



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/245700/>

Version: Published Version

Article:

D'Amico, Bernardino, Pomponi, Francesco, Arehart, Jay H. et al. (2026) Who cuts emissions, who turns up the heat? Causal machine learning estimates of energy efficiency interventions. Energy and buildings. 116613. ISSN: 0378-7788

<https://doi.org/10.1016/j.enbuild.2025.116613>

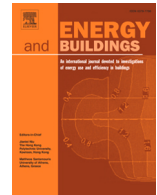
Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



Who cuts emissions, who turns up the heat? Causal machine learning estimates of energy efficiency interventions

Bernardino D'Amico^{a,b,*}, Francesco Pomponi^c, Jay H. Arehart^d, Lina Khaddour^{a,b}

^a Digital Built Environment Group (DiBEG), Edinburgh Napier University, Edinburgh, EH10 5DT, UK

^b School of Computing, Engineering and the Built Environment (SCEBE), Edinburgh Napier University, Edinburgh, EH10 5DT, UK

^c York School of Architecture, University of York, York, YO10 5DD, UK

^d Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, Boulder, 80309-0428, Colorado, USA

ARTICLE INFO

Keywords:

Energy efficiency

Fuel poverty

Causal machine learning

Do-calculus

ABSTRACT

Reducing domestic energy demand is central to climate mitigation and fuel poverty strategies, yet the impact of energy efficiency interventions is highly heterogeneous. Using a causal machine learning model trained on nationally representative data of the English housing stock, we estimate average and conditional treatment effects of wall insulation on gas consumption, focusing on distributional effects across energy burden subgroups. While interventions reduce gas demand on average ($\approx 19\%$), low energy burden groups achieve substantial savings, whereas those experiencing high energy burdens see little to no reduction. This pattern reflects a behaviourally-driven mechanism: households constrained by high costs-to-income ratios ($>10\%$) reallocate savings toward improved thermal comfort rather than lowering consumption. Far from wasteful, such responses represent rational adjustments in contexts of prior deprivation, with potential co-benefits for health and well-being. These findings call for a broader evaluation framework that accounts for both climate impacts and the equity implications of domestic energy policy.

1. Introduction

Domestic energy use accounts for a substantial portion of global greenhouse gas (GHG) emissions, making its reduction a cornerstone of climate mitigation policies worldwide [1,2]. In pursuit of decarbonisation, governments and international authorities have mandated laws, standards, subsidies and several other programmes and incentives to improve on energy efficiency of the domestic building stock [3–5]. As such, energy efficiency remains a critical pillar in these efforts for the foreseeable future, both as a means of reducing GHG emissions, and as a cost-effective strategy to accelerate progress toward national and international net-zero targets set for mid-century milestones [6–8].

Improving energy efficiency in residential buildings is widely perceived as a particularly attractive climate policy goal because it promises to deliver two key objectives at once: lower GHG emissions and improve household wellbeing [9,10]. This latter promise rests on the expectation that efficiency upgrades make it easier and more affordable to maintain adequate indoor temperatures, thereby reducing exposure to cold, damp, and fuel-related financial strain—all of which have been closely linked to respiratory illness [11], cardiovascular conditions [12], anxiety and depression [13] and more generally identified as determin-

ing factor of excess winter deaths [14,15]. This situation is commonly referred to as *fuel poverty*, defined as a household's inability to afford adequate energy services to maintain a healthy and comfortable living environment.

As such, policy co-benefits expected from energy efficiency interventions [16] are further reinforced by the broader health advantages of reduced fossil fuel use, including improved air quality and reduced pressure on public healthcare.

Existing research has shown however that domestic energy efficiency interventions often lead to highly heterogeneous cross-population reductions in energy demand [17,18]. In many cases, the actual energy savings fall short of the projections made by engineering evaluations based on building physics models—a phenomenon commonly referred to as the *rebound* effect [19]. While the rebound effect was initially dismissed as technical anomaly, it has been increasingly attributed to behavioural factors [20].

Notably, socioeconomic factors (particularly income) appear to play a central role in shaping post-retrofit energy behaviour [17], with higher-income households tending to achieve stronger and more sustained reductions in energy consumption following efficiency upgrades. In contrast, lower-income households often exhibit considerably

* Corresponding author.

E-mail addresses: b.damico@napier.ac.uk (B. D'Amico), francesco.pomponi@york.ac.uk (F. Pomponi), jay.arehart@colorado.edu (J.H. Arehart), l.khaddour@napier.ac.uk (L. Khaddour).

<https://doi.org/10.1016/j.enbuild.2025.116613>

Received 11 August 2025; Received in revised form 1 October 2025; Accepted 18 October 2025

Available online 24 October 2025

0378-7788/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

weaker, and in some cases negligible, reductions in energy use [18] post intervention.

While behavioural factors undoubtedly influence how households respond to energy efficiency interventions, a robust causal explanation must also consider the extent to which these behaviours are shaped by the intervention itself. A central mechanism in this regard is the *prebound* effect [21], where pre-intervention energy use is lower than engineering model estimates due to suppressed heating behaviours, as a result of affordability constraints. This effect is particularly pronounced among fuel-poor households, who may limit heating to unsafe or unhealthy levels in order to manage high energy costs [22,23]. For instance, recent ONS data report that a substantial proportion of UK households refrain from adequately heating their homes even in cold conditions; as of late 2022, more than 6 out of 10 (63 %) adults indicated they were consuming less gas and electricity due to rising living costs, with over 9 out of 10 (96 %) of these individuals specifically reducing their heating frequency [24].

In such cases, it is plausible that energy efficiency upgrades do not eliminate 'excess' energy use but instead enable a shift toward more adequate and healthy indoor conditions. While the prebound effect has been documented in observational studies, robust quantification of its causal impact across heterogeneous households remains limited. Our study contributes by estimating these causal effects using a causal machine learning framework.

Building on previous work [17], we adopt the hypothesis that when a dwelling undergoes an energy efficiency upgrade which, in addition to reducing required energy use, also lowers associated energy costs, this may induce behavioural changes to household heating patterns. Specifically, the hypothesis implies that exactly the same intervention alleviates energy burden (i.e., the proportion of income spent on energy bills) much more significantly, in relative terms, for lower-income households. As a result, these households are more likely to reallocate the meaningful financial relief gained from efficiency improvements toward increased thermal comfort rather than maximising energy savings, given that their heating behaviour was previously constrained to suboptimal comfort levels [22] due to high energy burden pressures. As such, this behaviourally-driven energy rebound is consistent with rational household responses under prior deprivation, highlighting the need to account for heterogeneity in evaluating intervention outcomes.

Previous attempts to infer the causal impact of energy efficiency interventions on household energy use rely on observational data sourced from the National Energy Efficiency Data (NEED) framework and employ econometric techniques such as difference-in-differences and statistical matching [17,18]. The NEED dataset is particularly valuable for its large-scale, longitudinal records of meter-level energy consumption. However, it lacks detailed socioeconomic data (most notably, income) at household-unit level. Furthermore, it covers records of energy efficiency interventions carried out through government schemes, to which households (or their landlord, social or private) have voluntarily adhered to. As a result, treatment assignment in the NEED sample dataset is subject to potential self-selection bias [25] i.e. households that choose to participate in retrofit programmes may differ systematically from non-participants in ways that also affect their energy consumption patterns, thus limiting the generalisability and internal validity of causal estimates based solely on this dataset.

Building on established findings regarding prebound and rebound effects, our study introduces two key methodological and data-driven innovations. First, we develop a causal graphical model [26] —a technique within the broader field of so-called causal machine learning (ML), whose aim is to estimate causal effects rather than mere predictions from observed data. By using a (causal) graphical model approach we are able to explicitly formalise the assumptions underlying our causal effect identification strategy, hence obtaining estimations of both average treatment effects (ATE) and conditional average treatment effects (CATE). This approach offers greater transparency than conventional

matching or difference-in-differences techniques, especially when addressing complex confounding and heterogeneity. Second, we source data from the English Housing Survey Fuel Poverty (EHS-FP) dataset series, which provides detailed unit-level information on household income, housing characteristics as well as modelled energy costs. While the reliance on modelled (rather than metered) energy costs introduces a degree of uncertainty —a limitation we directly address in Section 2.1— the EHS-FP includes nationally representative sampling weights for each unit. This allows us to generalise our findings to the entire stock of over 22 million dwellings in England. Crucially, the availability of directly observed household income at the unit level enables precise calculation of fuel burden, in contrast to reliance on indirect proxies. The only two existing works attempting to identify effect heterogeneity of energy efficiency interventions across household types [17,18] rely on area-level indicators such as the Index of Multiple Deprivation (IMD), which fails to capture important within-area income heterogeneity [27,28].

2. Methods

Our methodology is grounded in the concept of *probabilistic causation* [29], thus interpreting an external intervention as affecting the change in probability for an outcome to occur, rather than as a deterministic mechanism. Specifically, we adopt a causal graphical model approach [26], in which the system of interest is represented by a directed acyclic graph (DAG) that encodes assumed causal relationships among variables. Each node in the DAG corresponds to a random variable, and arrowed edges represent direct causal influences between variable pairs (Fig. 1).

This formalism relies on two key assumptions: the *causal Markov* condition (Section 2.4) and the *faithfulness* condition. The latter ensures that statistical independencies implied by the DAG structure are reflected in the observed data. Together, these assumptions enable the use of the graph to identify appropriate adjustment sets and to infer causal effects from observational distributions.

Causal graphical models offer a systematic and transparent framework for causal inference from observational data, which is particularly valuable in settings where randomised controlled trials are unethical, infeasible (as in this case), or prohibitively expensive. Although the formalisation of graph-based causal inference emerged from the field of artificial intelligence —most notably through the foundational work of Pearl [26]— the framework has since been widely adopted across diverse disciplines concerned with causal effect estimation. In epidemiology and public health research for instance, DAGs are established tools to represent confounding, identify adjustment variables, and support causal claims [30–32]. Similarly, in econometrics, graph-based causal reasoning has become an established tool for policy evaluation and treatment effect estimation [33–35].

Building on these foundations, recent developments of causal ML have extended the principles of graphical modelling to high-dimensional and data-rich environments. For example, in the Earth system sciences, causal discovery methods have been used to identify drivers of complex climate phenomena from observational time series [36]. In healthcare and medicine, causal ML has improved diagnostic accuracy and personalised treatment strategies beyond correlational approaches [37–39] such as those based on deep (yet opaque [40]) neural networks. More recently, foundation models to predict missing values in tabular data have shown predictive performance beyond current state-of-the-art when based on graphical (structural) causal models [41].

These methodological foundations provide the framework we relied upon to analyse the causal relationship between insulation interventions, household energy use, and behavioural adjustments. Before specifying the causal graph, we first describe the data on which the model is trained and the procedures used to prepare it.

2.1. Data sourcing and preparation

Data used to train the causal graphical model were obtained from the English Housing Survey - Fuel Poverty (EHS-FP) dataset collection [42], a yearly, nationally representative survey released by the UK government that collects information on household and dwelling characteristics necessary for assessing fuel poverty in England. To ensure methodological consistency, we selected EHS-FP datasets from the pre-pandemic period (2015–2018). During this period, data collection followed standard procedures involving in-person interviews and property inspections, ensuring comparability across survey years. Each yearly EHS-FP dataset contains between 11,853 and 12,024 household records, with the exact number varying by reference year. Approximately 16 % of these records were excluded for not listing gas as the main fuel type. Each remaining household record includes a sampling weight variable [43], which indicates the household's representativeness within the national domestic building stock population of approximately 22.6 million dwellings. These weights were used to perform stratified resampling, generating a weighted dataset of 60,000 household units for each of the four yearly datasets. Testing larger resampled datasets showed no appreciable difference in the estimated results. The final dataset used for training the model parameters therefore consists of 240,000 weighted units in total and is representative of the English domestic building stock across the years 2015–2018. Compared to other datasets such as NEED (which tracks households participating in energy efficiency programmes), the EHS-FP provides a randomly selected sample of the overall (English) housing stock. NEED is longitudinal but primarily includes programme participants, potentially introducing self-selection bias, whereas the EHS-FP provides a representative baseline for country-wide causal inference.

A detailed description of all variables, including definitions and measurement units, is provided in section S0.3 of the supplementary material.

2.2. Causal graph specification

The DAG constructed for our model is shown in Fig. 1. Here, the X symbol designates the intervention variable *walls insulation*, whereas Y_0 and W respectively indicate the outcome variable *gas consumption for space heating* and the covariate *household energy burden*. This latter is taken as the ratio between gas costs for space heating V_1 (plus all other gas and electricity costs, V_{10}), and household income V_7 , i.e. $w = (v_1 + v_{10})/v_7$.

To formally describe the core causal mechanism from a temporal perspective, we begin with the following causal sequence:

$$X \rightarrow Y_0^t \rightarrow V_1 \rightarrow W \rightarrow Y_0^{t+\Delta t} \quad (1)$$

illustrating how an external intervention entailing an energy efficiency measure to the dwelling $X = x$ —i.e. installation of external walls insulation in the specific case—initially leads to a reduction in energy (gas) consumption Y_0 at time t . This decrease in consumption subsequently lowers gas cost expenditures (V_1), thereby reducing the household's energy burden (W). The change in energy burden, in turn, influences occupants behaviour, potentially affecting subsequent gas consumption (Y_0) at a later time $t + \Delta t$.

Following the initial reduction in gas consumption at t , there may be a compensatory behavioural response leading to an increase in consumption at $t + \Delta t$. Consequently, the causal effect triggered by X enters a feedback loop involving the variables Y_0 , V_1 , W ultimately returning to Y_0 . As a result, both gas consumption and associated costs oscillate around (and eventually settle on) an equilibrium state post-intervention, which is constrained by the governing price relationship correlating cost and consumption of gas for a household:

$$\text{Gas cost } (V_1) \text{ [£]} \approx \text{Gas consumption } (Y_0) \text{ [kWh]} \times \text{Gas unit price } (V_9) \text{ [£/kWh]}$$

Capturing this higher-order feedback loop effect is beyond the scope of this analysis. Instead, we focus on identifying the direct physical effect

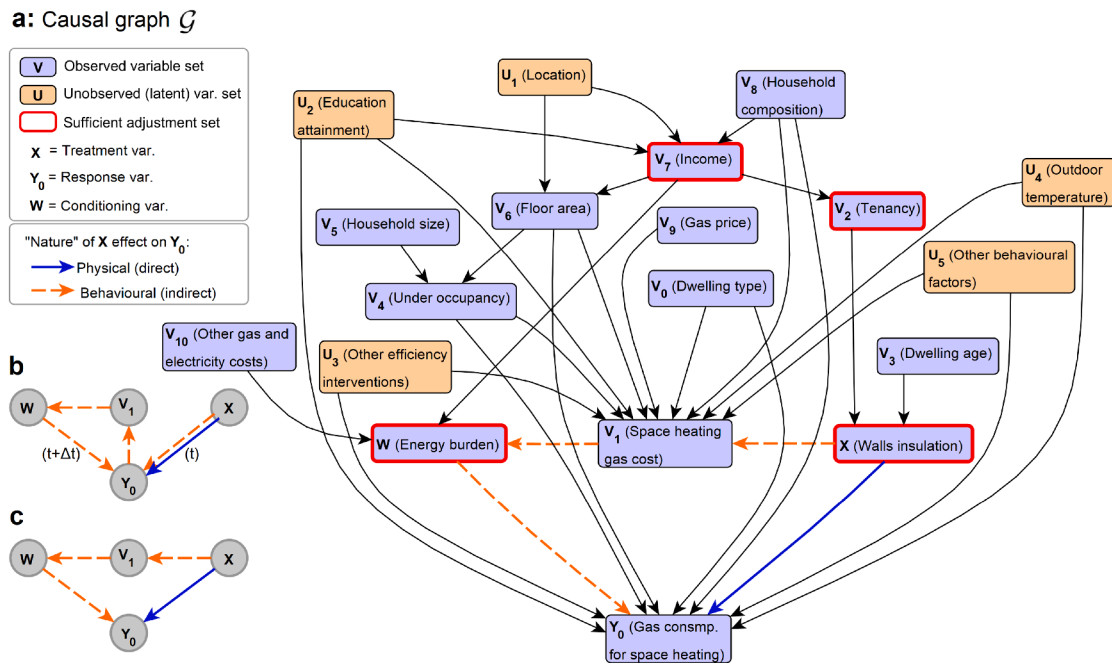


Fig. 1. Causal graphs specification representing the structure of assumed causal relationships. **a:** full directed acyclic graph (DAG) $\mathcal{G}$, incorporating the intervention variable, X , the outcome, Y_0 , and additional covariates representing potential confounders and mediators such as energy burden, W . **b:** the core causal mechanism is first modelled as a cyclic temporal graph, thus capturing the dynamic feedback loop initiated by the intervention X and mediated through gas consumption Y_0 at time t , then gas cost (V_1), and household energy burden (W), eventually influencing future gas consumption at time $t + \Delta t$. **c:** modified acyclic version of the core causal graph, derived by unfolding the feedback loop into two temporally undifferentiated causal pathways, i.e. a direct physical effect ($X \rightarrow Y_0$) and an indirect behavioural effect ($X \rightarrow V_1 \rightarrow W \rightarrow Y_0$), as formalised in Eq. (2). This transformation from cyclic to acyclic ensures compatibility with the requirement of DAG-based causal inference modelling.

of the efficiency intervention ($X \rightarrow Y_0^t$) and the subsequent *behavioural* adjustment mediated by changes in energy cost and burden for the household ($X \rightarrow \dots \rightarrow Y_0^{t+\Delta t}$).

By deliberately omitting second-order feedback effects, we can simplify the graphical representation of the response variable to energy efficiency interventions such as external wall insulation. Specifically, we replace the temporally directed cyclic chain (shown in Fig. 1-b and Eq. (1)) with an equivalent acyclic graph involving two atemporal paths, both starting at X and eventually reaching the response variable Y_0 from separate ways:

$$\begin{aligned} X &\rightarrow Y_0 \\ X &\rightarrow V_1 \rightarrow W \rightarrow Y_0 \end{aligned} \quad (2)$$

The first pathway represents the *physical* causal mechanism through which an energy efficiency intervention affects gas consumption for space heating. In the specific case of wall insulation, this direct effect operates by reducing the thermal transmittance (U-value) of the building envelope, thereby decreasing the rate of heat loss. The term *direct effect* refers here to the expected change in gas consumption, Y_0 , resulting from changing the state of wall insulation, X , from *false* (no insulation) to *true* (insulation installed), while holding constant the mediating variables—in this case, gas cost and energy burden—at the levels they would have attained under the external intervention. This notion of a direct effect corresponds to what Robins and Greenland termed the *pure* direct effect [44] and what Pearl later formalised as the *natural direct effect* within the framework of structural causal models [45].

In contrast, the second pathway in Eq. (2) captures the *behavioural* mechanism through which the intervention X affects gas consumption Y_0 , mediated by changes in gas costs and energy burden. This behavioural pathway is obtained by removing the mediating variable Y_0^t from Eq. (1) and introducing an edge from the intervention X pointing directly to gas costs V_1 (Fig. 1-c). This modification transforms the initial cyclic graph into an equivalent acyclic graph, that is a methodological prerequisite for causal graphical modelling [26].

2.3. Accounting for confounders

The core acyclic graph structure, shown in Fig. 1-c, is further extended by introducing additional variables which may act as potential confounders of the relationship between the intervention (X) and the outcome (Y_0). A confounder is a variable that influences both the treatment and the outcome, hence failing to account for such variables (e.g. via statistical adjustment) can result in biased estimates of causal effects.

To mitigate this risk, we expanded the graphical model to include variables that are theoretically expected to confound the primary causal pathways. Importantly, the selection of these variables was guided by theoretical (expert knowledge) understanding of the problem at hand, rather than by data-driven criteria such as statistical association or availability of variables' data in the first place. This approach aligns with best practices in causal DAG construction [32,46,47], which emphasise the primacy of causal assumptions and substantive knowledge over empirical correlations when specifying the model structure. A detailed description of such causal assumptions is reported in section S0.4 of supplementary material. By prioritising a theory-driven model specification, we aim to ensure that the set of conditioning variables used in the subsequent analysis appropriately blocks non-causal backdoor paths, thereby enabling valid identification of the causal effect of interest.

This theory-first strategy also addresses a core limitation of discovering causal relations exclusively from data [48], namely: the possibility of fitting multiple DAGs encoding the same set of conditional independencies among observed variables. These DAGs form what is known as a *Markov equivalence class*, i.e. a set of directed acyclic graphs that are statistically indistinguishable based solely on the full joint probability distribution they factorise into (via Eq. (3)), even though they may imply fundamentally different causal mechanisms (arrow orientations). In the absence of prior causal assumptions or experimental data to rely on,

it is therefore not possible to identify the true causal graph within this equivalence class, a point that echoes Cartwright's dictum “*no causes in, no causes out*” [49], highlighting that causal knowledge cannot emerge from statistical information alone.

By leveraging substantive, theory-driven knowledge of the data-generating process under study, we define a credible DAG from within the equivalence class, which conforms to the statistical properties of the data and also reflects defensible causal relationships grounded in domain expertise.

A key consequence of including variables in the causal graph regardless of their data availability is the appearance of unmeasured variables, denoted U_i , which represent latent confounders. Unlike fully observed models, i.e. where all probabilistic dependencies among variables are explicitly captured, the presence of unobserved common-cause confounders can complicate the identification of causal effects because such latent variables cannot be conditioned on (as we have no observational data for them) and therefore identification of unbiased causal effects is not always guaranteed in such settings. We deal with this problem in Section 2.5.

The final set of variables in the model consists of both observed and latent components. The observed set is given by $V = \{X, Y_0, W, V_0, \dots, V_{10}\}$, while the unobserved (latent) set is represented as $U = \{U_0, \dots, U_5\}$. A detailed description of all observed variables is provided in section S0.3 of the supplementary material. The final, fully specified graph, incorporating both observed and latent variables, is shown in Fig. 1-a.

2.4. Modularity assumption

As for Bayesian networks, in a causal graphical model each observed variable $V_i \in V$ is associated with a conditional probability distribution, denoted as $P(V_i = v_i \mid \mathbf{Pa}(V_i))_{\mathcal{G}}$, which describes the probability of $V_i = v_i$ given its parent variables $\mathbf{Pa}(\cdot)$ in the graph $\mathcal{G}$ encoding the underlying causal relationships.

A fundamental principle governing causal graphical models is the *causal Markov condition* (also known as modularity assumption), stating that, conditional on its direct parents in the causal graph, each variable is statistically independent of its non-descendants, meaning that once the immediate causal influences of a variable are accounted for (via conditioning), no additional information from its ancestors or other non-descendant nodes contributes to its belief probability. For example, consider a causal chain $X \rightarrow Y \rightarrow Z$, where X denotes the amount of time a student spends studying, Y their resulting exam score, and Z the likelihood of receiving a scholarship. In this case, once we observe the exam score Y , knowing how many hours the student spent studying, X , provides no further information about their scholarship chances Z (formally: $P(z \mid y, x) = P(z \mid y)$). The effect of X on Z is entirely mediated through Y , and the Markov condition implies that $Z \perp\!\!\!\perp x \mid y \mid \mathcal{G}$.

As a direct consequence of this assumption, the joint probability distribution over the full set of observed variables V can be expressed as a product of *factors* representing local conditional probability distributions, each corresponding to a variable conditional on its parents in $\mathcal{G}$. This factorisation follows the structural dependencies encoded in the causal graph and is formally given by Koller and Friedman [50]:

$$P(v_1, \dots, v_n)_{\mathcal{G}} = \prod_{i=1}^n P(v_i \mid \mathbf{Pa}(V_i))_{\mathcal{G}} \quad (3)$$

2.5. Identifiability

In this graphical setting, inferring causal effects from observational data requires to establish whether such inferences are theoretically permissible under the assumed causal structure. In particular, if all relevant variables are observed—what is known as a *Markovian model*—then post-intervention distributions such as $P(Y \mid do(X = x))$, representing the effect on Y of an external intervention setting $X = x$, can always

be derived using a truncated version [26] of the standard Bayesian factorisation Eq. (3). Such procedure however relies on the requirement that all potential confounding variables are measurable which is rarely the case in most realistic settings (and our case makes no exception). This more general scenario involving unmeasured confounders U_i in the graph, and referred to as a *semi-Markovian model*, complicates causal effect identification, as identifiability is not always guaranteed in such setting.

To address this, we employ Pearl's *do-calculus* [51], which provides a formal procedure to assesses whether a causal effect can be identified from observational (non-interventional) data, given a specified semi-Markovian causal graph. If identifiability is established, *do-calculus* yields an explicit formula for estimating the causal effect using only observed quantities, which is reported in the following section.

2.6. Effects estimation

The covariate-specific causal effect of external walls insulation (X) on the response variable gas consumption for space heating (Y_0) in a subset of the sample population for which energy burden (W) achieves a specific value w after the intervention, it is estimated via the following formula:

$$P(y_0 | do(x), w) = \sum_{V_7} \left[P(y_0 | x, w, v_7) \left(\frac{\sum_{V_2} P(v_7, w | x, v_2) P(v_2)}{\sum_{V_2} P(w | x, v_2) P(v_2)} \right) \right] \quad (4)$$

A formal proof of the estimator formula in Eq. (4) based on *do-calculus* computations is provided in the supplementary material section S0.2.

Eq. (4) expresses the probability of observing a specific level of gas consumption Y_0 in the subgroup defined by energy burden $W = w$, under the intervention $do(X = x)$. The summation over V_7 averages over all possible household income levels. The fraction in round brackets acts as a weighting term: it determines how much each income group v_7 should contribute when combining the conditional probabilities $P(y_0 | x, w, v_7)$. These weights are obtained by marginalising over tenancy type (V_2). In the causal graph, tenancy is a pre-treatment covariate that mediates the effect of household income (V_7) on the probability of receiving insulation (X). Statistically, this re-weights the observed income distribution so that it reflects the population structure once insulation has been intervened upon. Intuitively, Eq. (4) combines two pieces of information: (i) the conditional relationship between energy use Y_0 and the observed attributes (X, W, V_7), and (ii) the weighted representation of how frequently each income group v_7 occurs in the post-intervention population with energy burden $W = w$, obtained by appropriately averaging over tenancy types. Together, this ensures that the resulting estimate corresponds to the distribution of Y_0 that would be observed if wall insulation were intervened upon in the real population.

The causal effect of external walls insulation (X) on the response variable (Y_0) across the entire sample population, $P(y_0 | do(x))$, can then be derived by applying the law of total probability:

$$P(y_0 | do(x)) = \sum_w P(y_0 | do(x), w) P(w | do(x)) \quad (5)$$

where the first *factor* on the r.h.s. expressed in terms of observational quantities is obtained via Eq. (4) whilst the second *factor* on the r.h.s., $P(w | do(x))$, is derived from observational probability distributions as follows:

$$P(w | do(x)) = \sum_{V_2} P(w | x, v_2) P(v_2) \quad (6)$$

A formal proof of Eq. (6) is provided in the supplementary material (Eq. (S16)). Eq. (5) aggregates the subgroup-specific distributions from Eq. (4) into the overall population-level post-intervention distribution. The term $P(y_0 | do(x), w)$ captures the effect of the intervention within the subgroup defined by burden level w , while $P(w | do(x))$ represents how

frequently that subgroup occurs in the population once insulation has been intervened upon. Multiplying and summing out across all values of w therefore combines the within-group causal effects into a single population-wide effect. Formally, this is an application of the law of total probability, but intuitively it means that the overall effect is obtained as a weighted average of subgroup effects, where the weights correspond to the prevalence of each subgroup after the intervention.

Notably, the unbiased population-wide causal effect $P(y_0 | do(x))$ can be obtained —as alternative to Eq. (5)— using the *back-door criterion* [26], which involves identifying a set of adjustment variables satisfying the following two conditions:

- No variable in the adjustment set is a descendant of X
- The adjustment set blocks every back-door path between X and Y_0

A back-door path is any path from X to Y_0 in the causal graph $\mathcal{G}$ that includes an arrow pointing into X , thus representing a non-causal path that may introduce confounding bias. For example, variable V_2 satisfies both conditions and thus qualifies as a valid (and sufficient) adjustment set to compute $P(y_0 | do(x))$ (see Fig. 1).

However, our goal extends to estimating the covariate-specific causal effect for subpopulations defined by a fixed energy burden value $W = w$. This requires that all back-door paths from X to Y_0 remain blocked after including W in the conditioning set [52]. Unfortunately, this is not feasible under the back-door criterion because W is a descendant of X , and including it in the adjustment set would violate the first of the two back-door conditions stated above. Indeed, cautionary guidance in the literature has long warned against conditioning on post-treatment variables due to risks of introducing collider bias [53].

To address this, we resort to *do-calculus*, which enables principled reasoning about such cases. Contrary to general intuition, the *do-calculus* framework establishes that conditioning on a descendant of X does not necessarily bias the estimate of the causal effect [54] as long as the independency condition stated by its three rules (see supplementary material section S0.1)) are being met. We derive the custom identification formula presented in Eq. (4) accordingly. This enables unbiased estimation of the W -specific causal effect of X on Y_0 , despite W being a post-treatment variable.

Interestingly, the covariate-specific estimator formula so obtained (Eq. (4)) requires observing only four variables $\{X, W, V_2, V_7\}$ in order to compute the post-intervention distribution $P(y_0 | do(x), w)$. This set forms a *sufficient* adjustment set, as it blocks all non-causal paths whilst conditioning on a collider node (W).

All of the observational conditional probabilities on the r.h.s. of Eq. (4), and r.h.s. of Eq. (6), are computed from the Bayesian network factorisation (i.e. Eq. (3)) of the joint $P(V)$ using Variable Elimination (VE), a recursive algorithm to perform *exact* inference in probabilistic graphical models [55]. VE simplifies the computation of marginal probabilities by iteratively eliminating variables and combining *factors* until only the queried marginal or conditional probability remains. Unlike *approximate* inference methods such as those based on Monte Carlo sampling, VE yields exact results that are not subject to approximation errors arising from sampling variability or convergence criteria.

2.7. Average and conditional treatment effects

Post-interventional probability distributions obtained via Eqs. (4 and 5) are then used to compute the corresponding expectation (sample's mean) value of gas consumption Y_0 given the external intervention on wall insulation: $X = \text{true}$ ($= \text{false}$):

$$\begin{aligned} \mathbb{E}(Y_0 | do(x)) &= \sum_{y_0} y_0 P(y_0 | do(x)) \\ \mathbb{E}(Y_0 | do(x), w) &= \sum_{y_0} y_0 P(y_0 | do(x), w) \end{aligned} \quad (7)$$

with n being the number of unique (discretised) assignment values y_0 for the real-valued variable Y_0 . Eq. (7) shows how post-intervention

probability distributions (from Eqs. (4) and (5)) are converted into expected values of gas consumption. These expected values are the building blocks of treatment effect estimation, since they provide the average outcome we would see under each intervention scenario.

Accordingly, both the average treatment effect (ATE), representing the treatment's impact on gas consumption across the entire sample population, and the conditional average treatment effect (CATE), representing the treatment's impact within subgroups of the population characterised by a specific level of energy burden $W = w$, can both be expressed as follows:

$$\begin{aligned} ATE_G &= \mathbb{E}(Y_0 \mid do(X = true))_G - \mathbb{E}(Y_0 \mid do(X = false))_G \\ CATE_G &= \mathbb{E}(Y_0 \mid do(X = true), w)_G - \mathbb{E}(Y_0 \mid do(X = false), w)_G \end{aligned} \quad (8)$$

Specifically, Eq. (8) formalise the average treatment effect and conditional average treatment effect as differences between the expected outcomes with and without insulation. The ATE captures the effect across the full population, while the CATE isolates the effect for households with a specific level of energy burden.

Of note, the subscript G in Eq. (8) indicates that treatment effects are inferred based on the factorisation of $P(V)$ compatible with the causal structure encoded in the graph G —i.e. the factorisation stated in Eq. (3). This means that both the physical causal effect of the wall insulation intervention on gas consumption ($X \rightarrow Y_0$) and the behavioural effect mediated by energy burden ($X \rightarrow V_1 \rightarrow W \rightarrow Y_0$) are taken into account in the estimation of ATE_G and $CATE_G$. A step-by-step illustrative example of ATE and CATE calculation is provided in the supplementary material section S0.7.

2.8. Handling selection bias of gas cost data

Household gas cost data for space heating, as reported in the EHS-FP dataset (V_1) are derived using the BRE domestic energy model (BREDEM), a rule-based building-physics engineering model developed by the Building Research Establishment [56]. BREDEM estimates heating requirements under standardised assumption: households are expected to maintain demand temperatures of 21°C in living rooms and 18°C elsewhere during heating session [57], with heating demand calculated based on several factors such as dwelling characteristics, heating system type, local weather, and household composition [57]. The methodology also accounts for demographic variation (e.g. longer heating sessions for elderly or households with young children, reduced heating of unused rooms in under-occupied dwellings).

As a result, reported gas costs reflect simulated rather than observed gas expenditures. These values are accurate insofar households heat in line with the BREDEM model assumptions. In practice, financially constrained households may underheat to reduce bills, meaning BREDEM-based costs may likely overestimate actual spending for high-burden households—an issue central to the present study, as we test the hypothesis that behavioural factors may lead to deviations from adequate heating regimes. This raises the question of whether such behavioural effects can be reliably detected by training our causal ML model using EHS-FP data for space heating cost.

To address this we reframe the issue as a problem of *sample selection bias* [58,59]. The bias arises because households that maintain adequate heating despite high energy burden are more likely to be represented in the BREDEM-based dataset, while those who underheat are systematically under-represented. Consequently, the EHS-FP dataset constitutes a biased sample of the target population. It can be interpreted as if (actual) gas cost data were collected more frequently from household units engaging in adequate heating while systematically excluding units engaging in financially driven energy saving behaviour.

To handle this inherent bias in the training dataset, we follow on works from Bareinboim and Tian [60], Bareinboim and Pearl [61], Bareinboim et al. [62] which provide causal graphical criteria to test whether it is possible to recover population-wide observational (or post-interventional) distributions from samples collected under preferential

selection. In our causal graphical model, we introduce a binary selection variable S , where $S = 1$ denotes household units were included in the sample dataset, and $S = 0$ indicates otherwise. To capture the underlying causal mechanism biasing the data-gathering process, we introduce an arrowed edge from W to S in the graph G , thus obtaining an augmented graph G_S in which $W \rightarrow S$ conveys the causal effect of energy burden on dataset inclusion/exclusion for the unit. This reflects the idea that household units experiencing a high energy burden ($W = high$) are more likely to adopt heating behaviours that deviate from the assumptions made in the energy simulation model, thus leading to their exclusion from the collected sample.

Our goal is to obtain the post-intervention probability distribution of Y_0 representative of the target population, i.e. $P(y_0 \mid do(x), w)_{G_S}$, but since the sample was collected under selection bias, only the distribution conditional on $S = 1$ is accessible for use, i.e. $P(y_0 \mid do(x), w, S = 1)_{G_S}$. As such, the requirement under which the population-wide causal effect is recoverable from the biased sample essentially entails for these two probability distributions to be equivalent. This can be verified by leveraging on rule 1 of Pearl's do-calculus [26] (see section S0.1 in supplementary material):

$$P(y_0 \mid do(x), w)_{G_S} = P(y_0 \mid do(x), w, S = 1)_{G_S} \text{ if: } (Y_0 \perp\!\!\!\perp X \mid W)_{G_{S,\bar{X}}} \quad (9)$$

Compliance to the conditional independency requirement stated in Eq. (9) can be checked graphically via the *d-separation* criterion [63], by inspecting the mutilated augmented graph $G_{S,\bar{X}}$, that is G_S after removing all edges pointing to X (see Figure S1 in supplementary material). It suffice here to note that by conditioning on W we block any statistical association between nodes S and Y_0 in $G_{S,\bar{X}}$, thus rendering them conditionally independent given X and W .

2.9. Methodological limitations

Our estimates rely on the assumed DAG, which is grounded in substantive theory and empirical evidence (section S0.4 in supplementary material). However the choice of its structure remains a modelling assumption. Misspecification of key causal pathways (for example, omitting a relevant confounder or incorrectly orienting an arrow) could affect the estimated causal effects. We have explicitly discussed this dependence in Section 2.3 and acknowledge it as a limitation of the study. Importantly, the adjustment set used in our analysis is sufficient under the specified DAG, meaning it blocks all non-causal backdoor paths to isolate the effect of X on Y_0 . Nevertheless, readers should interpret the results in the context of the assumptions encoded in the DAG. This limitation is not unique to causal graphical models; all causal inference methods (e.g. including matching and difference-in-differences) crucially rely on untestable assumptions about the underlying data-generating process.¹ Our theory-driven DAG specification increases transparency (and hence defensibility) of these assumptions.

A further limitation concerns the use of BREDEM-simulated rather than observed gas expenditures. Simulated costs provide a consistent benchmark but may overstate actual spending for high-burden households who underheat, leading to sample selection bias. Our approach (following Bareinboim and Tian [60], Bareinboim and Pearl [61], Bareinboim et al. [62]) addresses this issue by using causal graphical criteria to establish when population-level effects can still be recovered

¹ Matching methods rely on the Conditional Independence Assumption, implying that all confounding variables have been included and controlled for in the model. In the language of causal graphical models, this is equivalent to finding a *sufficient* adjustment set of covariates that blocks all non-causal paths between treatment and outcome (see Section 2.6 and Fig. 1). Difference-in-Differences (DiD) methods relies on the Parallel Trends Assumption, that is, the treated and control groups would have followed the same outcome trend without the intervention.

despite preferential sampling. This provides a theoretical guarantee that bias remains low under the stated assumptions. However, without observed (metered) gas expenditures of the same household units, the extent of any residual bias cannot be empirically quantified. Importantly, using metered gas expenditure data from other datasets would not automatically resolve this issue, because these data would come from households different from those in the EHS-FP sample. Without a shared unique identifier (such as a household ID) across the datasets, a standard data linkage or merge is unfeasible. In a statistical modelling context, the two populations may also lack sufficient common support in their overall characteristic distributions, which would prevent the reliable comparison of energy use patterns. Future work could address this limitation through data fusion methods, provided that either uniquely linkable or otherwise strongly comparable household-level data are available.

3. Results

Our analysis aims to estimate the causal impact of energy efficiency interventions on energy use, which in turn influences energy costs and, consequently, the economic burden on households. To do so, we focus on a specific intervention: the installation of external wall insulation, which we denote as the treatment variable X in our causal model (Methods Section 2). While other retrofit interventions may significantly reduce energy demand, not every efficiency intervention necessarily leads to cost savings as well. That is the case for example of replacing a gas boiler with an electric heat pump: a typical domestic heat pump delivers between 3 to 5 units of heat for every unit of electricity used [64], far outperforming even the most efficient gas boilers, which operate at around 95 % efficiency. However, the higher retail unit price of electricity relative to gas (as of January 2025, $\approx 0.25\text{£/kWh}$ for electricity versus 0.06£/kWh for gas; UK averages) offsets much of this efficiency gain. As a result, households transitioning from gas to heat pumps would observe much lower energy demand but not proportionally lower energy bills, especially in the absence of time-of-use tariffs or government subsidies. By contrast, interventions such as building fabric upgrades reduce energy required to maintain indoor thermal comfort regardless of the energy vector and heating system technology installed, making them well-suited to evaluate our central hypothesis around fuel burden alleviation.

For this reason, we selected external wall insulation as our treatment variable (X) throughout the analysis. The estimated effects presented in the following subsections are measured in terms of annual gas consumption levels for space heating (in kWh/yr) denoted by the variable Y_0 .

3.1. Average treatment effect (ATE)

We begin by reporting the population-wide Average Treatment Effect (ATE) of external wall insulation on annual gas consumption for space heating. Here, the ATE is defined as the difference in expected outcomes $\mathbb{E}(Y_0 | do(X))$ under two scenarios: one in which all dwellings in the population sample are externally insulated ($X = \text{true}$), and another in which none receives the intervention treatment ($X = \text{false}$) (see Section 2.7).

Fig. 2-a shows the estimated post-intervention distributions $P(Y_0 | do(X))$ under both treatment conditions, along with their corresponding expectations $\mathbb{E}(Y_0 | do(X))$.

It is important here to distinguish the causal quantity $P(Y_0 | do(X))$ from its observational counterpart $P(Y_0 | X)$. While $P(Y_0 | X)$ tells us how gas consumption differs between dwellings that happen to be insulated and those that are not, it does so within the structure of the existing system, where insulation status may be influenced by other factors such as income, property type, occupant behaviour, etcetera (Section 2.3). By contrast, $P(Y_0 | do(X))$ represents a hypothetical intervention, capturing what the distribution of gas consumption would look like if we were to externally impose insulation (or lack thereof) across all dwellings, breaking any natural dependencies that normally determine who receives the intervention. In this sense, conditioning on X reflects correlation within

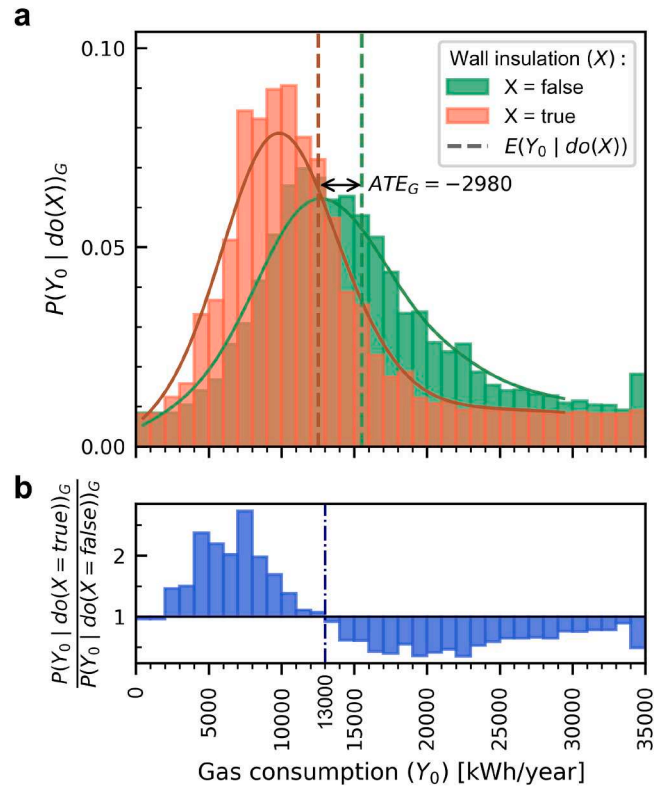


Fig. 2. Estimated causal effect of walls insulation intervention, X , on annual gas consumption for space heating, Y_0 . **a:** post-intervention distributions of gas consumption under the two interventional scenarios: $P(Y_0 | do(X = \text{true}))$ i.e. treatment group, and $P(Y_0 | do(X = \text{false}))$ i.e. control group. The downward shift in expectation $\mathbb{E}(Y_0 | do(X))$ under treatment $X = \text{true}$ indicates a clear average reduction in energy use (≈ -2980 kWh/year). **b:** probability ratio, $PR(Y_0)$, plotted across the range of annual gas consumption levels. $PR(Y_0) > 1$ suggests an increased likelihood of remaining in a low-consumption regime post-insulation, while $PR(Y_0) < 1$ indicates a reduced likelihood of persisting in high-consumption states. A clear pattern emerges, with higher-consuming households ($> 13,000$ kWh/yr) consistently more likely to reduce their energy use compared to lower-consuming households.

the current world, while intervening with $do(X)$ imagines an altered world in which we directly set the value of X and observe the resulting changes in Y_0 . This distinction is foundational to causal inference, and it underpins our interpretation of the average treatment effect reported here [26].

According to our findings, external wall insulation results in a population-wide average reduction of approximately 2980 kWh/year in gas consumption for space heating, from a baseline average annual consumption of $\approx 15,400$ kWh/year. This corresponds to a mean decrease of about 19 % relative to the untreated group.

The estimated 19 % reduction in gas consumption is substantially higher than the effect sizes reported in previous causal inference (difference-in-differences) evaluations of domestic insulation in England. Peñasco and Anadón [18] find that cavity wall insulation reduces annual gas use by approximately 6.9 % in the short term, with the effect tapering off over a four-year period. Similarly, Adan and Fuerst [65] report average reductions of around 10.5 %. While these figures provide benchmarks for comparison, they refer specifically to cavity wall insulation which typically applies to relatively modern housing, with better thermal performance at baseline. By contrast, our analysis does not distinguish between cavity and solid wall types, as our focus is on the overall effect of insulation regardless of wall construction. This approach includes a broader sample, capturing more older, less energy-efficient homes, particularly those with solid walls, where insulation typically

leads to greater gas savings. Additionally, both studies rely on data from the LEED dataset, which is constructed from observed participants in government retrofit schemes. As such, the sample may be subject to self-selection bias, since households opting into such schemes may differ systematically from the general population. In contrast, our use of a randomly selected (EHS) dwelling stock sample avoids this particular source of bias. Finally, it remains unclear in the literature whether reported percentage changes in metered gas consumption, based on the LEED dataset, refer specifically to space heating or to total household gas use, i.e. including other end-uses such as cooking and water heating. This ambiguity may contribute to discrepancies in reported effect sizes, especially given that insulation interventions only affect the share of gas consumption for the specific use of space heating.

Beyond average effects, the distributional impact of external wall insulation can be further explored through analysis of the probability ratio $PR(Y_0)$, defined as the ratio of post-intervention outcome probabilities under treatment and control: $PR(Y_0) = \frac{P(Y_0|do(X=true))}{P(Y_0|do(X=false))}$. As shown in Fig. 2-b, this ratio captures how the likelihood of a given level of gas consumption changes under the interventional scenario compared to the baseline (i.e. untreated) scenario. The analysis reveals a clear threshold effect at around 13 MWh/year. For households with baseline gas consumption below this value, the probability ratio consistently exceeds 1, indicating that external wall insulation increases the likelihood of remaining in a low-consumption regime. In contrast, for households with higher baseline usage (above 13 MWh/year), $PR(Y_0) < 1$, thus implying a reduced likelihood of persisting in high-consumption states post-intervention. This indicates a general redistribution of the probability mass, shifting toward lower levels of gas consumption under treatment. It also indicates a clear distributional effect trend of the intervention across the sample population, in that high-consuming households are systematically more likely to reduce their gas consumption levels post-treatment compared to those with a pre-treatment consumption level below the 13 MWh/year threshold.

3.2. Conditional average treatment effect (CATE)

While the ATE provides a summary statistic of the effect of external wall insulation across the entire population, it does not reveal how the intervention impacts households facing different levels of energy vulnerability, an aspect that is instrumental to the main hypothesis being tested. To address this, we estimate the Conditional Average Treatment Effect (CATE), defined as the difference in expectation outcomes $E(Y_0 | do(X), W)$ for the interventional scenario $X = true$ and baseline scenario $X = false$, now conditional on observing specific levels of energy burden (see Section 2.7). This latter is denoted by W , and defined as the ratio of overall energy costs to household income, a widely used metric to capture vulnerability to fuel poverty.

Fig. 3-a presents the estimated post-intervention outcome distributions $P(Y_0 | do(X), W)$ for households grouped into discrete energy burden strata, ranging from 2% ($\pm 0.5\%$) up to 11%. For each stratum, confidence intervals (40–60%) are shown for both treatment conditions, together with the corresponding medians and conditional expectations $E(Y_0 | do(X), W)$. We observe that average gas consumption rises with energy burden, irrespective of treatment status. However, as revealed by the distance between treated and untreated distributions, the effect of insulation declines steadily as energy burden increases. This results in greater overlap between the two distributions for high-burden groups, signalling a reduced intervention impact for these households.

This heterogeneity is more clearly illustrated in Fig. 3-b, which shows CATE values across energy burden levels. Households with relatively low burden (1.5–2.5% of income spent on energy) experience the largest treatment effect, with an average gas reduction of approximately 3720 kWh/year ($\approx -26\%$ relative to pre-intervention level). This benefit decreases progressively, reaching only ≈ 500 kWh/year ($\approx -3\%$) for households in the highest burden category ($W > 11\%$).

Furthermore, analysis of the probability ratio across energy burden subgroups (Fig. 3-c) provides additional evidence of the uneven distributional effects. Among households with low energy burden ($W = 2\% \pm 0.5\%$), those with mid-to-high baseline gas consumption (i.e., $> 9,000$ kWh/yr) are consistently more likely to reduce their demand following the insulation intervention, as indicated by probability ratios < 1 . However, this pattern progressively fades as energy burden increases. At higher burden levels, such as $W = 11\% (\pm 0.5\%)$, no clear threshold between high and low gas consumers is observed.

Here, probability ratios flatten and converge toward unity across the board, independently of pre-intervention consumption levels, thus suggesting a reduced differential impact of the intervention. This confirms that for households facing higher energy burden, the likelihood of achieving substantial reductions in gas consumption diminishes regardless of their absolute consumption levels. One likely explanation is that these households are already limiting their energy use to prioritise other essential expenditures (e.g. food, rent, debt repayments). As a result, enhancing the energy efficiency of their dwelling does not lead to further meaningful reductions in gas consumption, since energy demand was already suppressed prior to the intervention, that is strong evidence of prebound effect. Arguably, for the most vulnerable groups, financial constraints (rather than energy inefficiency) is the primary barrier preventing them to lower their carbon footprint further.

3.3. Discussion

Energy efficiency upgrades lead to significant energy reductions for households with low energy burden. For high-burden households, reductions are more modest, suggesting financial constraints and/or behavioural adaptation may limit the estimated carbon savings.

A critical implication of our results is the emergence of a structural trade-off between mitigating climate change and improving household well-being. Energy efficiency upgrades lead to significant energy reductions for households with low energy burden. In contrast, for high-burden households, the intervention yields relatively modest gas consumption reductions. This suggests that financial constraints limit further carbon savings for this group. The explanation lies in behavioural adaptation post-intervention: households reallocate the financial savings from the intervention toward increasing indoor temperatures, thereby alleviating thermal discomfort. While this behavioural shift may limit the net carbon savings predicted by engineering models, if households use the resulting financial savings to increase indoor temperatures, this could imply social co-benefits for occupants, although we do not directly measure such outcomes.

Our analysis did not directly measure the causal effect on indoor temperature levels across the household population pre- and post-intervention, meaning we cannot provide direct empirical evidence that efficiency interventions lead to adequate temperature levels for subgroups that may engage in underheating behaviours. We acknowledge the possibility that some households might simply engage in wasteful heating behaviours (e.g., leaving windows open while heating) once they can afford it, due to exogenous behavioural factors. While our causal graphical model is specifically designed to account for additional behavioural effects of latent confounders such as educational attainment (U_2) or other exogenous variables (U_5) affecting both gas consumption and the likelihood of high energy burden, future research could complement this approach by including household-level measurements of indoor temperature or other comfort indicators before and after interventions. Such data would allow empirical validation of whether the estimated behavioural adaptation indeed improves well-being. As explicitly visualised in Fig. 1-a, our model ensures that any statistical association measured between the intervention and the outcome, when mediated via energy burden (W), can be interpreted in causal terms. This means our findings represent the isolated causal effect of financial relief resulting from the efficiency intervention ($X \rightarrow V_1 \rightarrow W \rightarrow Y_0$),

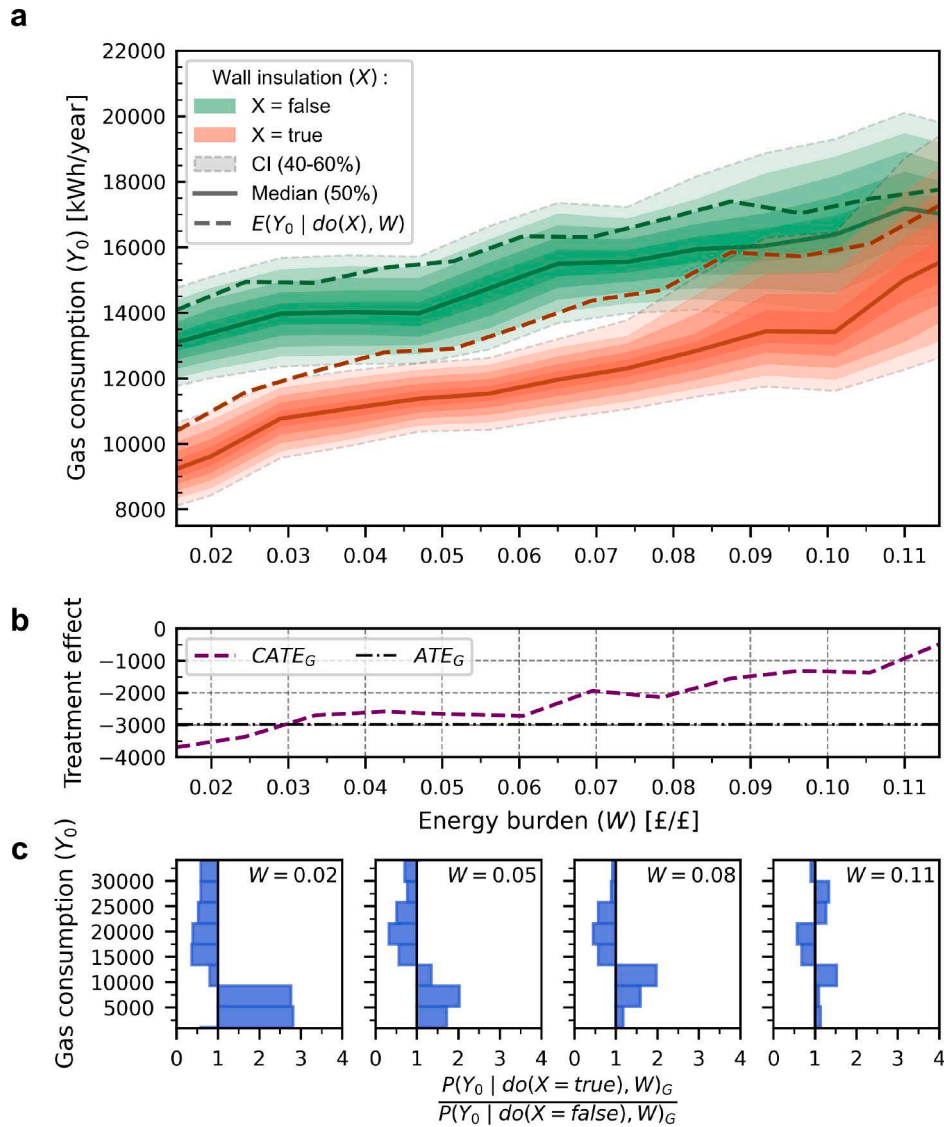


Fig. 3. Conditional average treatment effects (CATE) of external wall insulation on annual gas consumption by energy burden subgroups. **a:** estimated post-intervention outcome distributions $P(Y_0 | do(X), W)$ by energy burden level W , defined as the share of income spent on energy. Confidence intervals (40–60%), medians, and conditional expectations $E(Y_0 | do(X), W)$ are reported for both treated and untreated scenarios. **b:** plots of the Conditional Average Treatment Effect (CATE) as a function of W , revealing a declining trend in treatment efficacy with increasing energy burden: households spending 2% ($\pm 0.5\%$) of income on energy see reductions of ≈ 3720 kWh/year ($\approx -26\%$), whereas the most burdened ($W > 11\%$) only achieve ≈ 500 kWh/year ($\approx -3\%$). **c:** probability ratios $PR(Y_0)$ of population subgroups experiencing different levels of energy burden, showing that low-burden households with medium-to-high baseline gas consumption are consistently more likely to reduce demand post-intervention ($PR(Y_0) < 1$). This pattern disappears at higher burden levels (e.g. $W = 0.11 \pm 0.5\%$), where all households show limited post-treatment change, regardless of initial gas consumption.

insofar the specified causal graph *faithfully* captures the data generating process being modelled (Methods section).

Although our analysis is limited to the UK, the underlying behavioural mechanism (whereby energy efficiency interventions enable households with high energy burden to shift from energy deprivation to improved thermal comfort) it is likely to have broader relevance. In particular, similar dynamics may be observed in countries of the Global South, where energy poverty is widespread and housing quality is often poor, with limited access to affordable energy services. As these regions undergo rapid urbanisation and face increasing cooling demand due to climate change, energy efficiency policies should anticipate the same tension between carbon mitigation and the need to ensure liveable indoor environments. Future policy and research efforts should also account for cultural and climatic variability in household energy practices and thermal expectations, which may shape both the direction and magnitude of behavioural rebound effects in diverse contexts.

3.4. Policy implications

Our findings pose a significant challenge to prevailing methodologies in building energy modelling and policy evaluation. Energy and climate policies are frequently framed with the explicit intent of delivering dual benefits: simultaneously contributing to GHG emission reduction targets and addressing social inequalities, particularly fuel poverty and its associated health impacts [66–68]. However, our study reveals a crucial nuance. Academic literature and policy assessments typically project energy efficiency intervention impacts assuming idealised baseline conditions, resulting in uniform energy savings across diverse occupant populations, often through the use of simplified building archetype models [69,70] that do not capture behavioural heterogeneity due to pre-existing energy deprivation. Consequently, conclusions drawn from these methods offer only a partial understanding of actual energy savings and their distribution, potentially leading to overestimations of

carbon reductions while significantly undervaluing the crucial social benefits related to improved liveability.

The “win-win” narrative that energy efficiency simultaneously delivers substantial carbon emission reductions and improved household well-being requires critical re-evaluation. Our results reveal a structural trade-off in real-world applications of domestic energy efficiency interventions, particularly in the context of fuel poverty. For households with low energy burdens, insulation effectively delivers climate gains. These households are able to reduce gas consumption substantially without compromising indoor comfort, translating into tangible carbon savings aligned with engineering model predictions.

However, for those experiencing high energy burdens, the primary benefit of the intervention shifts from maximising carbon savings to potentially alleviating thermal discomfort and enhancing well-being. These households rationally reallocate financial relief to achieve healthier indoor temperatures, resulting in an increase in energy use aimed at improving comfort, which, while profoundly beneficial socially (e.g., reducing respiratory/cardiovascular illnesses and excess winter mortality [9–12]), significantly diminishes the intervention's net carbon mitigation potential for this group. These crucial social welfare benefits are rarely captured in carbon accounting models informing climate policy, thus underestimating the total societal value of energy efficiency interventions.

This nuanced trade-off is particularly pertinent given recent energy price surges (2021–2023) across the UK and Europe, which have pushed more households into higher energy burden strata [71]. In this evolving landscape, the role of energy efficiency interventions may increasingly be understood as a means to alleviate thermal discomfort and fuel poverty, rather than solely as a tool to drive down GHG emissions. Policies that do not account for these divergent outcomes and the behavioural adaptations they trigger risk misrepresenting their true impact and inadvertently overlooking the complex needs of vulnerable populations [72]. For example, UK schemes like the Energy Company Obligation (ECO) [68], which target low-income, fuel-poor households, may miscalculate their carbon contribution if they solely rely on models that assume uniform energy savings. Policymakers could refine such programmes by explicitly considering household energy burden profiles, targeting support to high-burden households with interventions that simultaneously alleviate fuel poverty and account for expected behavioural adaptation.

We advocate for an integrated policy framework that explicitly incorporates social welfare objectives, such as thermal comfort and public health indicators, into carbon accounting models. For instance, carbon accounting frameworks could weight avoided emissions differently for households where energy efficiency upgrades also deliver measurable health and well-being improvements, or include metrics such as reductions in cold-related morbidity and excess winter mortality. This approach would ensure a holistic assessment of energy efficiency measures, valuing both their climate and crucial equity contributions. Positioning energy efficiency as part of a broader strategy to reduce both emissions and social inequalities will ultimately contribute to a more just and sustainable transition to a low-carbon economy.

3.5. Directions for future research

In this study, we focused on income and energy burden as the primary sources of heterogeneity, as these dimensions directly capture distributional concerns central to energy poverty and efficiency policy. We acknowledge that additional factors—such as dwelling type, geographic region, tenure status, or urban/rural distinctions—could offer further insights into the differential impacts of insulation interventions. However, some of these dimensions cannot be fully explored with the current dataset; for instance, urban/rural location and fine-grained geographic identifiers are not available in the publicly accessible version of the English Housing Survey. For other factors, such as dwelling type and tenure, the causal model already incorporates related structural features

(e.g., wall type, ownership vs. rental) within the adjustment set, though these are not analysed as separate heterogeneity dimensions. Extending the heterogeneity analysis to cover all suggested dimensions would require additional data or an alternative study design, which lies beyond the scope of the present work. We highlight this as a valuable direction for future research.

4. Refutation

Causal effects estimation relies on critical assumptions regarding the data-generating process (see Methods Section 2). As a result, any violation of these assumptions can lead to significant inaccuracies in the estimated effects. In contrast to ML models aimed at purely predictive tasks, where cross-validation offers a comprehensive measure of model performance, causal inference lacks a universally applicable validation method. To tackle this issue, we draw on recent work [73] grounded in the philosophical principle of *falsifiability* [74]—positing that scientific claims cannot be conclusively proven, but gain credibility by resisting empirical attempts at refutation—we conduct refutation checks to test the reliability of our causal estimates. Aim of these tests is to check whether the estimated treatment effect conforms to the theoretical expectation required for the estimator to produce valid causal inferences. Such expectations will depend on the specific test, as explained in each of the following subsections.

4.1. Placebo treatment test

First, we perform a placebo treatment test [75] by randomly permuting the values of the treatment variable X in the training dataset. This involves assigning to each household unit a treatment value x drawn at random from the original dataset, effectively reallocating treatment values across units. This process replaces the original treatment variable X with a randomised version, X_{placebo} , thereby ‘breaking’ any causal association between X and its children variables $\text{ch}(X)$, while preserving its marginal probability distribution:

$$\begin{aligned} P(X_{\text{placebo}}) &= P(X) \\ P(\text{ch}(X_{\text{placebo}}) | X_{\text{placebo}}) &\neq P(\text{ch}(X) | X) \end{aligned} \quad (10)$$

We then evaluate whether the causal estimator (Eqs. (4–8)) detects any significant association under this placebo setting. Under valid causal assumptions, the theoretical expectation is that the estimator should yield a null average treatment effect for the randomised treatment: $ATE_{\text{placebo}} \approx 0$. While this test does not directly assess the estimated treatment effect, detecting an effect where none should exist would raise concerns about the model's validity (and hence about its estimate). To formally assess this, we define the following null hypothesis: the estimated true treatment effect ATE_C is not statistically distinguishable from a (random noise) placebo treatment ATE_{placebo} .

To test this hypothesis we repeat the random reallocation process n times, thus generating $n = 1600$ individual placebo datasets and compute the placebo treatment effect $ATE_{i,\text{placebo}}$ for every i -th dataset. These are then used to construct a reference distribution under the null. A p-value is then computed by comparing the magnitude of ATE_C to the empirical distribution of placebo effects (see section S0.5 in supplementary material).

Summary statistics for the placebo distribution are presented in the third column of Table 1. As shown, both the mean and median of ATE_{placebo} are close to zero (around -5 kWh/year), whereas the estimated true effect ATE_C is substantially larger by comparison. The very low p-value suggests that the observed effect is highly unlikely under the null hypothesis, providing strong evidence to reject it.

4.2. Data subsample test

The second refutation strategy assesses the robustness of the estimated average treatment effect to variations in the training data by

Table 1

Robustness checks for the estimated average treatment effect (ATE_G). The placebo test (third column) assesses whether spurious effects emerge when treatment values are randomly reassigned, simulating a null scenario. The subsample test (fourth column) evaluates estimate stability across random subsamples ($\approx 40\%$ of the full training dataset). Reported are means, medians, 1–99% confidence intervals, p-values, and sample sizes. The estimated true effect, shown in the final column, is substantially larger than placebo estimates and consistent with the empirical distribution of treatment effect obtained from random subsamples of the training dataset.

	Unit	ATE_{placebo}	ATE_{sub}	ATE_G (baseline) [kWh/yr]
Mean	[kWh/yr]	–5.2	–2935.5	
Median	[kWh/yr]	–5.5	–2935.6	
CI (1%; 99%)	[kWh/yr]	(–66.5; 51.6)	(–3004.1; –2872.2)	–2980.1
p-value	–	0.000625	0.1274	
Sample size (n)	–	1600	1800	

employing repeated random subsampling. Specifically, we generate $n = 1800$ randomly drawn subsamples of the training dataset, each comprising $\approx 40\%$ of the units in the full dataset, and compute the corresponding treatment effect estimate, $ATE_{i,\text{sub}}$, for each subsample i . This process yields an empirical distribution of subsample-based ATE s, which is then compared to the ATE estimated on the full dataset, i.e. ATE_G .

The intuition behind this test is that, if the full-sample estimate is stable and not unduly influenced by any particular segment of the data, it should lie somewhere within the empirical distribution formed by the subsample estimates. In contrast, a large deviation may signal sensitivity to specific data partitions, calling into question the generalisability of the result.

Formally, the null hypothesis for this test states that: the treatment effect estimated from the full dataset is not significantly different from the ATE s estimated on random subsamples. Thus, in this context, a high p-value is desirable (see section S0.6 in supplementary material), as it implies that any observed deviation of ATE_G from the distribution of $ATE_{i,\text{sub}}$ can plausibly be attributed to random sampling variation. Retaining the null hypothesis in this case strengthens the credibility of the causal estimate, indicating that it is not overly sensitive to the particular composition of the training data. Summary statistics for this test are reported in the forth column of Table 1, where it can be seen how the estimated true effect (–2980.1 kWh/yr) is well within the confidence interval of the empirical distribution for ATE_{sub} .

4.3. Sensitivity analysis to unobserved confounder

While the placebo and subsample tests assess robustness empirically within the observed data, a more fundamental concern is potential unmeasured confounding. By definition, such confounding cannot be directly tested, yet it may still bias causal estimates if an omitted variable influences both treatment assignment and outcomes ($X \leftarrow U \rightarrow Y_0$). To quantify this risk, we conduct a sensitivity analysis, which evaluates how strong such a confounder would have to be in order to fully explain away the estimated treatment effect ATE_G . Although the existence of such a confounder is hypothetical and most importantly it's unknowable (it might represent one, several, or none of the underlying factors), we will refer to it as *energy consciousness* of occupants, so to make the following exposition more intuitive. Let say conscious households are more likely to choose residing in insulated dwellings (through selective renting, purchasing, or investing in upgrades) and may also consume less gas for heating, independently of insulation. We formalise this using a standard linear omitted-variable bias (OVB) formula. Let $U \in \{0, 1\}$ denote a binary energy-consciousness indicator, with $U = 1$ indicating “conscious” households, and suppose U affects the probability of dwelling insulation X by:

$$\pi = P(X = \text{true} \mid U = 1) - P(X = \text{true} \mid U = 0) \quad (11)$$

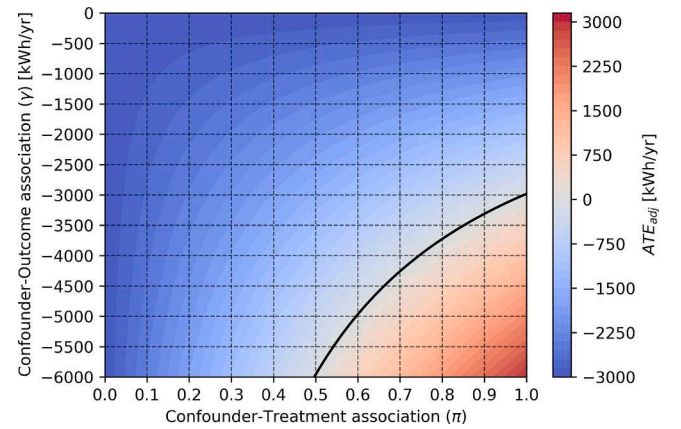


Fig. 4. Sensitivity analysis of the estimated average treatment effect (ATE_G) to a hypothetical unmeasured confounder U . The heat-map shows the adjusted treatment effect, $ATE_{\text{adj}} = ATE_G - \pi \cdot \gamma$, where π denotes the confounder-treatment association (increase in probability of dwelling insulation attributable to the confounder), and γ denotes the confounder-outcome association (change in annual gas consumption attributable to the confounder, independent of insulation). Shading represents the magnitude of the adjusted ATE. The solid black contour represents the tipping line, i.e. the locus of (π, γ) combinations that would exactly explain away the estimated effect of insulation ($ATE_G = -2980.1$ kWh/year).

whereby π represents the confounder-treatment association strength (i.e., the effect of energy consciousness on the probability of insulation). Similarly, let U affect the expected annual gas consumption Y_0 by:

$$\gamma = E(Y_0 \mid U = 1) - E(Y_0 \mid U = 0) \quad (12)$$

where γ represents the confounder-outcome association strength (i.e., the effect of energy consciousness on gas consumption, independent of insulation). If U exists, the estimated ATE_G will be biased by the quantity $\gamma\pi$, so that the adjusted effect is expressed as follows Wooldridge [76]:

$$ATE_{\text{adj}} = ATE_G - \gamma\pi \quad (13)$$

Fig. 4 shows ATE_{adj} across a grid of (π, γ) value pairs. Shading represents the magnitude of the adjusted effect, while the solid contour marks the tipping line where $ATE_{\text{adj}} = 0$. This line identifies all combinations of confounder-treatment and confounder-outcome association strengths sufficient to nullify the estimated effect of insulation ($ATE_G = -2980.1$ kWh/y). To illustrate, if energy-conscious households were three times more likely to choose living in insulated dwellings ($\pi = 0.75 - 0.25 = 0.5$) then energy consciousness would also need to reduce consumption independently by -5960 kWh/yr ($\approx 40\%$ of average baseline demand) to eliminate the insulation effect. For more modest confounder-treatment associations the confounder-outcome association required to explain away² the estimated effect becomes unrealistically large. For instance, assuming conscious households are only $1/3$ more likely to choose insulated dwellings compared to uncaring ones ($\pi = 0.6 - 0.4 = 0.2$) then the tipping line implies that consciousness independently drives these households to save as much as $\gamma \approx -14900$ kWh/year in gas, which exceeds the average annual baseline gas consumption of English households. Similarly, if energy consciousness, U , reduces consumption by a more modest $\gamma \approx -3000$ kWh/yr, U would need to be an almost deterministic predictor of insulation, i.e., $\pi = P(X = \text{true} \mid U = 1) - P(X = \text{true} \mid U = 0) \approx 1$. More generally, any realistic value for π would require an implausibly large effect of U on Y_0 , and any realistic effect of U on

² To ‘explain away’ an estimated effect means to show that the entire observed association between two variables can be attributed to (or is fully accounted for by) a third, unobserved variable (a confounder). In this context, it implies driving the estimated true causal effect to zero.

Y_0 would require an implausibly strong association with insulation (π). Thus, within plausible ranges for both π and γ , unmeasured confounding cannot fully explain away the estimated effect ATE_G . This sensitivity analysis therefore provides further quantitative evidence for the robustness of our causal conclusions.

5. Conclusions

The study advances understanding of the causal impacts in domestic energy efficiency interventions by combining a novel methodological framework with nationally representative data. Specifically, we employed a (semi-Markovian) causal Bayesian network and do-calculus, to estimate the causal effect of external wall insulation on household gas demand for space heating. Building on earlier studies using econometric methods (such as difference-in-differences and matching techniques) with the National Energy Efficiency Data (NEED), our methodology explicitly formalises causal assumptions through a graphical model. This enables transparent causal effect identification based on a *sufficient* adjustment set and provides a flexible basis for robust estimation of both average and Conditional Treatment Effects.

Crucially, this work also demonstrates the feasibility of using the English Housing Survey - Fuel Poverty (EHS-FP) dataset for causal inference at a national scale. Because the EHS-FP includes household-level income data and provides a nationally representative random sample, our estimates generalise to the entire English housing stock of approximately 22.6 million dwellings. This represents a significant advancement over earlier work, which may be limited by potential self-selection bias from programme-based samples (i.e., NEED datasets) and by the use of area-level deprivation proxies rather than household-level data.

Our findings indicate that external wall insulation reduces household gas consumption by approximately 19% on average relative to pre-intervention gas consumption levels (≈ -2980 kWh/year). However, Conditional Average Treatment Effect (CATE) estimates reveal pronounced heterogeneity depending on pre-existing levels of energy burden (defined as the share of household income spent on energy). Households in the lowest burden strata (1.5–2.5% of income spent on energy) achieve the largest savings (≈ -3720 kWh/year, $\approx -26\%$). In contrast, those in the highest energy burden category ($> 11\%$) realise only marginal reductions (≈ -500 kWh/year, $\approx -3\%$). These divergent outcomes are best understood as behaviourally rational responses to economic constraints: households already underheating prior to the intervention likely reallocate the financial relief from reduced bills toward higher thermal comfort rather than maximising energy savings. Far from being wasteful, such behavioural responses constitute rational adjustments in contexts of fuel poverty, with likely co-benefits for health and well-being.

The contribution of this study is therefore twofold. Substantively, it provides country-wide causal evidence that the impacts of energy efficiency retrofits are strongly conditioned by household-level energy burden, revealing a structural trade-off between carbon mitigation and social welfare gains. Methodologically, it demonstrates the applicability of causal graphical models for energy policy evaluation for the built environment, and that the EHS-FP dataset can serve as a robust basis for causal inference across the 22.6 million dwellings constituting the English housing stock.

Policy implications follow directly. For households with low energy burden, insulation measures deliver the expected operational carbon reductions in line with engineering model predictions. For high-burden households, however, the principal benefits lie in alleviating thermal deprivation and improving well-being, with only limited net carbon savings. This finding reveals a structural trade-off that challenges the prevailing “win-win” framing of energy efficiency as simultaneously maximising climate and social objectives. We argue that policy evaluation frameworks should explicitly integrate social welfare indicators such as thermal comfort, health outcomes, and equity alongside carbon accounting. Recognising and valuing these co-benefits is essential to accurately

capture the societal impact of energy efficiency interventions across the full housing stock and to support a more just and sustainable low-carbon transition.

CRedit authorship contribution statement

Bernardino D'Amico: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization; **Francesco Pomponi:** Writing – original draft, Resources, Methodology, Investigation; **Jay H. Arehart:** Writing – original draft, Formal analysis, Data curation; **Lina Khaddour:** Writing – original draft, Investigation.

Data and code availability

All data and code developed for this study are publicly available and can be accessed via the referenced GitHub repository [77].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

B.D. and L.K. express their gratitude to the School of Computing, Engineering & the Built Environment (SCEBE) at Edinburgh Napier University for providing the support and resources necessary for the completion of this research.

Supplementary material

Supplementary material associated with this article can be found in the online version at [10.1016/j.enbuild.2025.116613](https://doi.org/10.1016/j.enbuild.2025.116613).

References

- [1] J. Barrett, S. Pye, S. Betts-Davies, O. Broad, J. Price, N. Eyre, J. Anable, C. Brand, G. Bennett, R. Carr-Whitworth, et al., Energy demand reduction options for meeting national zero-emission targets in the United Kingdom, *Nat. Energy* 7 (8) (2022) 726–735. <https://doi.org/10.1038/s41560-022-01057-y>
- [2] C. Xia-Bauer, S. Gokarakonda, S. Guo, F. Filippidou, S. Thomas, J.R. Maheshwari, S.S. Vishwanathan, Comparative analysis of residential building decarbonization policies in major economies: insights from the EU, China, and India, *Energy Effic. 17* (5) (2024) 46. <https://doi.org/10.1007/s12053-024-10225-w>
- [3] E.Ó. Broin, J. Nässén, F. Johnsson, Energy efficiency policies for space heating in EU countries: a panel data analysis for the period 1990–2010, *Appl. Energy* 150 (2015) 211–223. <https://doi.org/10.1016/j.apenergy.2015.03.063>
- [4] M. Economidou, V. Todeschi, P. Bertoldi, D. D'Agostino, P. Zangheri, L. Castellazzi, Review of 50 years of EU energy efficiency policies for buildings, *Energy Build.* 225 (2020) 110322. <https://doi.org/10.1016/j.enbuild.2020.110322>
- [5] R. McAndrew, R. Mulcahy, R. Gordon, R. Russell-Bennett, Household energy efficiency interventions: a systematic literature review, *Energy Policy* 150 (2021) 112136. <https://doi.org/10.1016/j.enpol.2021.112136>
- [6] International Energy Agency, Net Zero by 2050: A Roadmap for the Global Energy Sector, 2022, (IEA Report). Available at: <https://www.iea.org/reports/net-zero-by-2050>.
- [7] I.P. o.C. Change, Climate Change 2023: Synthesis Report, 2023, (IPCC Sixth Assessment Report). Available at: <https://www.ipcc.ch/report/ar6/syr/>.
- [8] HM Government, UK Net Zero Strategy: Build Back Greener, 2021, (UK Government Policy Paper). Available at: <https://www.gov.uk/government/publications/net-zero-strategy>.
- [9] V. Ballesteros-Arjona, L. Oliveras, J.B. Muñoz, A.O. de Labry Lima, J. Carrere, E.M. Ruiz, A. Peralta, A.C. León, I.M. Rodríguez, A. Daponte-Codina, et al., What are the effects of energy poverty and interventions to ameliorate it on people's health and well-being?: A scoping review with an equity lens, *Energy Res. Soc. Sci.* 87 (2022) 102456. <https://doi.org/10.1016/j.erss.2021.102456>
- [10] X. Mei, B.K. Seo, The relationships among housing, energy poverty, and health: a scoping review, *Energy Sustain. Develop.* 83 (2024) 101568. <https://doi.org/10.1016/j.esd.2024.101568>
- [11] C. Moorcroft, A. Whitehouse, J. Grigg, Damp and mouldy home: impact on lung health in childhood, *Arch. Dis. Child.* (2025). <https://doi.org/10.1136/archdischild-2023-326035>
- [12] A. Singh, A. Mizdrak, L. Daniel, T. Blakely, E. Baker, L. Fleitas Alfonso, R. Bentley, Estimating cardiovascular health gains from eradicating indoor cold in Australia, *Environ. Health* 21 (1) (2022) 54. <https://doi.org/10.1186/s12940-022-00865-9>

- [13] J. Harris, J. Hall, H. Meltzer, R. Jenkins, T. Oreszczyn, S. McManus, Health, mental health and housing conditions in England, 2010, NatCen. Available at: <https://beatcold.org.uk/wp-content/uploads/2011/08/health-mental-health-and-housing-conditions-in-england.pdf>.
- [14] Office for National Statistics, Excess Winter Mortality in England and Wales: 2020 to 2021 (Provisional) and 2019 to 2020 (Final), 2021, (<https://www.ons.gov.uk/releases/excesswintermortalityinenglandandwales2020to2021provisionaland2019to2020final>). Accessed: May 2025.
- [15] A. Lee, I. Sinha, T. Boyce, J. Allen, P. Goldblatt, Fuel Poverty, Cold Homes and Health Inequalities, 2022, Institute of Health Equity, London. Available at: <https://www.instituteofhealthequity.org/resources-reports/fuel-poverty-cold-homes-and-health-inequalities-in-the-uk/read-the-report.pdf>.
- [16] A. Haines, A.J. McMichael, K.R. Smith, I. Roberts, J. Woodcock, A. Markandya, B.G. Armstrong, D. Campbell-Lendrum, A.D. Dangour, M. Davies, et al., Public health benefits of strategies to reduce greenhouse-gas emissions: overview and implications for policy makers, *Lancet* 374 (9707) (2009) 2104–2114. [https://doi.org/10.1016/S0140-6736\(09\)61759-1](https://doi.org/10.1016/S0140-6736(09)61759-1)
- [17] D. McCoy, R.A. Kotsch, Quantifying the distributional impact of energy efficiency measures, *Energy J.* 42 (6) (2021) 121–144. <https://doi.org/10.5547/01956574.42.6.dmc>
- [18] C. Peñasco, L.D. Anadón, Assessing the effectiveness of energy efficiency measures in the residential sector gas consumption through dynamic treatment effects: evidence from England and Wales, *Energy Econ.* 117 (2023) 106435. <https://doi.org/10.1016/j.eneco.2022.106435>
- [19] E. Aydin, N. Kok, D. Brounen, Energy efficiency and household behavior: the rebound effect in the residential sector, *RAND J. Econ.* 48 (3) (2017) 749–782. <https://doi.org/10.1111/1756-2171.12190>
- [20] D. Cali, T. Osterhage, R. Streblow, D. Müller, Energy performance gap in refurbished German dwellings: lesson learned from a field test, *Energy Build.* 127 (2016) 1146–1158. <https://doi.org/10.1016/j.enbuild.2016.05.020>
- [21] M. Sunikka-Blank, R. Galvin, Introducing the rebound effect: the gap between performance and actual energy consumption, *Build. Res. Inform.* 40 (3) (2012) 260–273. <https://doi.org/10.1080/09613218.2012.690952>
- [22] W. Anderson, V. White, A. Finney, Coping with low incomes and cold homes, *Energy Policy* 49 (2012) 40–52. <https://doi.org/10.1016/j.enpol.2012.01.002>
- [23] J. San Miguel-Bellod, P. González-Martínez, A. Sánchez-Ostiz, The relationship between poverty and indoor temperatures in winter: determinants of cold homes in social housing contexts from the 40s–80s in Northern Spain, *Energy Build.* 173 (2018) 428–442. <https://doi.org/10.1016/j.enbuild.2018.05.022>
- [24] ONS, The impact of winter pressures on adults in Great Britain: December 2022, 2022, <https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/theimpactofwinterpressuresonadultsingreatbritain/december2022>.
- [25] J.J. Heckman, Selection bias and self-selection, in: *Econometrics*, Springer, 1990, pp. 201–224. https://doi.org/10.1007/978-1-349-20570-7_29
- [26] J. Pearl, *Causality – Models Reasoning and Inference*, Cambridge University Press, Cambridge, 2nd edition, Cambridge, 2009. <https://doi.org/10.1017/CBO9780511803161>
- [27] D. Clelland, C. Hill, Deprivation, policy and rurality: the limitations and applications of area-based deprivation indices in Scotland, *Local Econ.* 34 (1) (2019) 33–50. <https://doi.org/10.1177/026909421982>
- [28] I. Deas, B. Robson, C. Wong, M. Bradford, Measuring neighbourhood deprivation: a critique of the index of multiple deprivation, *Environ. Plann. C: Govern. Policy* 21 (6) (2003) 883–903. <https://doi.org/10.1068/c0240>
- [29] P. Suppes, A probabilistic theory of causality, *Brit. J. Philos. Sci.* 24 (4) (1973).
- [30] S. Greenland, J. Pearl, J.M. Robins, Causal diagrams for epidemiologic research, *Epidemiology* 10 (1) (1999) 37–48.
- [31] G.J. Griffith, T.T. Morris, M.J. Tudball, A. Herbert, G. Mancano, L. Pike, G.C. Sharp, J. Sterne, T.M. Palmer, G. Davey Smith, et al., Collider bias undermines our understanding of COVID-19 disease risk and severity, *Nat. Commun.* 11 (1) (2020) 5749. <https://doi.org/10.1038/s41467-020-19478-2>
- [32] P.W.G. Tennant, E.J. Murray, K.F. Arnold, L. Berrie, M.P. Fox, S.C. Gadd, W.J. Harrison, C. Keeble, L.R. Ranker, J. Textor, et al., Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations, *Int. J. Epidemiol.* 50 (2) (2021) 620–632. <https://doi.org/10.1093/ije/dyaa213>
- [33] P. Spirtes, Graphical models, causal inference, and econometric models, *J. Econ. Methodol.* 12 (1) (2005) 3–34. <https://doi.org/10.1080/1350178042000330887>
- [34] J. Heckman, R. Pinto, Econometric causality: the central role of thought experiments, *J. Econom.* 243 (1–2) (2024) 105719. <https://doi.org/10.1016/j.jeconom.2024.105719>
- [35] P. Hünermund, E. Bareinboim, Causal inference and data fusion in econometrics, *Econom. J.* 28 (1) (2025) 41–82. <https://doi.org/10.1093/ectj/utae008>
- [36] J. Runge, S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Glymour, M. Kretschmer, M.D. Mahecha, J. Muñoz-Marí, et al., Inferring causation from time series in Earth system sciences, *Nat. Commun.* 10 (1) (2019) 2553. <https://doi.org/10.1038/s41467-019-10105-3>
- [37] J.G. Richens, C.M. Lee, S. Johri, Improving the accuracy of medical diagnosis with causal machine learning, *Nat. Commun.* 11 (1) (2020) 3923. <https://doi.org/10.1038/s41467-020-17419-7>
- [38] P. Sanchez, J.P. Voisey, T. Xia, H.I. Watson, A.Q. O'Neil, S.A. Tsafaris, Causal machine learning for healthcare and precision medicine, *R Soc. Open Sci.* 9 (8) (2022) 220638. <https://doi.org/10.1098/rsos.220638>
- [39] S. Feuerriegel, D. Frauen, V. Melnychuk, J. Schweisthal, K. Hess, A. Curth, S. Bauer, N. Kilbertus, I.S. Kohane, M. van der Schaar, Causal machine learning for predicting treatment outcomes, *Nat. Med.* 30 (4) (2024) 958–968. <https://doi.org/10.1038/s41591-024-02902-1>
- [40] J. Pearl, The limitations of opaque learning machines, in: J. Brockman (Ed.), *Possible Minds: Twenty-Five Ways of Looking at AI*, Penguin Press, London, 2019.
- [41] N. Hollmann, S. Müller, L. Purucker, A. Krishnakumar, M. Körfer, S.B. Hoo, R.T. Schirmmeister, F. Hutter, Accurate predictions on small data with a tabular foundation model, *Nature* 637 (8045) (2025) 319–326. <https://doi.org/10.1038/s41586-024-08328-6>
- [42] Department for Business, Energy & Industrial Strategy, English Housing Survey: Fuel Poverty Dataset, 2018, 2021, [data collection]. <https://doi.org/10.5255/UKDA-SN-8655-2>
- [43] N. Security, Department for Energy and Zero, Fuel Poverty Dataset Documentation - End User Licence, Technical Report, UK Data Archive, 2023. https://doc.ukdataservice.ac.uk/doc/9243/mrdoc/pdf/9243_2021_fuel_poverty_dataset_documentation_eul.pdf
- [44] J.M. Robins, S. Greenland, Identifiability and exchangeability for direct and indirect effects, *Epidemiology* 3 (2) (1992) 143–155.
- [45] J. Pearl, Direct and indirect effects, in: *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA, 2001, pp. 411–420.
- [46] Z.M. Laubach, E.J. Murray, K.L. Hoke, R.J. Safran, W. Perng, A biologist's guide to model selection and causal inference, *Proceed. Royal Soc. B* 288 (1943) (2021) 20202815. <https://doi.org/10.1098/rspb.2020.2815>
- [47] L. Poppe, J. Steen, W.W. Loh, G. Crombez, F. De Block, N. Jacobs, P.W.G. Tennant, J.V. Cauwenberg, A.L.D. Paeppe, How to develop causal directed acyclic graphs for observational health research: a scoping review, *Health Psychol. Rev.* 19 (1) (2025) 45–65. <https://doi.org/10.1080/17437199.2024.2402809>
- [48] C. Glymour, K. Zhang, P. Spirtes, Review of causal discovery methods based on graphical models, *Front. Genet.* 10 (2019) 524. <https://doi.org/10.3389/fgene.2019.00524>
- [49] N. Cartwright, No causes in, no causes out, in: *Nature's Capacities and Their Measurement*, Oxford University Press, Oxford, 1994, pp. 39–90. <https://doi.org/10.1093/0198235070.003.0003>
- [50] D. Koller, N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT press, Cambridge, 2009.
- [51] J. Pearl, Causal diagrams for empirical research, *Biometrika* 82 (4) (1995) 669–688. <https://doi.org/10.1093/biomet/82.4.669>
- [52] J. Pearl, M. Glymour, N.P. Jewell, *Causal Inference in Statistics: a Primer*, John Wiley & Sons, Hoboken, NJ, Hoboken, NJ, 2016.
- [53] J.M. Montgomery, B. Nyhan, M. Torres, How conditioning on post-treatment variables can ruin your experiment and what to do about it, *Am. J. Pol. Sci.* 62 (3) (2018) 760–775. <https://doi.org/10.1111/ajps.12357>
- [54] J. Pearl, Conditioning on post-treatment variables, *J. Causal Inferen.* 3 (1) (2015) 131–137. <https://doi.org/10.1515/jci-2015-0005>
- [55] N.L. Zhang, D. Poole, A simple approach to Bayesian network computations, in: *Proc. of the Tenth Canadian Conference on Artificial Intelligence*, 1994. <https://hdl.handle.net/1783.1/757>
- [56] J. Henderson, J. Hart, BREDEM 2012 – A technical description of the BRE Domestic Energy Model, 2012, Technical report, Building Research Establishment, UK. <https://files.bregroup.com/bre-co-uk-file-library-copy/filelibrary/bredem/BREDEM-2012-specification.pdf>
- [57] BRE, Fuel Poverty – methodology handbook, 2017, Technical report, Department for Business, Energy and Industrial Strategy, UK. https://doc.ukdataservice.ac.uk/doc/8289/mrdoc/pdf/8289_fuel_poverty_methodology_handbook.pdf
- [58] J.J. Heckman, Sample selection bias as a specification error, *Econometr. J.* 1979 153–161. <https://www.jstor.org/stable/1912352>
- [59] D.S. Lee, Training, wages, and sample selection: Estimating sharp bounds on treatment effects, 2005, National Bureau of Economic Research Cambridge, Mass., USA. <https://ssrn.com/abstract=837164>
- [60] E. Bareinboim, J. Tian, Recovering causal effects from selection bias, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, 29, 2015. <https://doi.org/10.1609/aaai.v29i1.9679>
- [61] E. Bareinboim, J. Pearl, Causal inference and the data-fusion problem, *Proceed. Natl. Acad. Sci.* 113 (27) (2016) 7345–7352. <https://doi.org/10.1073/pnas.1510507111>
- [62] E. Bareinboim, J. Tian, J. Pearl, Recovering from selection bias in causal and statistical inference, in: *Probabilistic and Causal Inference: The Works of Judea Pearl*, 2022, pp. 433–450. <https://doi.org/10.1145/3501714.3501740>
- [63] D. Geiger, T. Verma, J. Pearl, d-separation: from theorems to algorithms, in: *Machine Intelligence and Pattern Recognition*, 10, Elsevier, Amsterdam, 1990, pp. 139–148. <https://doi.org/10.1016/B978-0-444-88738-2.50018-X>
- [64] J. Rosenow, D. Gibb, T. Nowak, R. Lowes, Heating up the global heat pump market, *Nat. Energy* 7 (10) (2022) 901–904. <https://www.nature.com/articles/s41560-022-01104-8>
- [65] H. Adan, F. Fuerst, Do energy efficiency measures really reduce household energy consumption? A difference-in-difference analysis, *Energy Effic.* 9 (2016) 1207–1219. <https://doi.org/10.1007/s12053-015-9418-3>
- [66] E. I. S.B. Department for Business, Net Zero Strategy: Build Back Greener, 2021, Accessed: 2025-06-20, <https://assets.publishing.service.gov.uk/media/6194dfa4d3bf7f055071b1b/net-zero-strategy-beis.pdf>
- [67] E. I. S.B. Department for Business, Heat and Buildings Strategy, 2021, Accessed: 2025-06-20, https://assets.publishing.service.gov.uk/media/61d450eb8fa8f54c14eb14e4/6.7408_BEIS_Clean_Heat_Heat_Buildings_Strategy_Stage_2_v5_WEB.pdf
- [68] Ofgem, Energy Company Obligation (ECO), 2025, (<https://www.ofgem.gov.uk/environmental-and-social-schemes/energy-company-obligation-eco>). Accessed: 2025-06-19.

- [69] B.K. Nayak, S.K. Sansaniwal, J. Mathur, T. Chandra, V. Garg, R. Gupta, A review of residential building archetypes and their applications to study building energy consumption, *Archit. Sci. Rev.* 66 (3) (2023) 187–200. <https://doi.org/10.1080/00038628.2023.2193167>
- [70] P. Shen, H. Wang, Archetype building energy modeling approaches and applications: a review, *Renew. Sustain. Energy Rev.* 199 (2024) 114478. <https://doi.org/10.1016/j.rser.2024.114478>
- [71] N. Security, Department for Energy and Zero, Subnational Electricity and Gas Consumption Summary Report, Statistical Report, UK Government, 2024. Accessed: 2025-05-08, <https://assets.publishing.service.gov.uk/media/6763dd7ebe7b2c675de30820/Subnational-electricity-and-gas-consumption-summary-report-2023.pdf>.
- [72] BBC News, Ministers scrap green heating plans for new homes, 2023, Accessed: 2025-05-08, <https://www.bbc.co.uk/news/articles/c3e40pe9185o>.
- [73] A. Sharma, V. Syrgkanis, C. Zhang, E. Kiciman, Dowhy: addressing challenges in expressing and validating causal assumptions, *arXiv preprint arXiv:2108.13518*. (2021). <https://doi.org/10.48550/arXiv.2108.13518>
- [74] K.R. Popper, *The Logic of Scientific Discovery*, Routledge, London, London, 1934. <https://doi.org/10.4324/9780203994627>
- [75] A.C. Eggers, G. Tuñón, A. Dafoe, Placebo tests for causal inference, *Am. J. Pol. Sci.* 68 (3) (2024) 1106–1121. <https://doi.org/10.1111/ajps.12818>
- [76] J.M. Wooldridge, *Introductory Econometrics: A Modern Approach*, South-Western Cengage Learning, Boston, MA, 5 edition, Boston, MA, 2013.
- [77] B. D'Amico, Github repository, 2025, Accessed: 2025-04-02, https://github.com/bernardinodamico/Households_energy_wellbeing.