delestre, Peter Levell, and Helen Miller, "Tax policies to help achieve net zero carbon emissions," in Carl Emmerson, Paul Johnson, and Ben Zaranko, eds., *IFS Green Budget 2021*, IFS, 2022, chapter 8, pp. 355–391.

Ahlrichs, Jakob, Simon Wenninger, Christian Wiethe, and Björn Häckel, "Impact of socio-economic factors on local energetic retrofitting needs - A data analytics approach," *Energy Policy*, 2022, *160* (February 2021).

Allcott, Hunt and Michael Greenstone, "Is there an energy efficiency gap?," *Journal of Economic perspectives*, 2012, *26* (1), 3–28.

_ **and** _ , "Measuring the welfare effects of residential energy efficiency programs," Technical Report, National Bureau of Economic Research 2017.

Andersson, Julius and Giles Atkinson, "The Distributional Effects of a Carbon Tax: the Role of Income Inequality," *Centre for Climate Change Economics and Policy Working Paper 378/Grantham Research Institute on Climate Change and the Environment Working Paper 349*, 2020, *5709* (378), 1–35.

Ari, Anil, Nicolas Arregui, Simon Black, Oya Celasun, Dora Iakova, Aiko Mineshima, Victor Mylonas, Ian Parry, Iulia Teodoru, and Karlygash Zhunussova, "Surging Energy Prices in Europe in the Aftermath of the War: How to Support the Vulnerable and Speed up the Transition Away from Fossil Fuels," 2022.

Attari, Shahzeen Z., Michael L. DeKay, Cliff I. Davidson, and Wändi Bruin De Bruin, "Public perceptions of energy consumption and savings," *Proceedings of the National Academy of Sciences of the United States of America*, 2010, *107* (37), 16054–16059.

Bachmann, Rüdiger, Moritz Kuhn, Rüdiger Bachmann, David Baqaee, Christian Bayer, Moritz Kuhn, Andreas Löschel, Ben Mcwilliams, Benjamin Moll, Andreas Peichl, Karen Pittel, Moritz Schularick, and Georg Zachmann, "How it can be done," *EconTribute Working Paper*, 2022.

Bhattacharjee, Arnab, Max Mosley, and Adrian Pabst, "A ' Variable Energy Price Cap ' to Help Solve the Cost-of-Living Crisis," *NIESR Working Paper*, 2022, (September).

Blanco-Perez, Cristina and Abel Brodeur, "Publication bias and editorial statement on negative findings," *Economic Journal*, 2021, *130* (629), 1226–1247.

Brandon, Alec, Christopher M Clapp, John A List, Robert D Metcalfe, and Michael Price, "The Human Perils of Scaling Smart Technologies: Evidence from Field Experiments," Technical Report, National Bureau of Economic Research 2022.

BRE, "The Government's Standard Assessment Procedure for Energy Rating of Dwellings," Technical Report 2014.

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg, "Star wars: The empirics strike back," *American Economic Journal: Applied Economics*, 2016, *8* (1), 1–32.

_ , Nikolai Cook, and Anthony Heyes, "Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics," *American Economic Review*, 2020, *110* (11), 3634–3660.

Cassidy, Alecia, "How does mandatory energy efficiency disclosure affect housing prices?," *Journal of the Association of Environmental and Resource Economists*, 2023, *10* (3), 655–686.

Christensen, Peter, Paul Francisco, Erica Myers, and Mateus Souza, "Decomposing the wedge between projected and realized returns in energy efficiency programs," *The Review of Economics and Statistics*, 2021, pp. 1–46.

Cong, Shuchen, Destenie Nock, Yueming Lucy Qiu, and Bo Xing, "Unveiling hidden energy poverty using the energy equity gap," *Nature communications*, 2022, *13* (1), 2456.

Dalton, Ben and Franz Fuerst, "The 'green value' proposition in real estate: A meta-analysis," *Routledge Handbook of Sustainable Real Estate*, 2018, pp. 177–200.

Department for Business, Energy & Industrial Strategy, "Evaluation of the Domestic Private Rented Sector Minimum Energy Efficiency Standard Regulations," Technical Report 2020.

Department for Levelling Up, Housing & Communities, "A guide to energy performance certificates for the marketing, sale and let of dwellings," Technical Report 2017.

_ , "Energy Performance of Buildings Certificates: notes and definitions," Technical Report 2021.

Doremus, Jacqueline M, Irene Jacqz, and Sarah Johnston, "Sweating the energy bill: Extreme weather, poor households, and the energy spending gap," *Journal of Environmental Economics and Management*, 2022, *112*, 102609.

European Union, "Factsheet - Energy Performance of Buildings," Dec 2021.

Ferentinos, Konstantinos, Alex Gibberd, and Benjamin Guin, "Stranded houses? The price effect of a minimum energy efficiency standard," *Energy Economics*, 2023, *120*, 106555.

Fetzer, Thiemo, "Did Austerity Cause Brexit?," *American Economic Review*, nov 2019, *109* (11), 3849–3886.

_ , "Measuring the Epidemiological Impact of a False Negative: Evidence from a Natural Experiment," 2021, (596).

_ , "Subsidising the spread of COVID-19: Evidence from the UK'S Eat-Out-to-Help-Out Scheme*," *The Economic Journal*, oct 2021.

_ , "The Energy Price Guarantee: what next?," *CAGE Policy Briefing*, 2022, pp. 1–24.

_ , "Endogenous crime: Evidence from the policy response to the energy crisis in England," 2023.

_ , "Regulatory barriers to climate action: Evidence from Conservation Areas in England," *CAGE Working Paper*, 2023.

_ **and Thomas Graeber**, "Measuring the scientific effectiveness of contact tracing: Evidence from a natural experiment," *Proceedings of the National Academy of Sciences*, 2021, *118* (33), e2100814118.

_ **, Ludovica Gazze, and Menna Bishop**, "How large is the energy savings potential in the UK?," *CAGE wp 644*, 2022.

_ **, Srinjoy Sen, and Pedro C L Souza**, "Housing Insecurity and Homelessness: Evidence from the United Kingdom," *Journal of the European Economic Association*, nov 2022.

Gerarden, Todd D, Richard G Newell, and Robert N Stavins, "Assessing the energy-efficiency gap," *Journal of Economic Literature*, 2017, *55* (4), 1486–1525.

Ghosh, Arpita, Brendon McConnell, and Jaime Millán-Quijano, "Do Homebuyers Value Energy Efficiency? Evidence From an Information Shock," 2022.

Gregório, Vera and Júlia Seixas, "Energy savings potential in urban rehabilitation: A spatial-based methodology applied to historic centres," *Energy and Buildings*, 2017, *152*, 11–23.

Guertler, Pedro, Jack Carrington, and Antonia Jansz, "The Cold Man of Europe–2015," 2015.

Guin, Benjamin, Perttu Korhonen, and Sidharth Moktan, "Risk differentials between green and brown assets?," *Economics Letters*, 2022, *213*, 110320.

Guyon, Isabelle and Andre Elisseeff, "An Introduction to Variable and Feature Selection Isabelle," *Journal of Machine Learning Research*, 2003, *52* (3), 239–281.

Harari, Daniel, Brigid Francis-Devine, Paul Bolton, and Matthew Keep, "Research Briefing - Rising cost of living in the UK," *House of Commons Library*, 2022, pp. 4–5.

Hinson, Suzanna and Paul Bolton, "Fuel Poverty," May 2022.

Houde, Sébastien and Erica Myers, "Are consumers attentive to local energy costs? Evidence from the appliance market," *Journal of Public Economics*, 2021, *201*, 104480.

IEA, "Electricity Market Report - July 2022," Technical Report 2022.

Knittel, Christopher R and Samuel Stolper, "Using machine learning to target treatment: The case of household energy use," Technical Report, National Bureau of Economic Research 2019.

Labandeira, Xavier, José M Labeaga, and Xiral López-Otero, "A meta-analysis on the price elasticity of energy demand," *Energy Policy*, mar 2017, *102*, 549–568.

Leicester, Andrew and George Stoye, "Factors Associated with the Presence of Domestic Energy Efficiency Measures in England," *Fiscal Studies*, 2017, *38* (2), 331–356.

Levinson, Arik and Emilson Silva, "The electric gini: income redistribution through energy prices," *American Economic Journal: Economic Policy*, 2022, *14* (2), 341–65.

Lyubich, Eva, "The role of people vs. places in individual carbon emissions," Technical Report, HAAS Working Paper 324 2022.

Myers, Erica, "Are home buyers inattentive? Evidence from capitalization of energy costs," *American Economic Journal: Economic Policy*, 2019, *11* (2), 165–88.

__ , "Asymmetric information in residential rental markets: Implications for the energy efficiency gap," *Journal of Public Economics*, 2020, *190*, 104251.

__ , Steven Puller, Jeremy West et al., "Mandatory Energy Efficiency Disclosure in Housing Markets," *American Economic Journal: Economic Policy*, 2022, *14* (4), 453–87.

Ofgem, "Direct debits - What you need to know," Technical Report 2019.

— , "Ofgem updates price cap level and tightens up rules on suppliers," Technical Report 2022.

— , "Price cap – Decision on changes to the wholesale methodology," Technical Report 2022.

ONS, "Energy efficiency of housing in England and Wales: 2021," Technical Report 2021.

Petrov, Ivan and Lisa Ryan, "The landlord-tenant problem and energy efficiency in the residential rental market," *Energy Policy*, 2021, *157*, 112458.

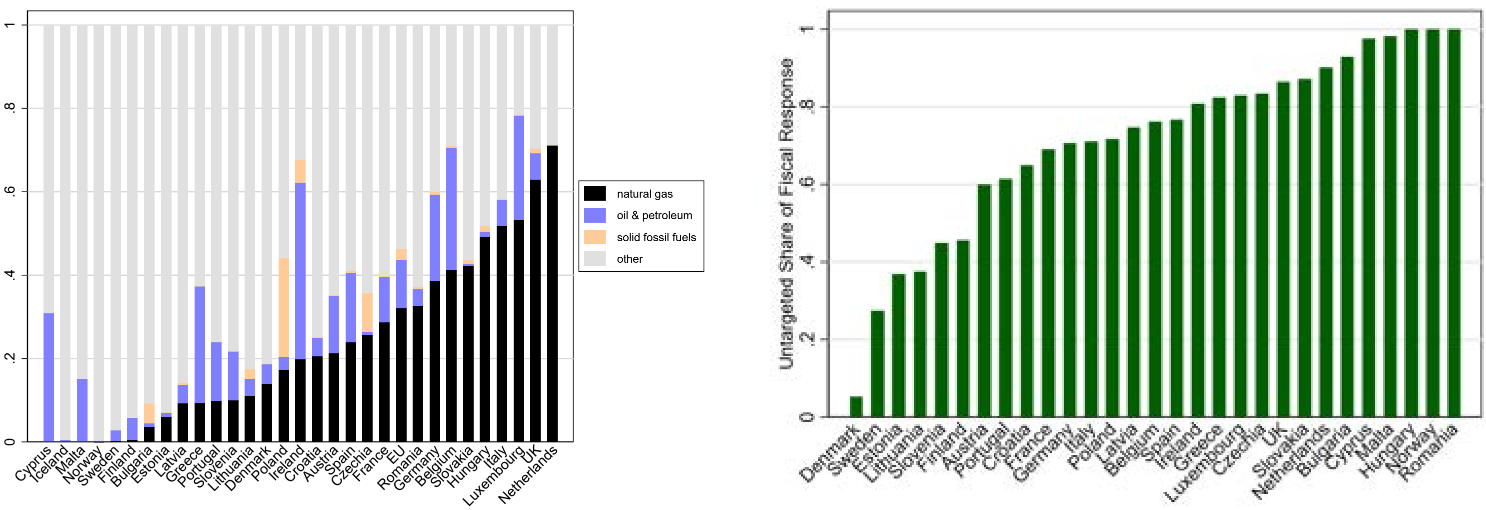
Ruhnau, Oliver, Clemens Stiewe, Jarusch Muessel, and Lion Hirth, "Gas demand in times of crisis: energy savings by consumer group in Germany," 2022.

Zhang, Tao, Peer Olaf Siebers, and Uwe Aickelin, "A three-dimensional model of residential energy consumer archetypes for local energy policy design in the UK," *Energy Policy*, 2012, *47*, 102–110.

Figures and tables

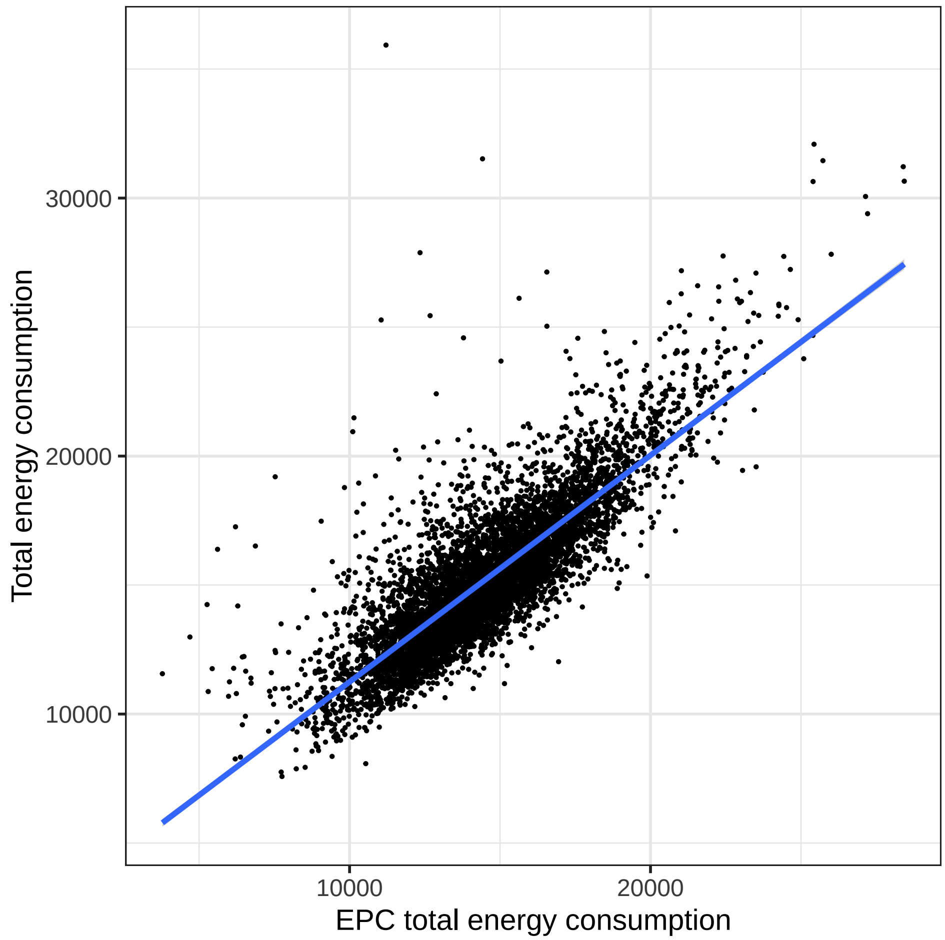
Figure 1: Fuel composition and untargeted policy response to energy crisis across EU countries

Panel A: Fuel composition | *Panel B:* Untargeted policy response



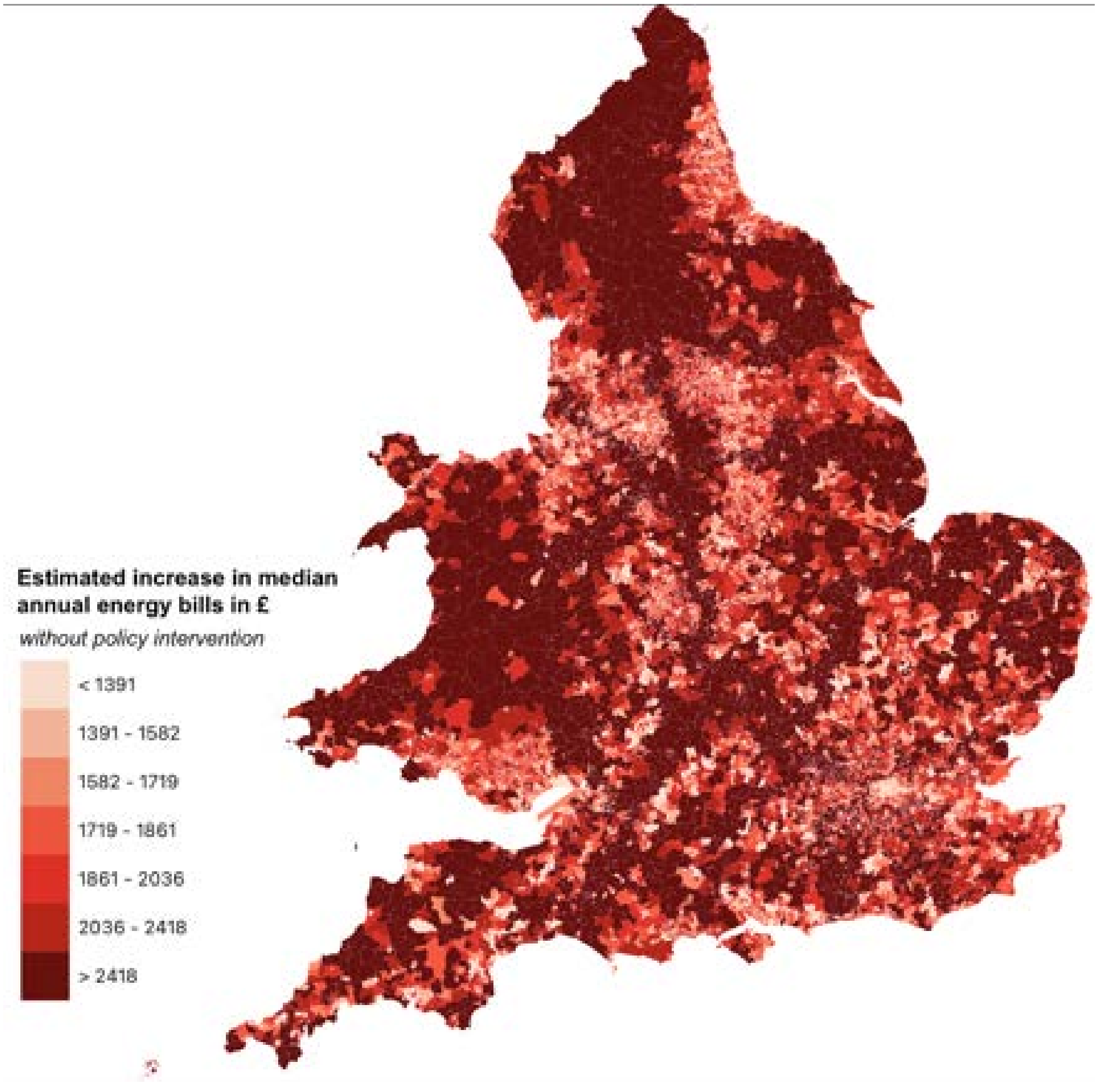
Notes: Panel A plots the share of each fuel in final energy consumption across EU countries in 2019 (source: Eurostat). Panel B plots the untargeted share of government budgets earmarked for measures to shield households from the energy crisis (Source: Bruegel, accessed on May 24, 2023).

Figure 2: Median property-level energy consumption at the MSOA-level compared with median imputed energy consumption measures from EPC-NEED data



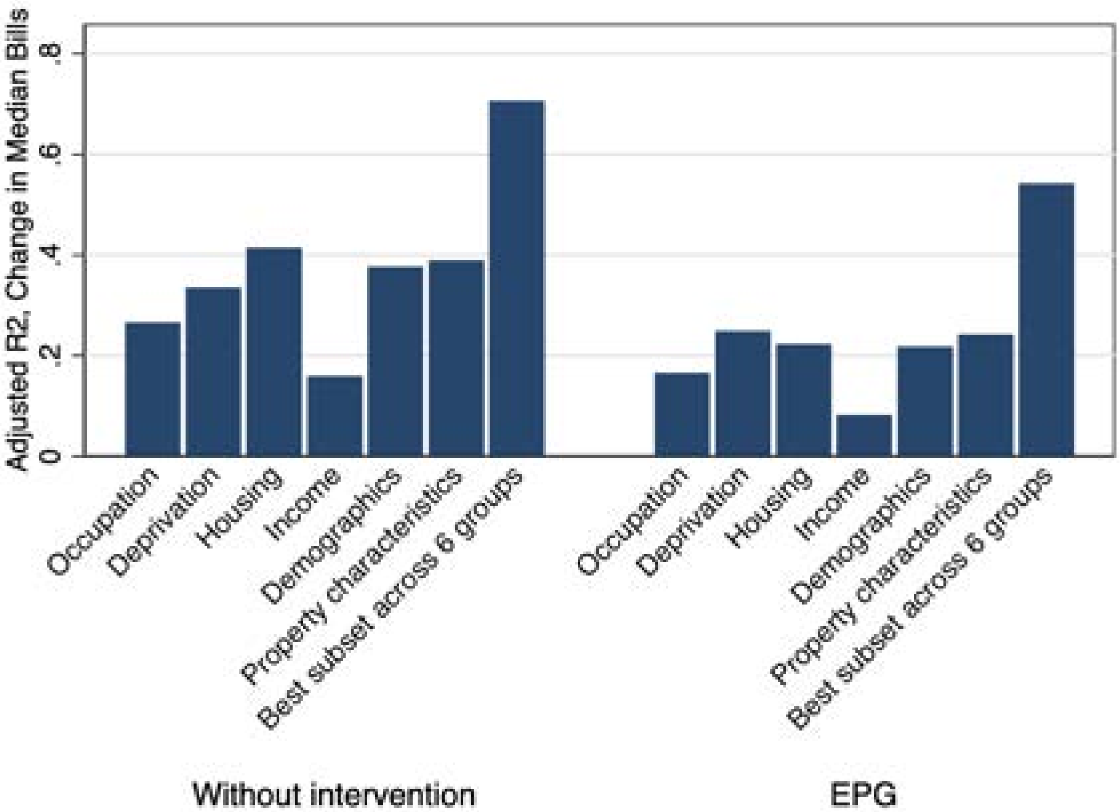
Notes: Figures provide a scatterplot of estimates of the median energy consumption per meter from published data at the MSOA-level (for metered electricity and gas only) on the vertical axis and the median of the ensemble imputed energy consumption measure on the horizontal axis.

Figure 3: Spatial distribution in estimated impact on annual energy bills for median household measured in £ without any policy intervention



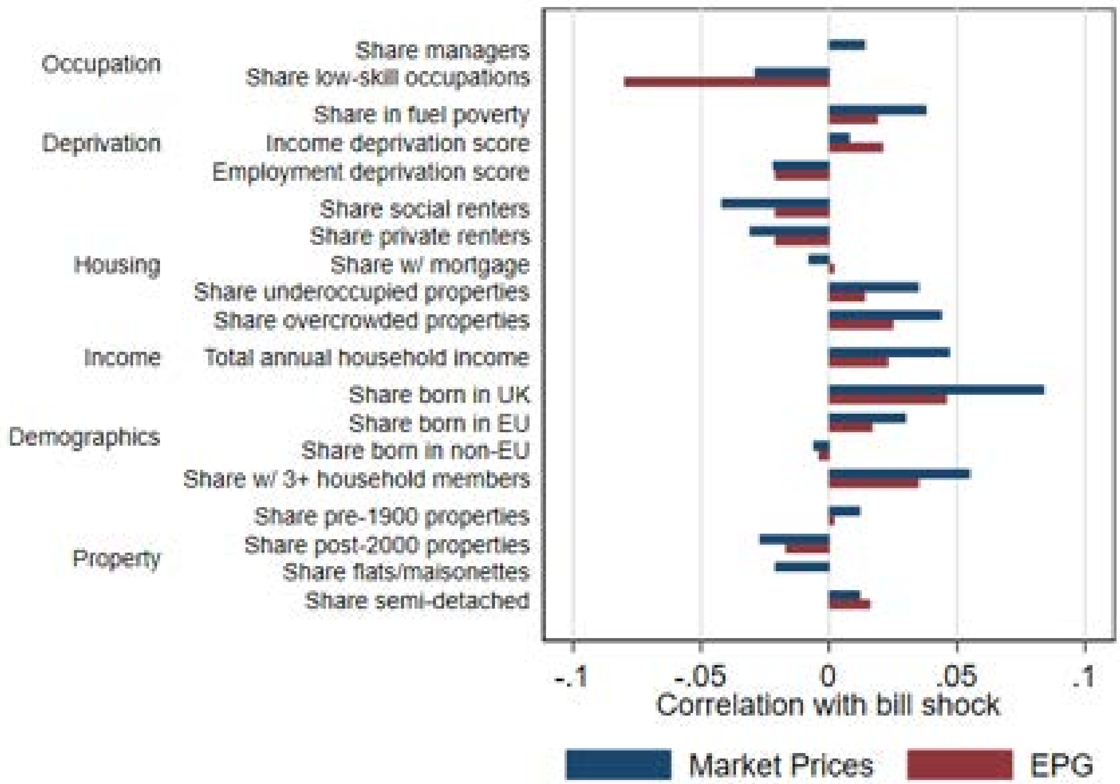
Notes: Figures plots the spatial distribution in energy bill expenditure for the median household across England and Wales.

Figure 4: Variation in the change in median energy bills between October 2022 and October 2021 explained by MSOA-level characteristics



Notes: the figure plots the adjusted R2 in Best Subset Selection regressions within each variable group without intervention and under the uniform price cap. The dependent variable is the change in the median energy bills in an MSOA between October 2022 and October 2021. Adjusted R2 are reported in Tables 1-6 and A8 (Without intervention) and Appendix Tables A9-A15 (EPG).

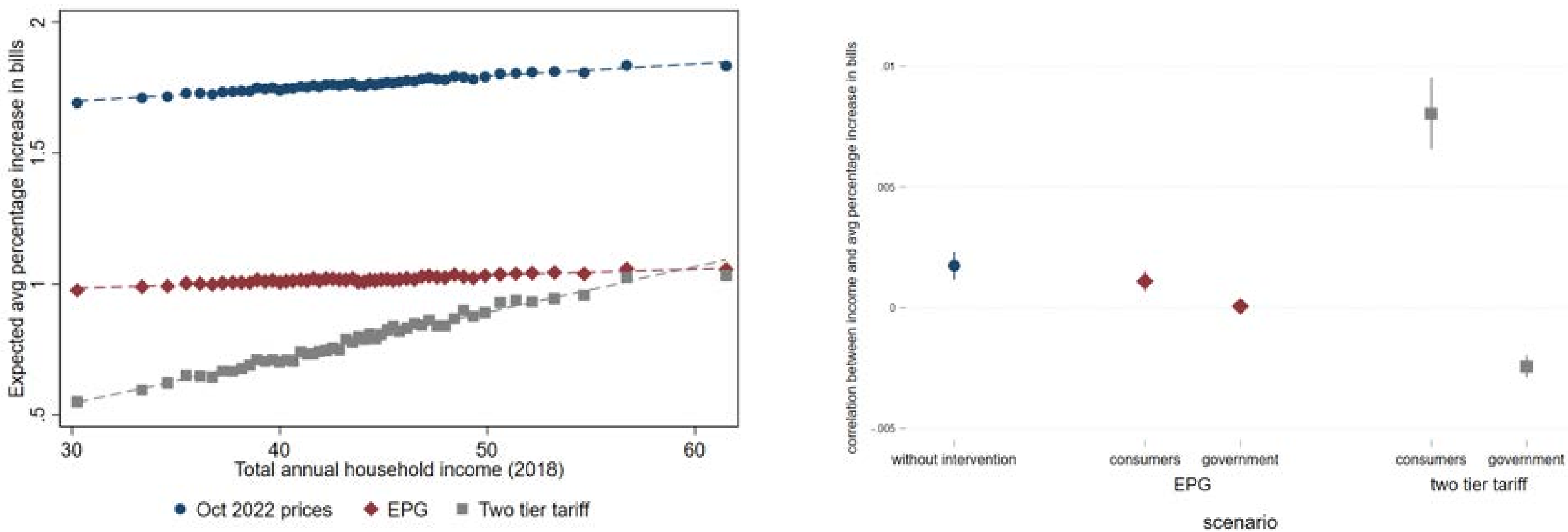
Figure 5: Notable changes in the correlation structure between the modelled median exposure of a household under no intervention and with the energy price guarantee



Notes: Figure plots some selected coefficients from Tables 1 - 6 and Appendix Tables A9 - A14. The chart highlights how the different MSOA level characteristics are correlated with the median households exposure to the energy price shock under market prices (navy bars) and under the implemented EPG (maroon bars) as it was implemented. For brevity, share managers refers to share managers, directors, and senior officials, share low-skill occupations refers to share elementary occupations, share w/ mortgage refers to share with mortgage or loan.

Figure 6: Comparison of income gradient in the energy price shock incidence under alternative price policies and decomposition consumer- and government-facing shocks

Panel A: Income gradient in the energy price shock incidence *Panel B:* Consumer- vs. government-facing shock



Notes: Panel A plots the relationship between average annual household income at the MSOA-level against the expected increase in the energy bills from October 2021 under three price scenarios to study the degree to which the specific measures are targeted in providing relief. Navy circles indicate the shock under October 2022 values without intervention, maroon diamonds indicate the shock under the EPG, and gray squares indicate the shock under a hypothetical two-tier tariff. Panel B presents regression results showing how the energy-price shock on average bills varies with income at the MSOA-level. The "without intervention" scenario shows the overall shock to bills between 2021 and 2022. Correlations under EPG and two-tier pricing are decomposed in consumer-facing shock and government subsidy. All regressions include district fixed-effects. Standard errors are clustered at the district level. Bars indicate 95% confidence intervals.

Table 1: Predictors of Percent Changes in Median Energy Bills Without Intervention: Occupational makeup of resident population

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of managers, directors and senior officials (2021)					0.014***	0.012***	0.014***
					(0.002)	(0.002)	(0.002)
Share of professional occupations (2021)							-0.010***
							(0.002)
Share of associate professional, technical and service occupations (2021)		-0.029***	-0.037***	-0.034***	-0.037***	-0.037***	-0.033***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Share of administrative and secretarial occupations (2021)			0.022***	0.020***	0.019***	0.021***	0.022***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of skilled trades, sales and customer service occupations (2021)						-0.006***	-0.010***
						(0.002)	(0.002)
Share of process, plant and machine operatives (2021)				0.012***	0.016***	0.018***	0.016***
				(0.002)	(0.002)	(0.002)	(0.002)
Share of elementary occupations (2021)	-0.023***	-0.030***	-0.026***	-0.035***	-0.029***	-0.029***	-0.029***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Best Subset							X
Observations	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.0765	.187	.243	.249	.26	.262	.266

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2: Predictors of Percent Changes in Median Energy Bills Without Intervention: Deprivation Indicators

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share in fuel poverty (2022)				0.021***	0.027***	0.035***	0.038***	0.037***	0.038***	0.038***
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Share unemployed (2021)	-0.034***						-0.011***	-0.009***	-0.011***	-0.012***
	(0.001)						(0.002)	(0.002)	(0.002)	(0.002)
Share inactive (2021)		0.028***	0.027***	0.029***	0.027***	0.024***	0.020***	0.022***	0.024***	0.024***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Income deprivation score rate (2019)		-0.026***								0.008
		(0.001)								(0.006)
Employment deprivation score rate (2019)			-0.028***	-0.042***	-0.044***			-0.011***	-0.015***	-0.022***
			(0.001)	(0.001)	(0.001)			(0.003)	(0.003)	(0.005)
Education deprivation score (2019)						-0.026***	-0.026***	-0.021***	-0.021***	-0.022***
						(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Health deprivation score (2019)						-0.028***	-0.024***	-0.018***	-0.020***	-0.019***
						(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Crime deprivation score (2019)									0.009***	0.009***
									(0.002)	(0.002)
Housing deprivation score (2019)			-0.021***	-0.022***	-0.019***	-0.023***	-0.020***	-0.020***	-0.019***	-0.020***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Living environment deprivation score (2019)					-0.014***	-0.017***	-0.016***	-0.016***	-0.016***	-0.017***
					(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Best Subset										X
Observations	6791	6791	6791	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.161	.208	.26	.29	.311	.328	.335	.337	.341	.341

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Predictors of Percent Changes in Median Energy Bills Without Intervention: Housing

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share social renters (2021)				-0.018***	-0.034***	-0.037***	-0.042***	-0.044***	-0.042***	-0.042***
				(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Share private renters (2021)					-0.021***	-0.028***	-0.033***	-0.033***	-0.031***	-0.031***
					(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Share home owners with mortgage or loan (2021)							-0.007***	-0.008***	-0.008***	-0.008***
							(0.002)	(0.002)	(0.002)	(0.002)
Share without a second address (2021)								0.018***	0.031***	0.031***
								(0.004)	(0.006)	(0.006)
Share with a second address in the UK (2021)						0.007***	0.007***	0.024***	0.035***	0.035***
						(0.001)	(0.001)	(0.004)	(0.005)	(0.005)
Share of properties in CT band A-B (2021)										-0.007***
										(0.002)
Share of properties in CT band C-D (2021)			-0.010***	-0.014***	-0.020***	-0.021***	-0.020***	-0.020***	-0.020***	-0.024***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of properties in CT band E-H (2021)									0.005***	0.000
									(0.001)	(.)
Share of underoccupied properties (2021)	0.046***	0.087***	0.091***	0.078***	0.047***	0.042***	0.038***	0.036***	0.035***	0.035***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Share of overcrowded properties (2021)		0.049***	0.054***	0.052***	0.045***	0.047***	0.045***	0.045***	0.044***	0.044***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Best Subset										X
Observations	6791	6791	6791	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.3	.402	.414	.436	.451	.457	.459	.461	.462	.462

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by “X”. Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Predictors of Percent Changes in Median Energy Bills Without Intervention: Income Indicators

	(1)	(2)	(3)
Total annual household income (2018)		0.039***	0.045***
		(0.002)	(0.002)
House prices - 10th percentile (2021)			-0.020***
			(0.003)
House prices - 90th percentile (2021)	-0.017***	-0.042***	-0.029***
	(0.001)	(0.003)	(0.004)
Best Subset			X
Observations	6791	6791	6791
Adjusted R2	.0401	.162	.171

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by “X”. Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Predictors of Percent Changes in Median Energy Bills Without Intervention: Demographics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share born in the UK (2021)			0.015***	0.022***	0.033***	0.028***	0.090***	0.084***
			(0.001)	(0.001)	(0.001)	(0.002)	(0.005)	(0.005)
Share born in EU countries (2021)							0.029***	0.030***
							(0.003)	(0.003)
Share born in non-EU european countries (2021)						-0.007***		-0.006***
						(0.001)		(0.001)
Share white (2021)							-0.040***	-0.037***
							(0.003)	(0.003)
Share aged 65 years and over (2021)	0.034***	0.055***	0.045***	0.044***	0.043***	0.043***	0.047***	0.047***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share with level 3 qualification or above (2021)				0.014***	0.048***	0.048***	0.049***	0.050***
				(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Share without qualifications (2021)					0.033***	0.032***	0.029***	0.029***
					(0.002)	(0.002)	(0.002)	(0.002)
Share with 3+ household members (2021)		0.041***	0.044***	0.049***	0.058***	0.057***	0.055***	0.055***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Best Subset								X
Observations	6791	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.164	.345	.362	.386	.409	.413	.426	.428

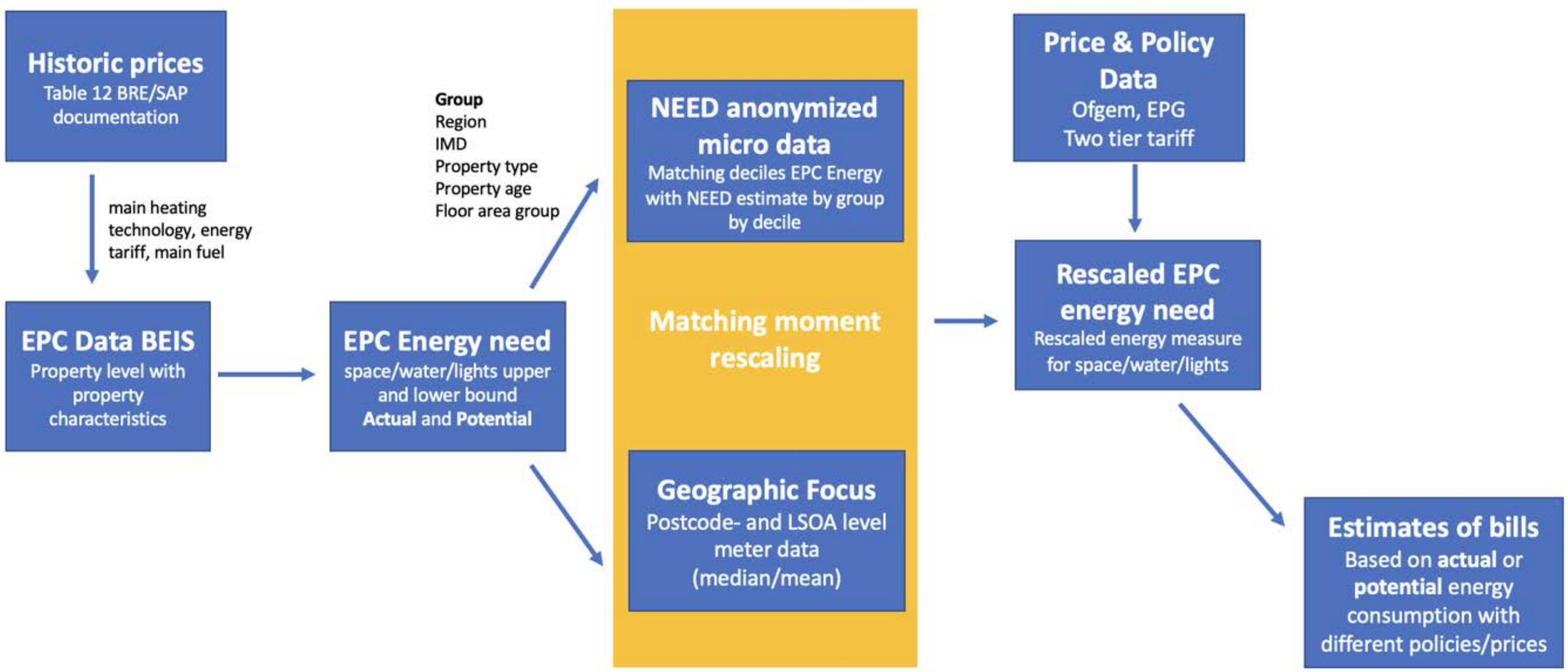
Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Predictors of Percent Changes in Median Energy Bills Without Intervention: Property Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of properties built pre-1900 (2021)					0.008***	0.012***	0.012***
					(0.001)	(0.001)	(0.001)
Share of properties built between 1900 and 1939 (2021)			0.028***	0.025***	0.026***	0.025***	0.024***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of properties built post-2000 (2021)		-0.029***	-0.034***	-0.029***	-0.027***	-0.027***	-0.027***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of flats/maisonettes in properties (2021)	-0.046***	-0.040***		-0.024***	-0.028***	-0.021***	-0.018***
	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)	(0.003)
Share of detached houses in properties (2021)			0.045***	0.030***	0.030***	0.033***	0.035***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Share of semi-detached houses in properties (2021)						0.012***	0.014***
						(0.001)	(0.002)
Share of terraced houses in properties (2021)							0.002
							(0.002)
Best Subset						X	
Observations	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.296	.414	.464	.518	.526	.537	.537

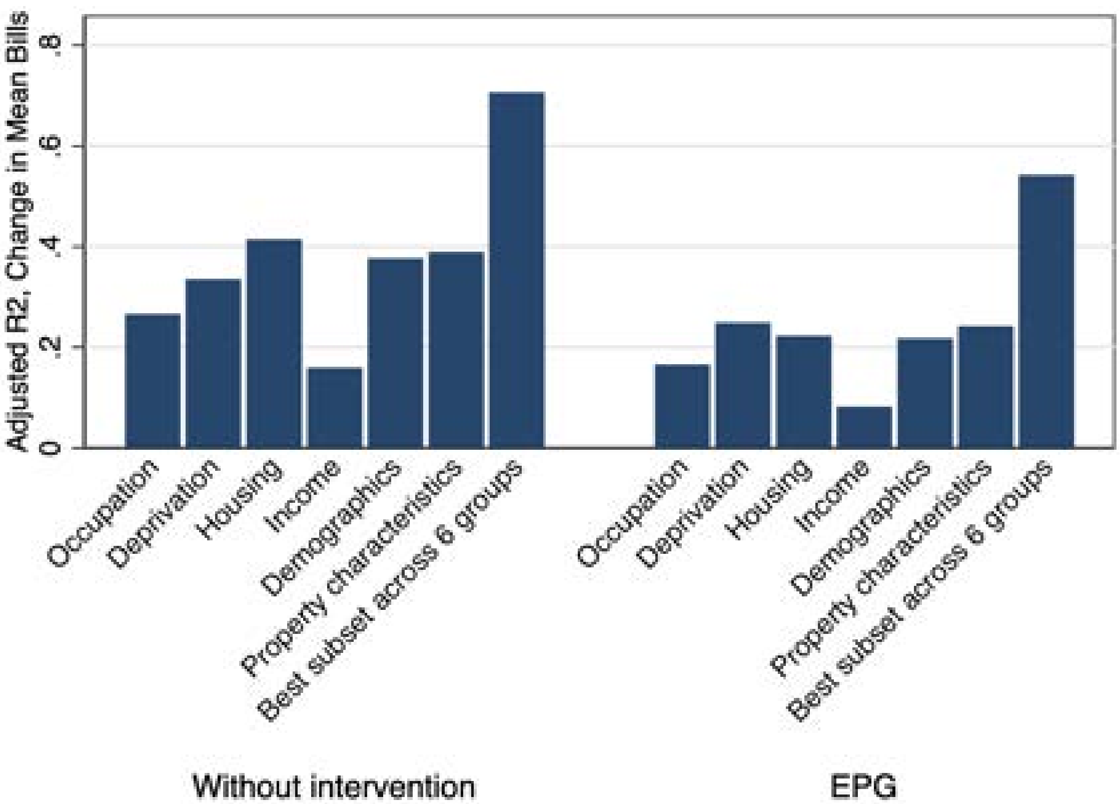
Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure A1: Schematic flowchart of the data processing pipeline to arrive at household-level energy price shock exposure measure



Notes: Figure provides a visual summary of the data construction process and the different steps and inputs that go into the derivation of the energy consumption and bill estimates.

Figure A2: Variation in the change in average energy bills between October 2022 and October 2021 explained by MSOA-level characteristics



Notes: the figure plots the adjusted R2 in Best Subset Selection regressions within each variable group without intervention and under the uniform price cap. The dependent variable is the change in the average energy bills in an MSOA between October 2022 and October 2021. Adjusted R2 are reported in Appendix Tables A1-A6 and A8 (Without intervention) and Appendix Tables A16-A22 (EPG).

Table A1: Predictors of Percent Changes in Average Energy Bills Without Intervention: Occupation Categories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of managers, directors and senior officials (2021)						0.007***	0.009***
						(0.002)	(0.002)
Share of professional occupations (2021)							-0.009***
							(0.002)
Share of associate professional, technical and service occupations (2021)		-0.022***	-0.028***	-0.030***	-0.025***	-0.027***	-0.024***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of administrative and secretarial occupations (2021)			0.033***	0.026***	0.024***	0.023***	0.024***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of skilled trades, sales and customer service occupations (2021)			-0.028***	-0.020***	-0.027***	-0.025***	-0.029***
			(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Share of process, plant and machine operatives (2021)					0.026***	0.027***	0.025***
					(0.002)	(0.002)	(0.002)
Share of elementary occupations (2021)	-0.025***	-0.030***		-0.016***	-0.033***	-0.030***	-0.030***
	(0.001)	(0.001)		(0.001)	(0.002)	(0.002)	(0.002)
Best Subset							X
Observations	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.0884	.156	.21	.232	.261	.263	.266

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2: Predictors of Percent Changes in Average Energy Bills Without Intervention: Deprivation Indicators

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share in fuel poverty (2022)				0.029***	0.034***	0.034***	0.035***	0.036***	0.036***	0.036***
				(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share unemployed (2021)								-0.006***	-0.007***	-0.008***
								(0.002)	(0.002)	(0.002)
Share inactive (2021)						0.010***	0.016***	0.014***	0.014***	0.015***
						(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Income deprivation score rate (2019)									0.004	0.017***
									(0.003)	(0.006)
Employment deprivation score rate (2019)										-0.015***
										(0.005)
Education deprivation score (2019)			-0.022***	-0.040***	-0.030***	-0.030***	-0.036***	-0.036***	-0.038***	-0.038***
			(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Health deprivation score (2019)		-0.025***			-0.019***	-0.020***	-0.029***	-0.027***	-0.028***	-0.026***
		(0.001)			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Crime deprivation score (2019)							0.018***	0.020***	0.019***	0.020***
							(0.001)	(0.002)	(0.002)	(0.002)
Housing deprivation score (2019)	-0.028***	-0.034***	-0.027***	-0.026***	-0.029***	-0.027***	-0.027***	-0.026***	-0.026***	-0.027***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Living environment deprivation score (2019)			-0.017***	-0.027***	-0.025***	-0.023***	-0.026***	-0.025***	-0.025***	-0.026***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Best Subset										X
Observations	6791	6791	6791	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.114	.199	.229	.276	.294	.308	.327	.328	.329	.33

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by “X”. Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3: Predictors of Percent Changes in Average Energy Bills Without Intervention: Housing

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share social renters (2021)			-0.014***	-0.053***	-0.054***	-0.046***	-0.038***	-0.035***	-0.035***	-0.035***
			(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Share private renters (2021)				-0.039***	-0.045***	-0.036***	-0.029***	-0.026***	-0.026***	-0.026***
				(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Share home owners with mortgage or loan (2021)							0.010***	0.010***	0.010***	0.010***
							(0.002)	(0.002)	(0.002)	(0.002)
Share without a second address (2021)					-0.009***	-0.008***	-0.009***	-0.008***	-0.008***	-0.005
					(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.006)
Share with a second address in the UK (2021)										0.002
										(0.006)
Share of properties in CT band A-B (2021)								-0.008***	0.000	-0.009***
								(0.002)	(.)	(0.002)
Share of properties in CT band C-D (2021)				-0.026***	-0.027***	-0.025***	-0.025***	-0.030***	-0.025***	-0.030***
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of properties in CT band E-H (2021)									0.005***	0.000
									(0.001)	(.)
Share of underoccupied properties (2021)	0.036***	0.075***	0.063***			0.017***	0.023***	0.025***	0.025***	0.024***
	(0.001)	(0.002)	(0.002)			(0.004)	(0.003)	(0.004)	(0.004)	(0.003)
Share of overcrowded properties (2021)		0.047***	0.044***	0.031***	0.035***	0.041***	0.043***	0.043***	0.043***	0.043***
		(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Best Subset								X		
Observations	6791	6791	6791	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.187	.279	.296	.349	.357	.36	.365	.368	.368	.368

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4: Predictors of Percent Changes in Average Energy Bills Without Intervention: Income Indicators

	(1)	(2)	(3)
Total annual household income (2018)	0.017***	0.049***	0.048***
	(0.001)	(0.002)	(0.002)
House prices - 10th percentile (2021)		-0.041***	-0.030***
		(0.002)	(0.003)
House prices - 90th percentile (2021)			-0.012***
			(0.003)
Best Subset			X
Observations	6791	6791	6791
Adjusted R2	.0414	.143	.148

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A5: Predictors of Percent Changes in Average Energy Bills Without Intervention: Demographics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share born in the UK (2021)				0.020***	0.031***	0.039***	0.077***	0.075***
				(0.001)	(0.002)	(0.002)	(0.005)	(0.005)
Share born in EU countries (2021)	-0.022***						0.024***	0.024***
	(0.001)						(0.003)	(0.003)
Share born in non-EU european countries (2021)								-0.003***
								(0.001)
Share white (2021)						-0.010***	-0.032***	-0.030***
						(0.002)	(0.004)	(0.004)
Share aged 65 years and over (2021)		0.041***	0.044***	0.032***	0.032***	0.032***	0.035***	0.035***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share with level 3 qualification or above (2021)			0.013***	0.018***	0.053***	0.051***	0.054***	0.054***
			(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)
Share without qualifications (2021)					0.035***	0.032***	0.031***	0.032***
					(0.002)	(0.002)	(0.002)	(0.002)
Share with 3+ household members (2021)		0.040***	0.044***	0.049***	0.058***	0.055***	0.056***	0.056***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Best Subset								X
Observations	6791	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.0678	.24	.263	.287	.314	.315	.324	.325

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A6: Predictors of Percent Changes in Average Energy Bills Without Intervention: Property Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of properties built pre-1900 (2021)						0.009***	0.009***
						(0.001)	(0.001)
Share of properties built between 1900 and 1939 (2021)			0.031***	0.027***	0.026***	0.027***	0.027***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of properties built post-2000 (2021)		-0.027***	-0.028***	-0.026***	-0.025***	-0.024***	-0.024***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of flats/maisonettes in properties (2021)	-0.034***	-0.029***					0.005*
	(0.001)	(0.001)					(0.003)
Share of detached houses in properties (2021)			0.037***	0.034***	0.038***	0.040***	0.044***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
Share of semi-detached houses in properties (2021)				0.015***	0.018***	0.022***	0.026***
				(0.001)	(0.001)	(0.001)	(0.002)
Share of terraced houses in properties (2021)					0.010***	0.010***	0.014***
					(0.001)	(0.001)	(0.002)
Best Subset							X
Observations	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.17	.274	.37	.4	.408	.417	.417

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by “X”. Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A7: Predictors of Percent Changes in Median Energy Bills Without Intervention: Blocked Variable Selection Approach

	(1)	(2)
Share of managers, directors and senior officials (2021)	0.015***	0.016***
	(0.002)	(0.003)
Share of professional occupations (2021)	-0.006**	-0.006**
	(0.003)	(0.003)
Share of skilled trades, sales and customer service occupations (2021)	-0.013***	-0.013***
	(0.002)	(0.002)
Share of process, plant and machine operatives (2021)	0.003**	0.003**
	(0.001)	(0.001)
Share of elementary occupations (2021)	0.008***	0.008***
	(0.001)	(0.001)
Share in fuel poverty (2022)	0.013***	0.013***
	(0.002)	(0.002)
Share unemployed (2021)	0.006***	0.006***
	(0.002)	(0.002)
Share inactive (2021)	-0.004	-0.005
	(0.003)	(0.004)
Employment deprivation score rate (2019)	-0.004*	-0.003
	(0.002)	(0.004)
Education deprivation score (2019)	-0.006***	-0.006***
	(0.002)	(0.002)
Health deprivation score (2019)	0.005**	0.005**
	(0.002)	(0.002)
Crime deprivation score (2019)	0.003**	0.003**
	(0.001)	(0.001)
Housing deprivation score (2019)	-0.010***	-0.010***
	(0.001)	(0.001)
Living environment deprivation score (2019)	-0.022***	-0.022***
	(0.002)	(0.002)
Share social renters (2021)	-0.033***	-0.032***
	(0.003)	(0.003)
Share private renters (2021)	-0.032***	-0.031***
	(0.003)	(0.004)
Share home owners with mortgage or loan (2021)	-0.015***	-0.015***
	(0.003)	(0.003)
Share without a second address (2021)	-0.013**	-0.015**
	(0.006)	(0.007)
Share with a second address in the UK (2021)	-0.010*	-0.012*
	(0.006)	(0.006)
Share of properties in CT band C-D (2021)	0.002**	0.002*
	(0.001)	(0.001)
Share of properties in CT band E-H (2021)	0.015***	0.015***
	(0.002)	(0.002)
Share of underoccupied properties (2021)	0.018***	0.019***
	(0.003)	(0.003)
Share of overcrowded properties (2021)	0.008***	0.009***
	(0.002)	(0.002)
Total annual household income (2018)	0.007***	0.008***
	(0.002)	(0.002)
House prices - 10th percentile (2021)	-0.013***	-0.013***
	(0.002)	(0.002)
House prices - 90th percentile (2021)	-0.010***	-0.011***
	(0.002)	(0.003)
Share born in the UK (2021)	0.030***	0.031***
	(0.005)	(0.005)
Share born in EU countries (2021)	0.008***	0.008***
	(0.002)	(0.002)
Share born in non-EU european countries (2021)	0.003***	0.003***
	(0.001)	(0.001)
Share white (2021)	-0.026***	-0.027***
	(0.003)	(0.004)
Share with level 3 qualification or above (2021)	0.014***	0.014***
	(0.004)	(0.004)
Share without qualifications (2021)	0.014***	0.014***
	(0.002)	(0.003)
Share with 3+ household members (2021)	0.010***	0.011***
	(0.002)	(0.002)
Share of properties built pre-1900 (2021)	0.016***	0.016***
	(0.001)	(0.001)
Share of properties built between 1900 and 1939 (2021)	0.014***	0.014***
	(0.001)	(0.001)
Share of properties built post-2000 (2021)	-0.021***	-0.021***
	(0.001)	(0.001)
Share of flats/maisonettes in properties (2021)	-0.004**	-0.005**
	(0.002)	(0.002)
Share of semi-detached houses in properties (2021)	0.002***	0.002***
	(0.001)	(0.001)
Observations	6791	6791
Adjusted R2	.7	.699

Notes: Table reports results from OLS regressions. Column 1 shows best subset across all 5 groups of variables analyzed in Tables 1 through 4. Column 2 is the full specification based on best subsets determined in Tables 1 through 4. For comparison, columns 3 through 6 re-display the optimal specifications from Tables 1 through 4. Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A8: Predictors of Percent Changes in Average Energy Bills Without Intervention: Blocked Variable Selection Approach

	(1)	(2)
Share of managers, directors and senior officials (2021)	0.015***	0.014***
	(0.002)	(0.002)
Share of professional occupations (2021)	-0.010***	-0.011***
	(0.002)	(0.003)
Share of associate professional, technical and service occupations (2021)	0.003***	0.003***
	(0.001)	(0.001)
Share of skilled trades, sales and customer service occupations (2021)	-0.021***	-0.021***
	(0.001)	(0.001)
Share of process, plant and machine operatives (2021)	0.004***	0.003**
	(0.001)	(0.001)
Share in fuel poverty (2022)	0.018***	0.017***
	(0.001)	(0.001)
Share unemployed (2021)	0.008***	0.008***
	(0.002)	(0.002)
Share inactive (2021)	-0.009***	-0.009**
	(0.002)	(0.004)
Income deprivation score rate (2019)	-0.014***	-0.014***
	(0.004)	(0.004)
Employment deprivation score rate (2019)	0.008**	0.007**
	(0.004)	(0.004)
Education deprivation score (2019)	-0.012***	-0.012***
	(0.002)	(0.002)
Health deprivation score (2019)	0.013***	0.013***
	(0.002)	(0.002)
Crime deprivation score (2019)	0.008***	0.008***
	(0.001)	(0.001)
Housing deprivation score (2019)	-0.018***	-0.018***
	(0.001)	(0.001)
Living environment deprivation score (2019)	-0.034***	-0.034***
	(0.001)	(0.001)
Share social renters (2021)	-0.039***	-0.040***
	(0.003)	(0.003)
Share private renters (2021)	-0.034***	-0.035***
	(0.003)	(0.003)
Share home owners with mortgage or loan (2021)	-0.010***	-0.010***
	(0.002)	(0.003)
Share of properties in CT band A-B (2021)	-0.023***	-0.024***
	(0.003)	(0.003)
Share of properties in CT band C-D (2021)	-0.017***	-0.017***
	(0.002)	(0.002)
Share of underoccupied properties (2021)	0.023***	0.023***
	(0.003)	(0.003)
Share of overcrowded properties (2021)	0.019***	0.019***
	(0.002)	(0.002)
Total annual household income (2018)	0.010***	0.009***
	(0.002)	(0.002)
House prices - 10th percentile (2021)	-0.014***	-0.014***
	(0.002)	(0.002)
House prices - 90th percentile (2021)	0.004**	0.004*
	(0.002)	(0.002)
Share born in the UK (2021)	0.025***	0.025***
	(0.005)	(0.005)
Share born in EU countries (2021)	0.011***	0.011***
	(0.002)	(0.002)
Share born in non-EU european countries (2021)	0.004***	0.004***
	(0.001)	(0.001)
Share white (2021)	-0.023***	-0.022***
	(0.003)	(0.003)
Share with level 3 qualification or above (2021)	0.016***	0.016***
	(0.003)	(0.004)
Share without qualifications (2021)	0.025***	0.025***
	(0.002)	(0.002)
Share with 3+ household members (2021)	0.008***	0.008***
	(0.002)	(0.002)
Share of properties built pre-1900 (2021)	0.018***	0.018***
	(0.001)	(0.001)
Share of properties built between 1900 and 1939 (2021)	0.014***	0.014***
	(0.001)	(0.001)
Share of properties built post-2000 (2021)	-0.014***	-0.014***
	(0.001)	(0.001)
Share of flats/maisonettes in properties (2021)	0.012***	0.012***
	(0.003)	(0.003)
Share of detached houses in properties (2021)	0.009***	0.009***
	(0.002)	(0.002)
Share of semi-detached houses in properties (2021)	0.012***	0.012***
	(0.002)	(0.002)
Share of terraced houses in properties (2021)	0.013***	0.014***
	(0.002)	(0.002)
Observations	6791	6791
Adjusted R2	.737	.737

Notes: Table reports results from OLS regressions. Column 1 shows best subset across all 5 groups of variables analyzed in Tables 1 through 4. Column 2 is the full specification based on best subsets determined in Tables 1 through 4. For comparison, columns 3 through 6 re-display the optimal specifications from Tables 1 through 4. Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A9: Predictors of Percent Changes in Median Energy Bills under the EPG: Occupation Categories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of managers, directors and senior officials (2021)							0.001
							(0.001)
Share of professional occupations (2021)						-0.008***	-0.008***
						(0.002)	(0.002)
Share of associate professional, technical and service occupations (2021)		-0.016***	-0.018***	-0.016***	-0.016***	-0.013***	-0.013***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of administrative and secretarial occupations (2021)		0.016***	0.013***	0.011***	0.014***	0.015***	0.015***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of skilled trades, sales and customer service occupations (2021)					-0.010***	-0.013***	-0.013***
					(0.002)	(0.002)	(0.002)
Share of process, plant and machine operatives (2021)				0.010***	0.014***	0.012***	0.012***
				(0.001)	(0.001)	(0.001)	(0.001)
Share of elementary occupations (2021)	-0.011***		-0.011***	-0.019***	-0.017***	-0.018***	-0.017***
	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Best Subset						X	
Observations	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.0316	.0838	.114	.122	.137	.141	.141

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A10: Predictors of Percent Changes in Median Energy Bills under the EPG: Deprivation Indicators

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share in fuel poverty (2022)				0.015***	0.018***	0.018***	0.018***	0.019***	0.019***	0.019***
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share unemployed (2021)					-0.010***			-0.006***	-0.005***	-0.008***
					(0.002)			(0.001)	(0.002)	(0.002)
Share inactive (2021)						0.008***	0.010***	0.008***	0.009***	0.010***
						(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Income deprivation score rate (2019)										0.021***
										(0.004)
Employment deprivation score rate (2019)									-0.003	-0.021***
									(0.002)	(0.004)
Education deprivation score (2019)				-0.019***	-0.017***	-0.013***	-0.015***	-0.015***	-0.014***	-0.016***
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Health deprivation score (2019)			-0.010***			-0.011***	-0.014***	-0.012***	-0.011***	-0.009***
			(0.001)			(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Crime deprivation score (2019)							0.005***	0.007***	0.007***	0.007***
							(0.001)	(0.001)	(0.001)	(0.001)
Housing deprivation score (2019)	-0.018***	-0.015***	-0.018***	-0.016***	-0.013***	-0.016***	-0.016***	-0.015***	-0.015***	-0.016***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Living environment deprivation score (2019)		-0.013***	-0.011***	-0.018***	-0.016***	-0.015***	-0.016***	-0.015***	-0.015***	-0.016***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Best Subset										X
Observations	6791	6791	6791	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.0923	.138	.159	.182	.198	.21	.213	.216	.216	.219

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by “X”. Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A11: Predictors of Percent Changes in Median Energy Bills under the EPG: Housing

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share social renters (2021)				-0.028***	-0.029***	-0.022***	-0.023***	-0.021***	-0.022***	-0.022***
				(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Share private renters (2021)				-0.027***	-0.030***	-0.024***	-0.023***	-0.021***	-0.022***	-0.022***
				(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Share home owners with mortgage or loan (2021)								0.002	0.002	0.002
								(0.002)	(0.002)	(0.002)
Share without a second address (2021)							0.013***	0.012***	0.010**	0.010**
							(0.003)	(0.003)	(0.005)	(0.005)
Share with a second address in the UK (2021)					0.006***	0.005***	0.017***	0.016***	0.014***	0.014***
					(0.001)	(0.001)	(0.003)	(0.003)	(0.004)	(0.004)
Share of properties in CT band A-B (2021)										0.001
										(0.001)
Share of properties in CT band C-D (2021)			-0.005***	-0.014***	-0.014***	-0.012***	-0.012***	-0.012***	-0.012***	-0.011***
			(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of properties in CT band E-H (2021)									-0.001	0.000
									(0.001)	(.)
Share of underoccupied properties (2021)	0.023***	0.047***	0.049***			0.014***	0.013***	0.014***	0.014***	0.014***
	(0.001)	(0.001)	(0.001)			(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Share of overcrowded properties (2021)		0.029***	0.031***	0.018***	0.020***	0.025***	0.025***	0.025***	0.025***	0.025***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Best Subset								X		
Observations	6791	6791	6791	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.146	.213	.22	.242	.248	.252	.254	.254	.254	.254

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A12: Predictors of Percent Changes in Median Energy Bills under the EPG: Income Indicators

	(1)	(2)	(3)
Total annual household income (2018)		0.020***	0.024***
		(0.001)	(0.001)
House prices - 10th percentile (2021)			-0.012***
			(0.002)
House prices - 90th percentile (2021)	-0.009***	-0.022***	-0.015***
	(0.001)	(0.002)	(0.002)
Best Subset			X
Observations	6791	6791	6791
Adjusted R2	.0225	.0863	.0921

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A13: Predictors of Percent Changes in Median Energy Bills under the EPG: Demographics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share born in the UK (2021)			0.012***	0.015***	0.021***	0.027***	0.050***	0.046***
			(0.001)	(0.001)	(0.001)	(0.002)	(0.004)	(0.004)
Share born in EU countries (2021)	-0.016***					0.006***	0.016***	0.017***
	(0.001)					(0.001)	(0.002)	(0.002)
Share born in non-EU european countries (2021)								-0.004***
								(0.001)
Share white (2021)							-0.019***	-0.017***
							(0.003)	(0.003)
Share aged 65 years and over (2021)		0.029***	0.021***	0.020***	0.020***	0.021***	0.022***	0.022***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share with level 3 qualification or above (2021)				0.007***	0.025***	0.027***	0.027***	0.027***
				(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Share without qualifications (2021)					0.019***	0.020***	0.017***	0.017***
					(0.002)	(0.002)	(0.002)	(0.002)
Share with 3+ household members (2021)		0.026***	0.028***	0.031***	0.036***	0.038***	0.035***	0.035***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Best Subset								X
Observations	6791	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.072	.205	.224	.234	.248	.25	.256	.258

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A14: Predictors of Percent Changes in Median Energy Bills under the EPG: Property Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of properties built pre-1900 (2021)						0.002**	0.002**
						(0.001)	(0.001)
Share of properties built between 1900 and 1939 (2021)			0.011***	0.014***	0.012***	0.012***	0.012***
			(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Share of properties built post-2000 (2021)		-0.019***	-0.016***	-0.017***	-0.018***	-0.017***	-0.017***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of flats/maisonettes in properties (2021)	-0.025***	-0.022***	-0.023***	-0.017***			-0.002
	(0.001)	(0.001)	(0.001)	(0.001)			(0.003)
Share of detached houses in properties (2021)				0.011***	0.024***	0.024***	0.022***
				(0.001)	(0.001)	(0.001)	(0.002)
Share of semi-detached houses in properties (2021)					0.015***	0.016***	0.015***
					(0.001)	(0.001)	(0.002)
Share of terraced houses in properties (2021)					0.012***	0.012***	0.011***
					(0.001)	(0.001)	(0.002)
Best Subset						X	
Observations	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.174	.267	.296	.313	.324	.325	.325

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A15: Predictors of Percent Changes in Median Energy Bills under the EPG: Blocked Variable Selection Approach

	(1)	(2)
Share of professional occupations (2021)	-0.007***	-0.006**
	(0.002)	(0.003)
Share of associate professional, technical and service occupations (2021)	0.004***	0.004***
	(0.001)	(0.001)
Share of administrative and secretarial occupations (2021)	0.002**	0.002**
	(0.001)	(0.001)
Share of skilled trades, sales and customer service occupations (2021)	-0.012***	-0.011***
	(0.001)	(0.001)
Share of process, plant and machine operatives (2021)	0.003**	0.003**
	(0.001)	(0.001)
Share of elementary occupations (2021)	0.004***	0.005***
	(0.001)	(0.001)
Share in fuel poverty (2022)	0.005***	0.005***
	(0.001)	(0.001)
Share unemployed (2021)	0.003**	0.003**
	(0.001)	(0.001)
Education deprivation score (2019)	-0.003*	-0.004**
	(0.002)	(0.002)
Health deprivation score (2019)	0.007***	0.006***
	(0.002)	(0.002)
Crime deprivation score (2019)	0.003***	0.003***
	(0.001)	(0.001)
Housing deprivation score (2019)	-0.010***	-0.010***
	(0.001)	(0.001)
Living environment deprivation score (2019)	-0.019***	-0.020***
	(0.002)	(0.002)
Share social renters (2021)	-0.020***	-0.021***
	(0.003)	(0.003)
Share private renters (2021)	-0.021***	-0.022***
	(0.003)	(0.003)
Share home owners with mortgage or loan (2021)	-0.008***	-0.008***
	(0.002)	(0.002)
Share of properties in CT band C-D (2021)	-0.003***	-0.003***
	(0.001)	(0.001)
Share of underoccupied properties (2021)	0.006**	0.006**
	(0.003)	(0.003)
Share of overcrowded properties (2021)	0.010***	0.011***
	(0.002)	(0.002)
Total annual household income (2018)	0.007***	0.008***
	(0.002)	(0.002)
House prices - 10th percentile (2021)	-0.003	-0.002
	(0.002)	(0.002)
House prices - 90th percentile (2021)	0.004**	0.003
	(0.002)	(0.002)
Share born in the UK (2021)	0.018***	0.020***
	(0.005)	(0.006)
Share born in EU countries (2021)	0.008***	0.008***
	(0.002)	(0.002)
Share white (2021)	-0.013***	-0.014***
	(0.003)	(0.004)
Share aged 65 years and over (2021)	0.007**	0.011***
	(0.003)	(0.004)
Share with level 3 qualification or above (2021)	0.019***	0.017***
	(0.002)	(0.004)
Share without qualifications (2021)	0.008***	0.008***
	(0.002)	(0.002)
Share with 3+ household members (2021)	0.010***	0.010***
	(0.002)	(0.002)
Share of properties built pre-1900 (2021)	0.012***	0.012***
	(0.001)	(0.001)
Share of properties built between 1900 and 1939 (2021)	0.009***	0.009***
	(0.001)	(0.001)
Share of properties built post-2000 (2021)	-0.010***	-0.010***
	(0.001)	(0.001)
Share of detached houses in properties (2021)	0.008***	0.008***
	(0.002)	(0.002)
Share of semi-detached houses in properties (2021)	0.008***	0.008***
	(0.001)	(0.001)
Share of terraced houses in properties (2021)	0.011***	0.011***
	(0.001)	(0.001)
Observations	6791	6791
Adjusted R2	.49	.489

Notes: Table reports results from OLS regressions. Column 1 shows best subset across all 5 groups of variables analyzed in Tables 1 through 4. Column 2 is the full specification based on best subsets determined in Tables 1 through 4. For comparison, columns 3 through 6 re-display the optimal specifications from Tables 1 through 4. Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A16: Predictors of Percent Changes in Average Energy Bills under the EPG: Occupation Categories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of managers, directors and senior officials (2021)						-0.009***	-0.008***
						(0.002)	(0.002)
Share of professional occupations (2021)			-0.020***		-0.013***		-0.007***
			(0.001)		(0.001)		(0.002)
Share of associate professional, technical and service occupations (2021)						-0.007***	-0.005***
						(0.001)	(0.001)
Share of administrative and secretarial occupations (2021)		0.015***	0.022***	0.015***	0.018***	0.020***	0.020***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of skilled trades, sales and customer service occupations (2021)	-0.017***	-0.021***	-0.034***	-0.030***	-0.036***	-0.031***	-0.034***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Share of process, plant and machine operatives (2021)				0.029***	0.023***	0.023***	0.022***
				(0.002)	(0.002)	(0.002)	(0.002)
Share of elementary occupations (2021)				-0.017***	-0.017***	-0.019***	-0.018***
				(0.001)	(0.001)	(0.001)	(0.001)
Best Subset							X
Observations	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.0612	.108	.149	.169	.182	.188	.191

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A17: Predictors of Percent Changes in Average Energy Bills under the EPG: Deprivation Indicators

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share in fuel poverty (2022)					0.016***	0.019***	0.018***	0.017***	0.018***	0.018***
					(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share unemployed (2021)									-0.005***	-0.004**
									(0.002)	(0.002)
Share inactive (2021)										0.002*
										(0.001)
Income deprivation score rate (2019)							0.015***	0.026***	0.033***	0.034***
							(0.002)	(0.004)	(0.005)	(0.005)
Employment deprivation score rate (2019)								-0.013***	-0.018***	-0.021***
								(0.004)	(0.004)	(0.004)
Education deprivation score (2019)			-0.008***	-0.019***	-0.029***	-0.024***	-0.030***	-0.030***	-0.031***	-0.031***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Health deprivation score (2019)						-0.012***	-0.017***	-0.015***	-0.014***	-0.014***
						(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Crime deprivation score (2019)				0.018***	0.017***	0.021***	0.017***	0.017***	0.018***	0.019***
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Housing deprivation score (2019)	-0.024***	-0.019***	-0.021***	-0.020***	-0.020***	-0.022***	-0.024***	-0.025***	-0.024***	-0.024***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Living environment deprivation score (2019)		-0.018***	-0.016***	-0.021***	-0.027***	-0.026***	-0.027***	-0.027***	-0.027***	-0.027***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Best Subset										X
Observations	6791	6791	6791	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.121	.182	.196	.238	.26	.27	.276	.278	.279	.279

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A18: Predictors of Percent Changes in Average Energy Bills under the EPG: Housing

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share social renters (2021)				-0.009***	-0.017***	-0.018***	-0.017***	-0.018***	-0.018***	-0.016***
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Share private renters (2021)					-0.015***	-0.019***	-0.020***	-0.021***	-0.021***	-0.018***
					(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Share home owners with mortgage or loan (2021)	0.022***	0.025***	0.033***	0.029***	0.019***	0.020***	0.021***	0.021***	0.021***	0.021***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Share without a second address (2021)						-0.007***	-0.018***	-0.022***	-0.022***	-0.024***
						(0.001)	(0.004)	(0.006)	(0.006)	(0.005)
Share with a second address in the UK (2021)							-0.011***	-0.015***	-0.015***	-0.016***
							(0.004)	(0.005)	(0.005)	(0.005)
Share of properties in CT band A-B (2021)								0.003	0.003	0.003
								(0.002)	(0.002)	(0.002)
Share of properties in CT band C-D (2021)		-0.010***	-0.014***	-0.016***	-0.017***	-0.018***	-0.019***	-0.017***	-0.017***	-0.017***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of properties in CT band E-H (2021)									0.000	0.000
									(.)	(.)
Share of underoccupied properties (2021)										0.005
										(0.003)
Share of overcrowded properties (2021)			0.013***	0.017***	0.024***	0.027***	0.026***	0.027***	0.027***	0.028***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Best Subset								X		
Observations	6791	6791	6791	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.103	.123	.151	.161	.184	.191	.192	.192	.192	.192

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A19: Predictors of Percent Changes in Average Energy Bills under the EPG: Income Indicators

	(1)	(2)	(3)
Total annual household income (2018)	0.011***	0.029***	0.029***
	(0.001)	(0.001)	(0.001)
House prices - 10th percentile (2021)		-0.023***	-0.024***
		(0.001)	(0.002)
House prices - 90th percentile (2021)			0.000
			(0.002)
Best Subset		X	
Observations	6791	6791	6791
Adjusted R2	.0243	.0718	.0717

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A20: Predictors of Percent Changes in Average Energy Bills under the EPG: Demographics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share born in the UK (2021)			0.020***	0.028***	0.022***	0.027***	0.043***	0.043***
			(0.001)	(0.001)	(0.001)	(0.002)	(0.005)	(0.005)
Share born in EU countries (2021)						0.006***	0.013***	0.013***
						(0.001)	(0.002)	(0.002)
Share born in non-EU european countries (2021)								0.000
								(0.001)
Share white (2021)							-0.012***	-0.012***
							(0.003)	(0.003)
Share aged 65 years and over (2021)		0.017***			0.009***	0.010***	0.011***	0.011***
		(0.001)			(0.001)	(0.001)	(0.001)	(0.001)
Share with level 3 qualification or above (2021)			0.010***	0.034***	0.033***	0.034***	0.034***	0.034***
			(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Share without qualifications (2021)				0.024***	0.024***	0.025***	0.023***	0.023***
				(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Share with 3+ household members (2021)	0.020***	0.029***	0.032***	0.038***	0.040***	0.042***	0.040***	0.040***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Best Subset							X	
Observations	6791	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.089	.133	.147	.166	.175	.176	.178	.178

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A21: Predictors of Percent Changes in Average Energy Bills under the EPG: Property Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Share of properties built pre-1900 (2021)							-0.002**
							(0.001)
Share of properties built between 1900 and 1939 (2021)		0.017***	0.018***	0.018***	0.017***	0.016***	0.016***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of properties built post-2000 (2021)	-0.022***	-0.018***	-0.015***	-0.016***	-0.015***	-0.016***	-0.017***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share of flats/maisonettes in properties (2021)			-0.014***			0.020***	0.021***
			(0.001)			(0.003)	(0.003)
Share of detached houses in properties (2021)				0.010***	0.015***	0.031***	0.032***
				(0.001)	(0.001)	(0.003)	(0.002)
Share of semi-detached houses in properties (2021)				0.011***	0.014***	0.028***	0.028***
				(0.001)	(0.001)	(0.002)	(0.002)
Share of terraced houses in properties (2021)					0.010***	0.023***	0.024***
					(0.001)	(0.002)	(0.002)
Best Subset							X
Observations	6791	6791	6791	6791	6791	6791	6791
Adjusted R2	.107	.165	.204	.22	.232	.243	.244

Notes: Table reports results from OLS regressions. Empirical models selected using best subset selection on the set of predictors using the AIC information criterion. Best subset marked by "X". Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A22: Predictors of Percent Changes in Average Energy Bills under the EPG: Blocked Variable Selection Approach

	(1)	(2)
Share of managers, directors and senior officials (2021)	0.005**	0.005**
	(0.002)	(0.002)
Share of professional occupations (2021)	-0.006**	-0.006*
	(0.002)	(0.003)
Share of associate professional, technical and service occupations (2021)	0.010***	0.010***
	(0.001)	(0.001)
Share of administrative and secretarial occupations (2021)	0.002**	0.002**
	(0.001)	(0.001)
Share of skilled trades, sales and customer service occupations (2021)	-0.017***	-0.017***
	(0.001)	(0.001)
Share of process, plant and machine operatives (2021)	0.004***	0.004***
	(0.001)	(0.001)
Share in fuel poverty (2022)	0.012***	0.012***
	(0.001)	(0.001)
Share unemployed (2021)	0.006***	0.006***
	(0.001)	(0.002)
Employment deprivation score rate (2019)	0.003	0.005
	(0.002)	(0.004)
Education deprivation score (2019)	-0.013***	-0.012***
	(0.002)	(0.002)
Health deprivation score (2019)	0.016***	0.016***
	(0.002)	(0.002)
Crime deprivation score (2019)	0.008***	0.008***
	(0.001)	(0.001)
Housing deprivation score (2019)	-0.019***	-0.019***
	(0.001)	(0.001)
Living environment deprivation score (2019)	-0.033***	-0.033***
	(0.001)	(0.001)
Share social renters (2021)	-0.032***	-0.032***
	(0.002)	(0.003)
Share private renters (2021)	-0.033***	-0.032***
	(0.002)	(0.003)
Share home owners with mortgage or loan (2021)	-0.005**	-0.004*
	(0.002)	(0.003)
Share without a second address (2021)	-0.030***	-0.031***
	(0.004)	(0.005)
Share with a second address in the UK (2021)	-0.024***	-0.025***
	(0.004)	(0.005)
Share of properties in CT band A-B (2021)	-0.012***	-0.012***
	(0.003)	(0.003)
Share of properties in CT band C-D (2021)	-0.011***	-0.011***
	(0.002)	(0.002)
Share of overcrowded properties (2021)	0.018***	0.018***
	(0.002)	(0.002)
Total annual household income (2018)	0.011***	0.011***
	(0.002)	(0.002)
House prices - 10th percentile (2021)	-0.003**	-0.003**
	(0.002)	(0.002)
Share born in the UK (2021)	0.017***	0.017***
	(0.006)	(0.006)
Share born in EU countries (2021)	0.011***	0.011***
	(0.002)	(0.002)
Share white (2021)	-0.008**	-0.009**
	(0.003)	(0.004)
Share with level 3 qualification or above (2021)	0.017***	0.017***
	(0.004)	(0.004)
Share without qualifications (2021)	0.021***	0.021***
	(0.002)	(0.003)
Share with 3+ household members (2021)	0.013***	0.013***
	(0.002)	(0.002)
Share of properties built pre-1900 (2021)	0.013***	0.013***
	(0.001)	(0.001)
Share of properties built between 1900 and 1939 (2021)	0.010***	0.010***
	(0.001)	(0.001)
Share of properties built post-2000 (2021)	-0.007***	-0.007***
	(0.001)	(0.001)
Share of flats/maisonettes in properties (2021)	0.011***	0.011***
	(0.003)	(0.003)
Share of detached houses in properties (2021)	0.009***	0.009***
	(0.002)	(0.002)
Share of semi-detached houses in properties (2021)	0.012***	0.012***
	(0.002)	(0.002)
Share of terraced houses in properties (2021)	0.018***	0.018***
	(0.002)	(0.002)
Observations	6791	6791
Adjusted R2	.61	.61

Notes: Table reports results from OLS regressions. Column 1 shows best subset across all 5 groups of variables analyzed in Tables 1 through 4. Column 2 is the full specification based on best subsets determined in Tables 1 through 4. For comparison, columns 3 through 6 re-display the optimal specifications from Tables 1 through 4. Robust standard errors are presented in parentheses, asterisks indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix to "Distributional and climate implications of policy responses to the energy crisis: Lessons from the UK"

For Online Publication

Thiemo Fetzer Ludovica Gazze Menna Bishop

A Step 1: Deriving a physical energy consumption measure for each property in the EPC data

An essential ingredient in our energy consumption calculations was the set of fuel prices faced by a given property for each type of energy consumption (space heating, water heating, and lighting). For example, while gas is the most common heating method across properties in the EPC data, many use either electricity or oil and therefore face different prices. Additional complexity follows from the range of possible tariffs used to price a household's electricity use. The prices used in EPC calculations dating back to 2013 for all possible fuel types are published by BRE.[1] We had to infer which of these had been applied to each energy consumption type for each property in order to estimate expenditures.

To decide the assignment of fuel prices, we consulted four variables from the EPC database: main heating system (`MAINHEAT_DESCRIPTION`), water heating system (`HOTWATER_DESCRIPTION`), type of fuel used to power the central heating (`MAIN_FUEL`), and electricity tariff (`ENERGY_TARIFF`). For example, if the main heating system was recorded as "boiler and radiators, mains gas", main fuel as "mains gas", hot water system as "from main system" and energy tariff as "single", a property was assigned the "mains gas" fuel price for space and water heating and the "standard tariff" price for lighting from the SAP price list. The raw data contain 9,796 unique combinations of these four variables, and so we restricted our attention to the 30 most common combinations, excluding those containing oil.[2] In total, these 30 com-

[1] Data are available here https://bregroup.com/sap/standard-assessment-procedure-sap-2012/

[2] We excluded properties using oil as there is no price cap for this fuel, which is our source of price data for gas and electricity (see the next section for details).

binations account for 85% of the sample. For the remaining 15%, we infer energy consumption using `ENERGY_CONSUMPTION_CURRENT`, a variable which estimates total energy consumption in kWh per metre squared of floor area. We scale this variable by `FLOOR_AREA` and multiply by the cost share of each energy use type to produce estimates of energy consumption for space and water heating and lighting.

We followed the SAP documentation to the best of our ability in the process of assigning fuel prices to energy consumption types, though in places the appropriate correspondence was not clear. Ambiguities also arose in interpreting how prices, which include a standing charge and price per kWh, had been applied to consumption to produce the spending estimates available in EPC data, complicating the reverse-engineering of this calculation. To account for this uncertainty, we have included a lower-bound estimate, which incorporates standing charges, as well as an upper-bound estimate, which excludes standing charges from consumption calculations. This inevitably introduces some measurement error, which we intended to tackle via spatial aggregation.

The consumption estimates we produce for water heating, space heating and lighting abstract from behavioral responses, as the underlying physical SAP model used to produce the EPC data considers three factors:

1. The physical characteristics of a property, such as build-type, insulation technology, floor area, window area, number of rooms and light fixtures
2. Time-invariant climatic factors that affect fuel consumption and are ultimately determined by property location
3. Fixed relationships between estimated use due to time-invariant estimates of likely occupation and use determined by the physical makeup of the property such as the number of bedrooms, the floor area, etc.

Consumption estimates stemming from EPC data will therefore not map one-to-one with consumption estimates which reflect actual patterns of energy consumption by residents, such as those produced from meter readings. Rather, EPC data is based on exogenous and typically fixed characteristics of the underlying buildings, a desirable feature for an econometrician. In this sense, our consumption estimates should be understood as *theoretical* as opposed to *real* consumption estimates.

B Step 2: Anchoring technically-required energy consumption measure with anonymized meter-level data

In step two, we produce a second consumption measure that incorporates anonymous micro data on energy and gas consumption from the National Energy Efficiency Data-Framework (NEED).[3] We refer to these as *real* consumption estimates as, unlike the EPC-based estimates, they reflect patterns of energy consumption behaviour by households. The sample includes four million properties and is designed to be representative of domestic properties in England and Wales. Data are available annually for years 2005-19, of which our analysis uses 2017-19. The data include estimates of energy and gas consumption which are derived from meter readings, alongside a number of property and area-level characteristics.

We use this *real* consumption data to develop a refinement of the *theoretical* consumption measure derived in Step 1. We match moments of the NEED meter-level data with moments from the EPC-derived consumption measure, in effect rescaling our theoretical consumption measure. This is possible because the NEED data include a range of property characteristics which are also present in the EPC data:

- property type (six categories)
- property age band (four categories)
- an indicator for whether gas is the main heating fuel (two categories)
- floor area band (five categories)
- a measure of the relative deprivation of the area in which a property is located, measured in 2019 (five categories each for Wales and England)
- region (nine categories for England, one for Wales).

In theory, there are $6 \times 4 \times 2 \times 5 \times 5 \times 9 = 10,800$ unique combinations of these features in England and $6 \times 4 \times 2 \times 5 \times 5 = 1,200$ unique combinations in Wales.

For each unique combination, we calculate the deciles of combined gas and electricity consumption in the NEED data, excluding combinations which contain

[3]The data can be found here `https://www.gov.uk/government/statistics/national-energy-efficiency-data-framework-need-anonymised-data-2021`.

300 properties or fewer out of the total 12 million (4 million for each of the years 2017-19).

We then replicate this exercise using the EPC consumption estimates derived in Step 1. When calculating total consumption, we take weighted averages of the upper and lower bounds for our light, water, and space energy consumption estimates, before summing over these to derive aggregate energy consumption. The weight assigned to the upper bound of each consumption estimate is 5 minus the floor area band (1-5), meaning a higher weight is assigned to the lower bound for larger properties.

Next, we match the NEED energy consumption deciles for each unique combination of property attributes to the corresponding EPC energy consumption deciles. For example, a property that is in the top decile of *theoretical consumption* (derived from EPC data) among properties with the same combination of property attributes will be assigned the top decile of *real consumption* (derived from NEED data) for properties with these same attributes. The latter provides us with a potentially more accurate representation of real consumption behaviour at the property level.

We then update our property-level estimates of theoretical consumption by multiplying by the ratio of real to theoretical energy consumption (both actual and potential) for a property's attributes and consumption decile.[4]

We then perform a second rescaling using postcode-level gas and electricity consumption data, again for the years 2017-19.[5] For each postcode, we compute the sum of median gas and electricity consumption across years. We then repeat this exercise for the theoretical consumption estimates developed in Step 1 as well as the estimates which were adjusted using NEED data. Next, we rescale the property-level theoretical and NEED-adjusted consumption estimates by multiplying by the ratio of the median postcode-level energy consumption from the postcode data to the corresponding value in the EPC data. We perform this rescaling of theoretical consumption estimates only for properties in postcodes with at least 25% coverage in the EPC data. Here, coverage is defined as the number of properties per postcode

[4]Note that energy consumption estimates for properties whose combinations of attributes included less than 300 properties were not rescaled.

[5]The electricity data can be found on `https://www.gov.uk/government/collections/sub-national-electricity-consumption-data`; the gas data is on `https://www.gov.uk/government/collections/sub-national-gas-consumption-data`.

in the EPC data relative to the number of energy meters used to form the energy consumption estimates in the postcode data.[6] We then perform this same rescaling for NEED-adjusted consumption estimates. We exclude from both rescaling exercises properties in postcodes with five or fewer properties in the EPC data or five of fewer energy meters in the postcode consumption estimates. For properties in postcodes which fail these coverage requirements, we rescale consumption using the same methods but with LSOA-level data. Here, we impose a looser restriction of 50% coverage of EPC properties in a property's LSOA.

C Step 3: Converting consumption measures to time-varying spending estimates

In our third step, we convert the *time-invariant* consumption estimates from Steps 1 and 2, measured in kWh, into *time-varying* estimates of actual spending in GBP. In practice, this is not straightforward as the energy prices faced by households, which consist of a unit price and standing charge, depend on the particular energy supply contract which they have entered into.

We are interested in four types of price scenario:

1. Energy price cap.
   The energy cap sets the maximum price that energy suppliers are allowed to charge customers, and is chosen by regulator Ofgem for gas and electricity prices to reflect the costs of supplying energy and to allow modest profits (Ofgem, 2022a). The cap has been updated every 6 months since its introduction in January 2019, but from October 2022 will be updated on a quarterly basis. The price cap was originally conceived to protect inattentive consumers from being charged unfair rates. In its early years, some energy contracts on the market were cheaper than the cap, but since the summer of 2021 the cap has been the cheapest rate available. This phenomenon is due to price increases between the time at which the price cap is set and the time at which it comes into effect (as of October 2022, this gap has been shortened from two

[6] The postcode-level data includes the number of meters used to form the estimates of median gas and electricity consumption respectively, and we use the highest of these two figures.

months to 25 working days) (Ofgem, 2022b). As such, the cap has been a more accurate reflection of the prices faced by households in recent months than in previous years. Our study incorporates price cap values from October 2021 and October 2022.

2. Energy Price Guarantee.
   In September 2022, the UK government announced the Energy Price Guarantee programme as a response to the ongoing energy crisis. The EPG reduces the maximum per unit rate below the level of the October 2022 price cap in an attempt to limit the average household energy bill to around £2,500. As discussed in Fetzer (2022), the standing charge is maintained at the level of the October 2022 price cap.

3. Historical average energy prices.
   The Department for Business, Energy and Industrial Strategy (BEIS) publishes data on average gas and electricity prices for 2010-2021.[7] These data are particularly valuable for estimating energy bills pre-2019, when the energy price cap had not yet been introduced.

4. Two-tier tariff.
   This is an alternative policy proposal to the energy-price guarantee that is discussed in more detail in Fetzer (2022). It consists of a two-tier tariff wherein the standing charge would be fixed at the level of the October 2021 price cap, as would unit prices for the first 9,500 kWh of natural gas consumption and the first 2,500 kWh of electricity consumption. As 50% of UK households consume less than 12,100 kWh of natural gas and 2,900 kWh of electricity, this would drastically limit energy price increases for the bulk of households.[8] The second tier of the energy tariff would be set at steeper levels which could be aligned with the EPG. For example, a second tier unit price of 20 pence per kWh for natural gas and 60 pence per kWh for electricity, together with the first tier described above, would have a similar cost to the government as the

[7] Data are available here `https://www.gov.uk/government/statistical-data-sets/annual-domestic-energy-price-statistics`

[8] See `https://www.gov.uk/government/statistics/national-energy-efficiency-data-framework-need-consumption-data-tables-2021`.

EPG. This would offer much more targeted support without undermining the incentive to save energy created by higher unit prices.

Energy prices consist of a standing charge and unit rate which differ according to region, payment method (for example, direct debit versus pre-paid), fuel type (electricity versus gas), and electricity metering arrangement (whether the electricity tariff varies by time of use). We use information on these dimensions from the EPC data to assign the appropriate price to each property. In the absence of data on payment method, we assume direct debit for all households.[9]

We then estimate energy expenditure for a given property and energy use (space heating, water heating, or lighting) as follows:

$$spend_{ierfmt} = cons_{ierfm} \times price_{rfmt} + charge_{rfmt}$$

Here, $cons_{ierfm}$ is the consumption estimate for energy use type e by property i in region r with fuel f and metering arrangement m, as calculateds in Steps 1 and 2. $price_{rfmt}$ and $charge_{rfmt}$ are the unit price and standing charge at which the cap has been set for their region, metering arrangement and fuel in period t (assuming payment by direct debit).

In essence, this spending calculation converts the consumption estimate in physical energy units, which reflects the physical characteristics of a property, back into energy cost estimates that are, in turn, exogenous to household-specific choices with respect to their energy supply contract. This data structure is also ideal for merging in different price scenarios to forecast their likely impact on household bills across different groups and regions within the UK.

Most households are on one-year fixed term contracts at the energy supply contracts.

[9]Direct debit is the most popular payment method (Ofgem, 2019).

D Step 4: Energy efficiency upgrade recommendations and its costing

Lastly, we also examine the specific energy efficiency upgrade investments recommended in the EPCs. We were not able to confirm how the costing of these recommendations is done. We thereby convert the estimates of the costs for specific measures, which typically include an upper- and a lower-bound range, to a further upper and lower bound based on an inflation rate estimate.

To do so, we construct a version of the cost estimate that is expressed in current GBP which effectively uprates the upper- and lower-bound cost estimate by what we judged to be the most appropriate inflation rate from the lodgement date (the date the EPC was drawn up) to the current date.

E Illustrating the goodness-of-fit of estimated consumption

This section describes how our derived property-level consumption measures fit actual energy consumption available at the MSOA level and presents a validation exercise. Appendix Figure A4 shows the goodness-of-fit of the underlying measures, highlighting that the crude E_p^{EPC} measure does a decent job at fitting the data but harnessing data on energy consumption improves the goodness-of-fit substantially.[10] Appendix Figure A5 further emphasises that rescaling improves goodness-of-fit across each of the three moments that we consider.

We next consider an additional out-of-sample validation exercise comparing empirical moments that were not used in the training step. For some local authorities, we have data that provide pairwise measures of both the mean and median electricity and gas consumption by district and by property-type and floor-area band.[11] We did not use these data in the rescaling as it is coarser than postcode-level data. We can therefore use these measures in an out-of-sample validation exercise.

[10] Appendix Figures A6 and A7 present corresponding scatterplots for estimates of average and total energy consumption at the MSOA-level respectively.

[11] Local authority table, England and Wales, `https://www.gov.uk/government/statistics/national-energy-efficiency-data-framework-need-consumption-data-tables-2021`.

We do so by estimating

$$E_{d,c}^{BEIS} = \alpha + \beta \times E_{d,c}^{ensemble} + x_{d,p} \times \nu + \epsilon_d$$

where $E_{d,c}^{BEIS}$ stands for the median or mean energy consumption of a property with a characteristic c in district d, that is, these are derived moments in the actual energy consumption data. We construct the corresponding moments in aggregated form, either the median or mean, at the district by property characteristic based on the property-level *ensemble* measure: $E_{d,c}^{ensemble}$. We control for property-level controls and district fixed effects, $x_{d,p}$.

Our attention will be on the estimated coefficient β. In the regressions that exclude other control variables or shifters, this coefficient should be close to one if there was a near one-to-one mapping of the EPC-derived consumption measures and the actually observed consumption data. A second focus will be on the combined R^2 of these regressions. If this R^2 is close to one, it would indicate that, on average, our approach to measure hypothetical consumption captures the variation in actual consumption quite well.

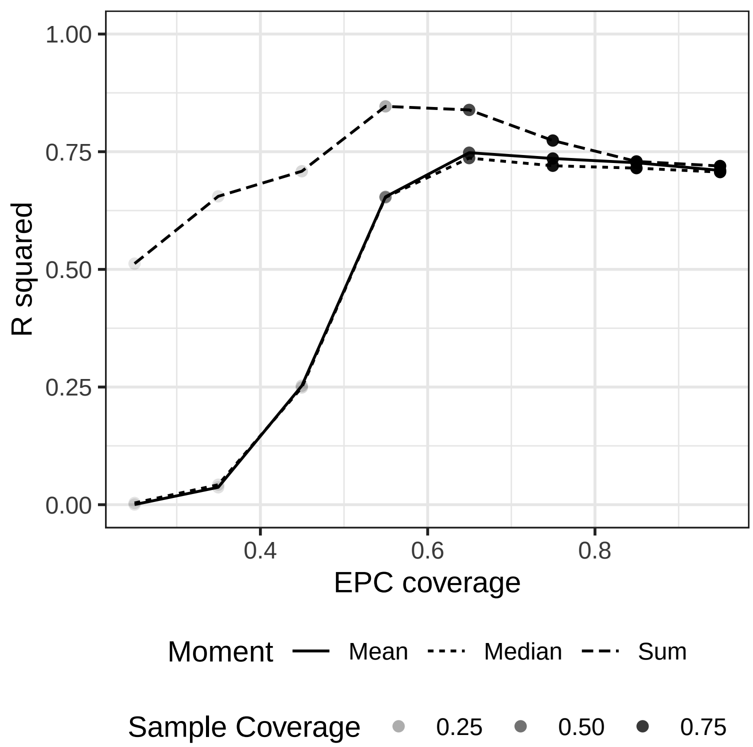
Lastly, we are interested in whether, after absorbing district fixed effects and property characteristics included in the vector $x_{d,p}$, our EPC-derived consumption measure $E_{d,c}^{j}$ carries signal over and above area- and property characteristic-specific idiosyncrasies. In other words, this exercise tests whether our two-way rescaling approach achieves its goal.

Unconditional fit. In Figure A8 we present the simple unconditional scatterplot of the two datasets. On the horizontal axis we plot the EPC-derived median energy ensemble predicted energy consumption at the district by floor area combination level in Panel A, and the district by property type level in Panel B, $E_{d,c}^{ensemble}$. The vertical axis plots the actual observed median consumption for 2019, $E_{d,c}^{BEIS}$.

We observe a tight fit even in the unconditional regressions. We next explore this validation more systematically.

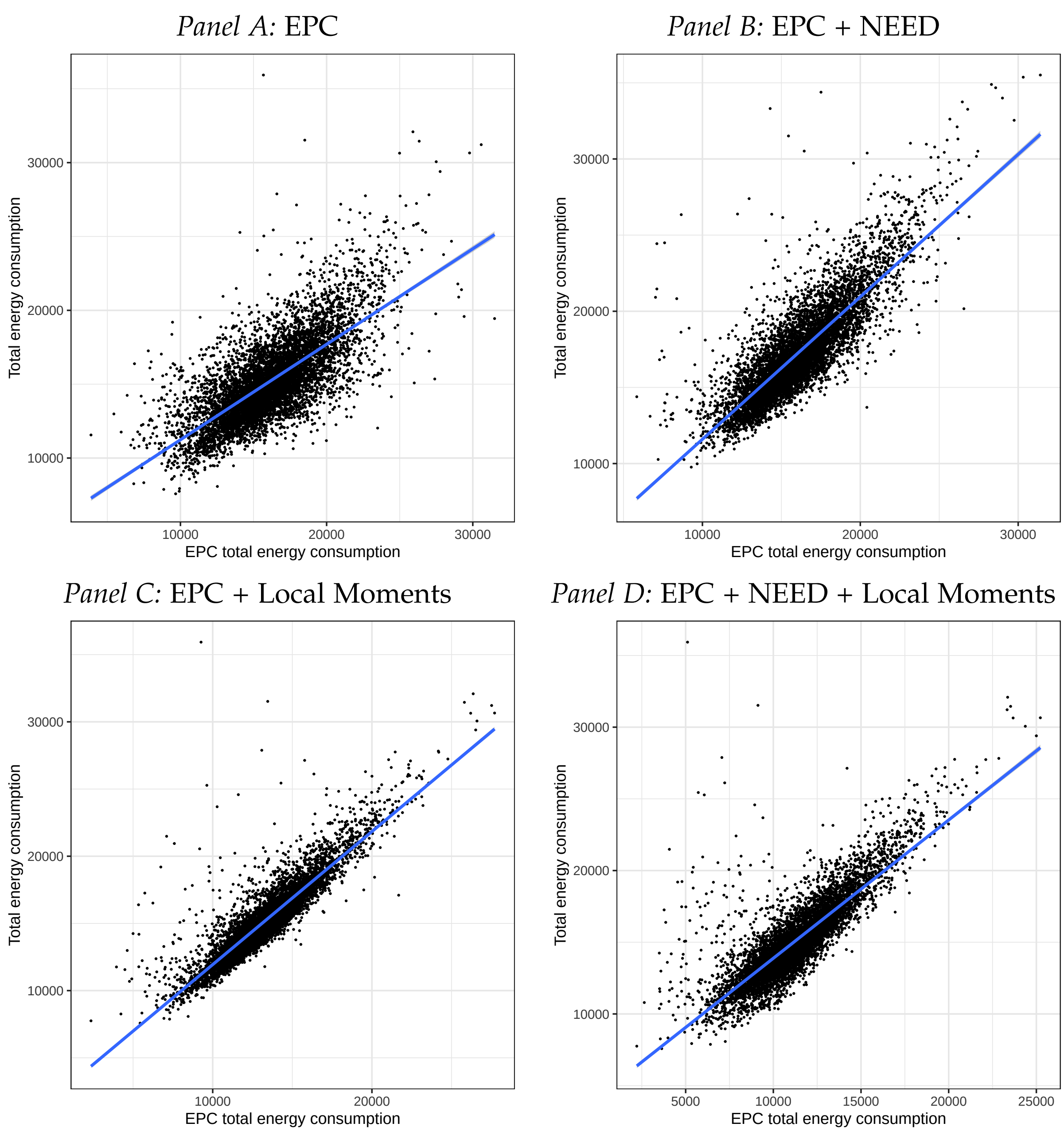
Conditional fit. We first compare the BEIS empirical moments of the median and the mean electricity and gas consumption with the five measures based on the EPC measures. Table A23 presents these results, adding control variables across panels. Based on the patterns in this table, we conclude that our empirical approach calibrates the EPC-derived data to actual consumption data well, which allows us to provide a richer view of the likely impact of the energy price shock. A similar picture emerges when studying the district-by-property-type empirical moments presented in Appendix Table A24.

Figure A3: Correlation between moments of derived consumption and moments from actual consumption data



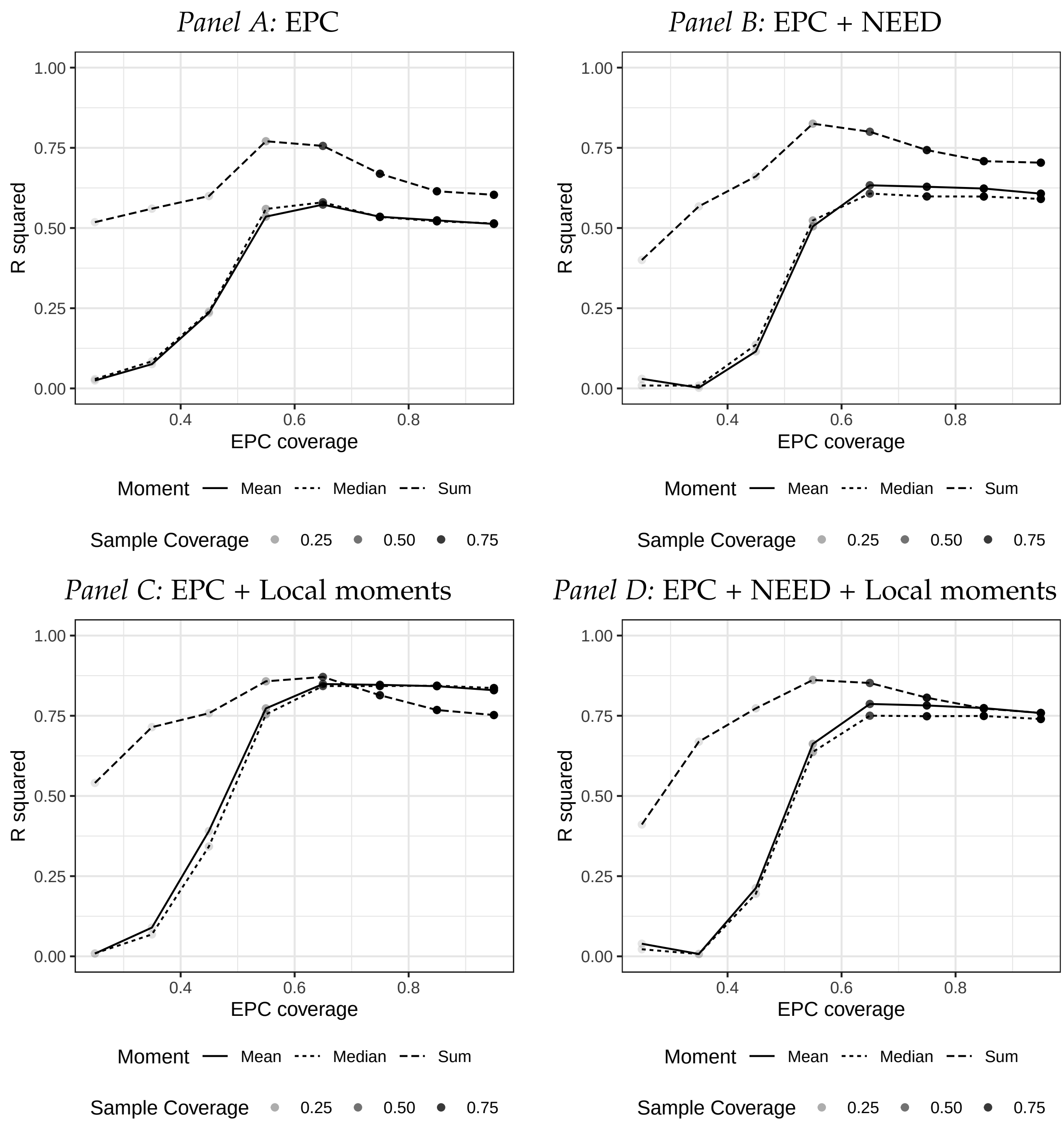
Notes: Figures plot the R^2 that is obtained from validating the derived ensemble consumption measure and three moments: the total, mean, and median consumption against actual consumption data that is published from gas and electricity meters across the country. We compare the goodness-of-fit of each derived moment against the corresponding moment from subnational statistics. The horizontal axis captures the ratio of the number of EPC properties against the population of properties in an area based on council tax data. A value of 0.4 on the axis implies that the estimating sample includes data from all MSOAs that have at most 40% of their building stock captured in the EPC data.

Figure A4: Median property-level energy consumption measures at the MSOA-level compared with median imputed energy consumption measures from EPC-NEED data



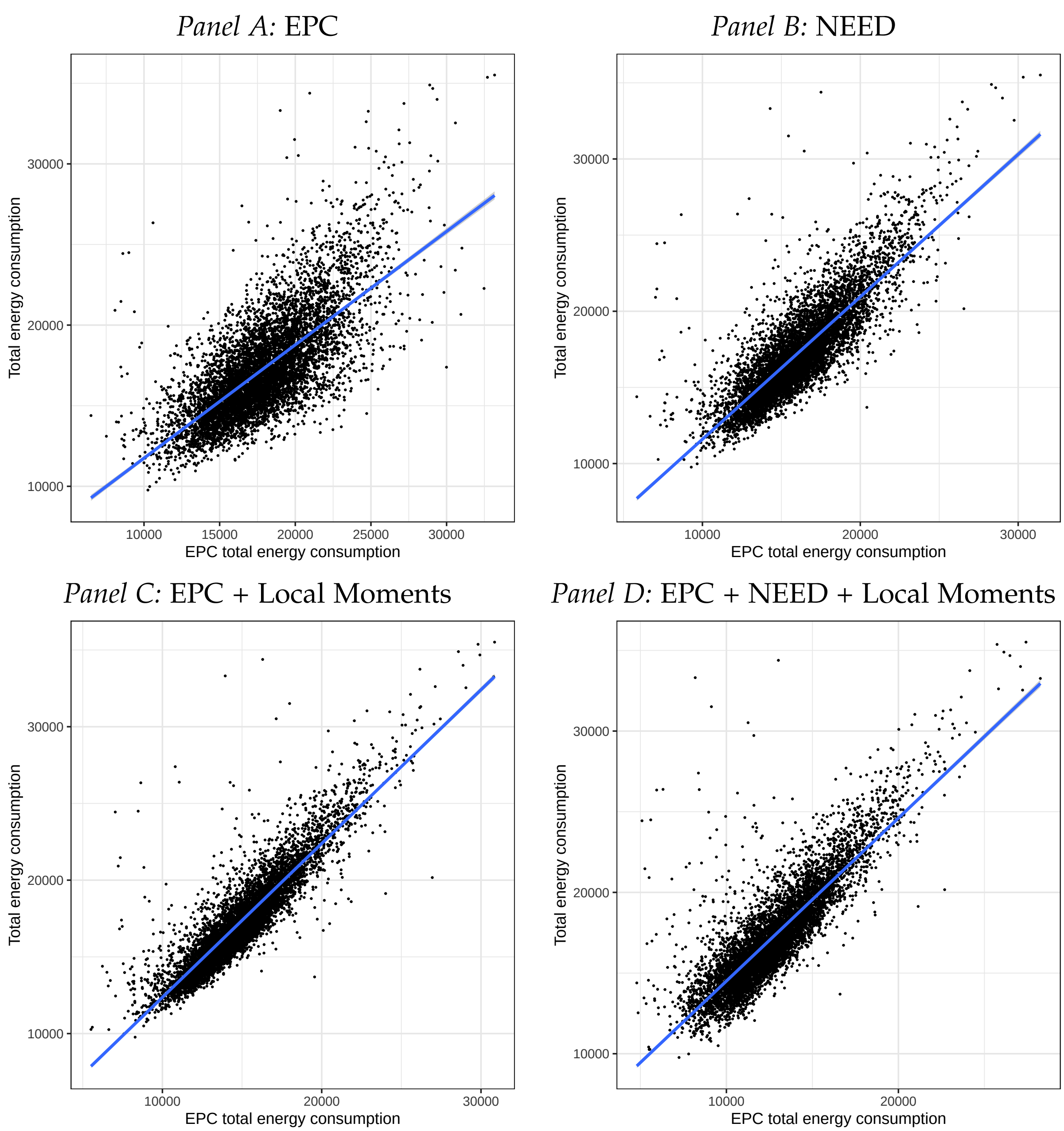
Notes: Figures provide a scatterplot of estimates of the median energy consumption per meter from published data at the MSOA-level (for metered electricity and gas only) on the vertical axis and the median of various imputed energy consumption measures that leverage different data on the horizontal axis. Panel A provides the implied consumption estimates from the EPC data as is. Panel B augments the EPC data with a matching-of-moments approach based on anonymized individual level meter reading data collected under the NEED framework. Panel C uses the EPC raw energy consumption estimates and augments it with matched granular area-specific moments. Panel D is the final measure that combines the EPC raw data, the property-specific moment-matching and the local area specific moment matching.

Figure A5: Correlation between moments of derived consumption proxy measures and moments from actual consumption data



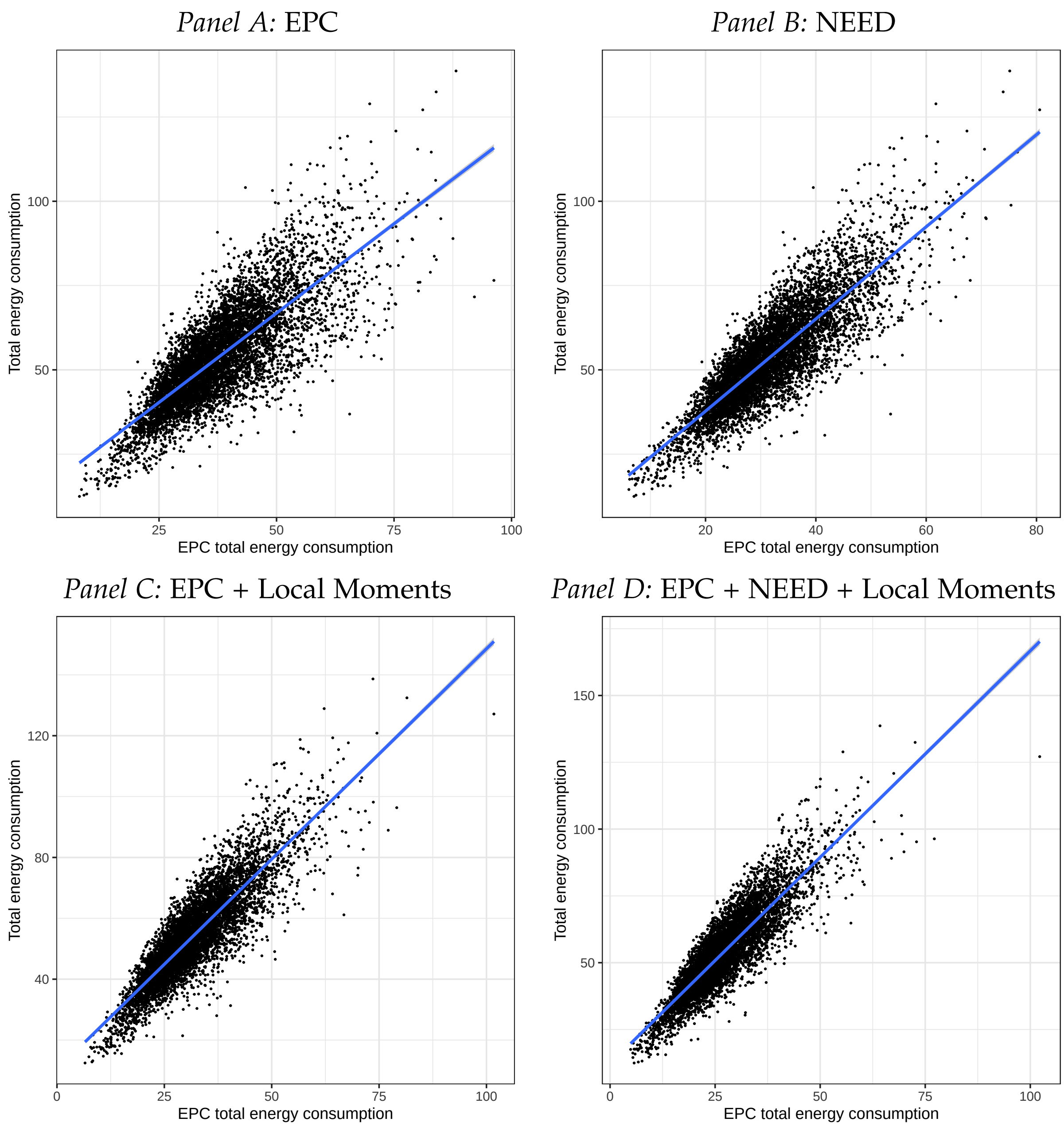
Notes: Figures plot the R^2 that is obtained from validating the derived implied consumption measures and three moments: the total consumption, the mean, and median consumption against actual consumption data that is published from gas and electricity meters across the country. For the four different derived measures, we compare the goodness-of-fit of the three moments against the corresponding moment from subnational statistics. The horizontal axis captures the ratio of the number of EPC properties against the population of properties in an area based on council tax data. A value of 0.4 on the axis implies that the estimating sample includes data from all MSOAs that have at most 40% of their building stock captured in the EPC data. We note that the goodness-of-fit remains stable across each of the moments when the estimating sample includes MSOAs with an EPC coverage of up to 60%.

Figure A6: Average property-level energy consumption measures at the MSOA-level compared with imputed energy consumption measures from EPC-NEED data



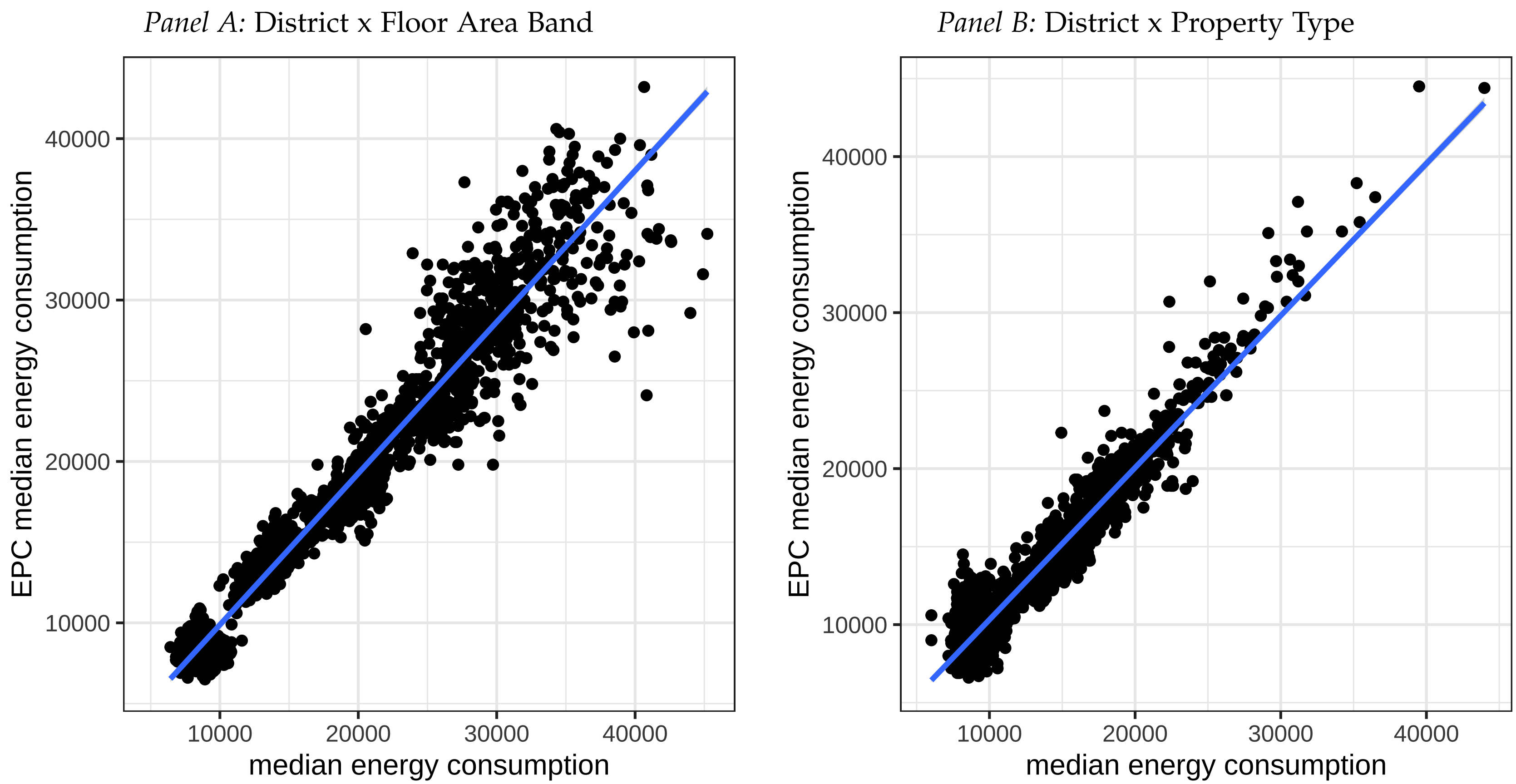
Notes: Figures provide a scatterplot of the mean energy consumption per meter estimates from published data at the MSOA level (for metered electricity and gas only) on the vertical axis and the average of various imputed energy consumption measures that leverage different data on the horizontal axis. Panel A provides the implied consumption estimates from the EPC data as is. Panel B augments the EPC data with a matching-of-moments approach based on anonymized individual level meter reading data collected under the NEED framework. Panel C uses the EPC raw energy consumption estimates and augments it with matched granular area-specific moments. Panel D is the final measure that combines the EPC raw data, the property-specific moment-matching and the local area specific moment matching.

Figure A7: Total property-level energy consumption measures at the MSOA-level compared with imputed energy consumption measures from EPC-NEED data



Notes: Figures provide a scatterplot of the total energy consumption per meter estimates from published data at the MSOA level (for metered electricity and gas only) on the vertical axis and the average of various imputed energy consumption measures that leverage different data on the horizontal axis. Panel A provides the implied consumption estimates from the EPC data as is. Panel B augments the EPC data with a matching-of-moments approach based on anonymized individual level meter reading data collected under the NEED framework. Panel C uses the EPC raw energy consumption estimates and augments it with matched granular area-specific moments. Panel D is the final measure that combines the EPC raw data, the property-specific moment-matching and the local area specific moment matching.

Figure A8: Unconditional raw scatter plot district-by-floor-area or district-by-property-type median energy consumption data vis-a-vis our EPC-derived ensemble consumption estimate



Notes: Figures plot a raw scatterplot of the median district-level energy consumption by floor area in Panel A or district-level energy consumption by property type in Panel B against the corresponding median constructed from our EPC-derived ensemble measure. The corresponding regression is presented in column 10 of Table A23 and Table A24 respectively.

Table A23: Comparison of district-by-floor-area BEIS-reported average and median electricity and gas consumption vis-a-vis corresponding EPC-derived and rescaled proxy measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	EPC		*NEED*		*EPC + Local*		*EPC + NEED + Local*		*Average*	
Panel A: No controls										
Derived energy consumption proxy	0.859***	0.857***	0.924***	0.929***	0.977***	1.067***	1.035***	1.129***	0.919***	0.938***
	(0.011)	(0.011)	(0.015)	(0.014)	(0.007)	(0.005)	(0.010)	(0.009)	(0.010)	(0.009)
R2	0.884	0.899	0.884	0.895	0.957	0.966	0.944	0.950	0.927	0.933
Observations	1650	1650	1650	1650	1650	1650	1650	1650	1650	1650
Panel B: Floor Area Band FE										
Derived energy consumption proxy	0.192***	0.298***	0.185***	0.298***	0.580***	0.806***	0.525***	0.714***	0.357***	0.484***
	(0.029)	(0.031)	(0.039)	(0.041)	(0.039)	(0.037)	(0.047)	(0.044)	(0.042)	(0.042)
R2	0.953	0.947	0.952	0.946	0.971	0.973	0.965	0.964	0.958	0.955
Observations	1650	1650	1650	1650	1650	1650	1650	1650	1650	1650
Panel C: District FE										
Derived energy consumption proxy	0.901***	0.890***	0.960***	0.955***	0.986***	1.071***	1.046***	1.137***	0.939***	0.952***
	(0.009)	(0.009)	(0.012)	(0.012)	(0.007)	(0.006)	(0.009)	(0.008)	(0.008)	(0.008)
R2	0.943	0.949	0.929	0.931	0.972	0.978	0.959	0.963	0.959	0.961
Observations	1650	1650	1650	1650	1650	1650	1650	1650	1650	1650
Panel D: District FE and Floor Area Band FE										
Derived energy consumption proxy	0.196***	0.286***	0.109***	0.185***	0.399***	0.606***	0.278***	0.448***	0.265***	0.367***
	(0.024)	(0.026)	(0.027)	(0.029)	(0.026)	(0.028)	(0.038)	(0.042)	(0.031)	(0.034)
R2	0.986	0.982	0.984	0.979	0.989	0.988	0.986	0.983	0.986	0.983
Observations	1650	1650	1650	1650	1650	1650	1650	1650	1650	1650
Moment:	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median

Notes: Table presents regression results comparing the district-level average electricity and gas consumption average from BEIS micro data by floor-area type with the measures that we constructed as part of our proxy variables. Across the panels more control variables are included. The goodness-of-fit improves and even after controlling for district- and floor-area band, the district specific measures carry strong signal. The observation that the coefficient is near one suggests that the calibration exercise is not producing a biased estimate of the population mean despite the data being from a subsample of the population of properties.

Table A24: Comparison of district-by-property-type BEIS-reported average and median electricity and gas consumption vis-a-vis corresponding EPC-derived and rescaled proxy measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	EPC		*NEED*		*EPC + Local*		*EPC + NEED + Local*		*Average*	
Panel A: No controls										
Derived energy consumption proxy	0.789***	0.859***	0.903***	0.975***	0.949***	1.071***	0.993***	1.121***	0.888***	0.975***
	(0.013)	(0.011)	(0.013)	(0.010)	(0.008)	(0.011)	(0.010)	(0.012)	(0.010)	(0.009)
R2	0.756	0.850	0.823	0.896	0.887	0.942	0.881	0.927	0.845	0.918
Observations	2303	2303	2303	2303	2303	2303	2303	2303	2303	2303
Panel B: Property Type FE										
Derived energy consumption proxy	0.697***	0.782***	0.810***	0.892***	0.989***	1.111***	0.969***	1.102***	0.891***	0.968***
	(0.038)	(0.037)	(0.032)	(0.029)	(0.017)	(0.028)	(0.022)	(0.032)	(0.024)	(0.024)
R2	0.822	0.888	0.846	0.906	0.899	0.947	0.882	0.928	0.871	0.929
Observations	2303	2303	2303	2303	2303	2303	2303	2303	2303	2303
Panel C: District FE										
Derived energy consumption proxy	0.807***	0.863***	0.930***	0.994***	0.937***	1.057***	1.004***	1.130***	0.888***	0.969***
	(0.009)	(0.008)	(0.011)	(0.010)	(0.007)	(0.009)	(0.010)	(0.011)	(0.008)	(0.008)
R2	0.807	0.891	0.853	0.920	0.888	0.947	0.881	0.932	0.863	0.933
Observations	2303	2303	2303	2303	2303	2303	2303	2303	2303	2303
Panel D: District FE and Property Type FE										
Derived energy consumption proxy	0.707***	0.729***	0.855***	0.893***	0.951***	1.056***	1.020***	1.150***	0.893***	0.922***
	(0.033)	(0.039)	(0.043)	(0.047)	(0.024)	(0.038)	(0.040)	(0.053)	(0.027)	(0.035)
R2	0.883	0.936	0.876	0.931	0.902	0.952	0.883	0.933	0.894	0.947
Observations	2303	2303	2303	2303	2303	2303	2303	2303	2303	2303
Moment:	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median

Notes: Table presents regression results comparing the district-level average electricity and gas consumption average from BEIS micro data by property type (detached, semi-detached, (mid/end) terraced, flat and/or bungalow) with the measures that we constructed as part of our proxy variables.

F Bounding the what, who, and how

It is inherently challenging to separate the drivers of energy consumption. Naturally, there is an interaction across at least three factors:

$$E_{i,p} = f(\text{What}_p, \text{Who}_{i,p}, \text{How}_{i,p})$$

We leverage anonymized meter-reading data from England and Wales at the property-level to bound the extent to which we can explain variation in energy use between the What_p. In the NEED anonymized microdata we observe a range of property characteristics that could drive variation in energy consumption.[12] We characterise the extent to which we can capture variation in the observed energy consumption data across properties (or households) saturating simple linear regression specifications of the form

$$E_{i,p,t} = x_{i,p,t} \times \beta + \epsilon_{i,p,t}$$

The features in $x_{i,p,t}$ include:

- Property characteristics: property type (six categories) such as detached, semi-detached, or flat; property age band (four bands) capturing the date range when a property was built; an indicator for whether gas is the main heating fuel; floor area bins (five categories) ranging from less than 50 square meters to over 200 square meters. Further, we also have measures capturing whether a property has had some energy efficiency measures such as cavity wall insulation or loft insulation installed.

- Socio-economic characteristics: quintiles of the English and Welsh indices of Multiple Deprivation (IMD) from 2019 and council tax bands. That is, for every property, we know the region (10 regions make up England and Wales) and whether a property falls into a region in a specific quintile of the English- or Welsh deprivation ranking.

[12]The data are a stratified random sample from the population of properties. Unfortunately, BEIS does not make the sampling weights available for each strata, which means we can not correct for the respective under- and oversampling. We have requested this information but are still awaiting a response.

In addition, we have a property identifier which will serve as a *property fixed effect* in some specification as the most demanding, but also least informative, way of trying to absorb property- and time-invariant resident-specific observable and unobservable characteristics.

To allow for potential non-linear interactions between different property characteristics driving energy consumption such as an interaction between floor area and property age, we construct a measure that captures the unique combinations of each of the property characteristics. That is, each unique combination of property characteristics is identifying an own *group* which we refer to as `Property`. There are 9,846 unique combinations in the data of these characteristics.

We follow the same procedure for the socioeconomic indices to exploit typical patterns of socio-economic segregation in residential choice. As with property characteristics, we combine these into a group variable that captures all unique combinations that exist in the data. We refer to this as `Socioeconomics`.

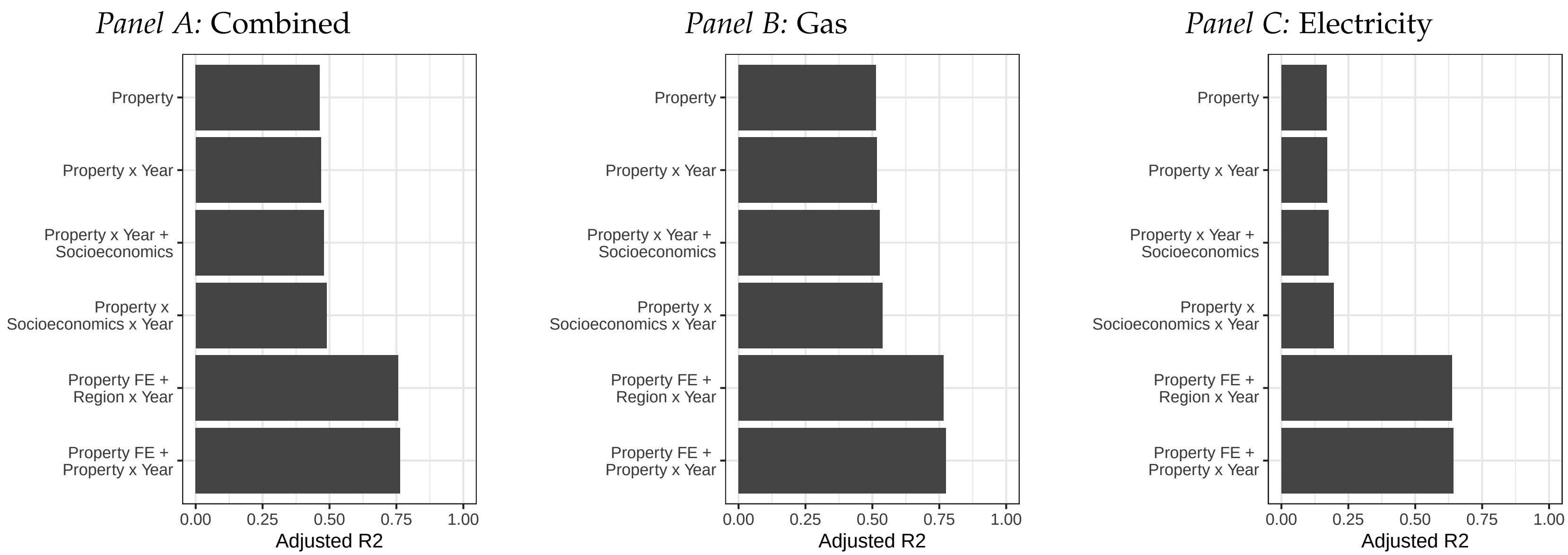
Lastly, we interact each of these variables with year fixed effects to allow for non-linear interactions of property characteristics and year-on-year unobservable shocks.

Results. We present the results from this characterisation exercise by plotting the estimated *adjusted* R^2 in Figure A9 showing both combined gas and electricity, along with gas and electricity consumption separately. We note that property and socio-economic characteristics can, at most, capture 50% of the variation in energy consumption. In particular, electricity consumption appears much more idiosyncratic compared to natural gas consumption. This finding is not surprising given that demand for natural gas is predominantly driven by space-heating and hot-water generation which do not vary much with household composition and tastes compared to electricity consumption. We note that the adjusted R^2 can reach up to around 75% in the specifications with property fixed-effects.

Interpretation. The results of this characterisation exercise suggest that property characteristics alone cannot explain much of the variation in energy consumption. At most, characteristics can explain around 50% of the variation in residential energy use. Moreover, the maximal goodness-of-fit attainable appears to be bounded around 75%, obtained when we control for property fixed-effects, which may capture some of the underlying unobservable socio-economics factors (who lives there) along with behavioural factors (how do they live). This unexplained variation could generate at least part of the difference between the engineering estimates of the benefits of energy efficiency investments and smart technologies and estimates based on actual energy use (Brandon et al., 2022).

Interestingly, our validation exercise for the property-level energy consumption measure we constructed produces a goodness-of-fit vis-a-vis statistical moments such as the mean and in particular, the median, that also achieves an adjusted R^2 of around 75%. This provides us with further confidence that our energy consumption measures can do a good job at picking up variation in the data.

Figure A9: Decomposition of variance in the anonymized individual property-level energy consumption data documenting to what extent different features can characterise the variation in energy consumption



Notes: Figures plot out the adjusted R^2 obtained from regressing combined, gas, and electricity anonymized property-level consumption data against a set of features.

G Data used for correlational analysis

The following covariates were sourced from the 2021 UK census:

Category	Covariates
Demographics	Highest qualification obtained, ethnicity, county and country of birth, age, household size, occupation
Deprivation	Unemployment, inactivity
Housing	Tenure, second homes, council tax band, occupancy
Property characteristics	dwelling type, dwelling age

These data were supplemented by the following variables:

Household income. Model-based income estimates at the MSOA-level are produced by the Office of National Statistics (ONS).[13] Our analysis used estimates of average total annual income for the year 2018.

Fuel poverty. Annual statistics on the number of individuals in fuel poverty at the LSOA-level are produced by the Department for Business, Energy & Industrial Strategy (BEIS).[14]. These adopt the Low Income Low Energy Efficiency (LILEE) metric of fuel poverty, which considers householda fuel poor if they live in an energy inefficient property and have disposable income below the poverty line. Our analysis uses figures for the year 2022. These were aggregated from Lower Layer Super Output Area (LSOA) level to the MSOA level using population estimates.

Median property prices. We compute price per square foot in each MSOA in 2021 using open data from the HM Land Registry and compute relevant percentiles.

Indices of Deprivation (IoD) . English Indices of Deprivation (IoD) are published by the Department for Levelling Up, Housing & Communities.[15] These are relative measures of deprivation which incorporate the seven following domains: income;

[13]Data are avaliable here `https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/datasets/smallareaincomeestimatesformiddlelayersuperoutputareasenglandandwales`

[14]Data are avaliable here `https://www.gov.uk/government/collections/fuel-poverty-sub-regional-statistics`

[15]Data are available here `https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019`

employment; health deprivation and disability (in our analysis, we refer to this as health); education, skills and training (education); crime; barriers to housing and services (housing and services) and living environment. Our analysis uses rankings along these dimensions for each LSOA for the year 2019. These were aggregated to the MSOA-level using population estimates.