But why?
The need for a causal understanding of changes in energy use

Phil Grunewald
Department of Engineering Science
University of Oxford
Oxford OX1 3PJ, UK

Abstract

Change is at the heart of the energy transition. Technologies, buildings and people - all change. But why?

Some changes happen naturally, other changes are the result of deliberate or random interventions. Some result from centrally devised policies, others are personal choices.

For effective and deliberate change to be achieved, a sound understanding of what causes the change can be helpful in developing the right mechanisms, and to avoid wasting time and effort on ineffective ones.

The importance of 'understanding the causes of change' seems self-evident, but the practical implementation for studies that provide evidence of causes (rather than mere correlation) are challenging. Ambition is required. Even after the event, it can be difficult to attribute effects to causes. Was a demand reduction the result of price, policy or products?

We present findings from ten years of deliberate research design intended to get a handle on 'causes' in changing energy demand. Inspired by the work of Judea Pearl (Pearl and Mackenzie 2018), we present an iterative three step process to understand causation. 1) Causal model 2) Observation and 3) Intervention. The causal model creates a framework for the research design and helps to systematically hypothesise about causal pathways and confounding variables. It informs what experimental design is required and what variables need to be observed. In some cases deliberate and controlled intervention are needed to test and revise the hypothesis.

Examples of the successful implementation are presented for behavioural interventions (demand response), technology interventions (heat pumps) and market interventions (price elasticity).

This approach is currently being scaled up as part of the UK Energy Demand Observatory and Laboratory (EDOL), which establishes statistically robust control groups to act as counterfactuals for interventions, which are tested in matched laboratories. Only with such longitudinal data at sufficient scale can causes be understood.

Introduction

Correlation is not causation. This well-established truism is haunting many energy researchers with a desire to affect change. Cockerels crowing is highly correlated to sunrise, but that does not mean that cockerels cause the sun to rise. That is not a question of missing data. No matter how many sunrises are being observed and how strong the correlation to crowing is, these data will not permit any causal inference. A policy to ban cockerels – no matter how well enforced – will have little impact on sunrise.

The limitation of statistics to make any judgement on such a fundamental understanding of the world has been addressed in the work of Judea Pearl (ibid), who devised a system for causal inferences. This paper gives a much-abbreviated version of some of the potential use cases and applications in energy research.

The topic has serious implications for machine learning and the ability of AI solutions to 'reason' in a way that seems to come natural to humans, but was hitherto inaccessible mathematically. The right choice of data and research design could equip models to infer robust causal understandings, where thus far only correlations could be observed.

Approaches to understanding causation

Causes and blame

An instinct in wanting to understand causes is often related to a desire to attribute blame. This is not the intention of this paper. We are interested in understanding causes to understand how to change things.It is important to note that establishing a cause is in any case not sufficient to attribute blame. Causes can run in long and complicated causal chains. It may be possible to establish that Y caused Z, but Y itself may have been caused by X. Whether to blame X or Y for Z becomes a moral question and is certainly beyond the scope of this paper.

Counterfactuals

Counterfactuals are at the heart of understanding causation. A cause is what makes the difference between a world in which that cause is present and one where it is not (all else being equal). If - as in our case - only one world is available, the other has to be imagined as a counterfactual.

The human mind is exceptionally good at creating counterfactuals. We can imagine with ease what might have happened if we had taken different choices. Equally, we are able to imagine numerous possible futures. This is how humans make decisions: they weigh up the imagined futures arising from their choice options, often unconsciously.

The perceived ease of cognitive counterfactual creation can mislead into thinking that this is a simple thing to do. Only when trying to replicate the process with artificial means, does it become apparent how much skill is involved. Just like picking up a raspberry seems trivial, until one witnesses a robot clumsily failing at the task time and again, despite considerable engineering effort to master this 'simple' task, so does 'counterfactual creation' reveal its challenging features when trying to let artificial intelligence have a go. Humans still have the edge. Taking good decisions is difficult, because creating good counterfactuals is difficult.

To understand if carbon emissions cause climate change, counterfactuals can be created with climate models. Such models simulate global temperatures and weather events for any number of CO_2 concentrations. All but one of these are counterfactuals, because they did not actually materialise. So long as the one model with the actual CO_2 concentrations accurately predicts the observed effects, one can explore the counterfactuals to see what would have happened with a different CO_2 concentration and attribute the difference to the CO_2.

This approach can also be used to establish the probability of a specific weather event having been the result of climate change. The likelihood of the same weather occurring in a counterfactual world without climate change allows to quantify the probability of the event having been caused by climate change.

The mathematical notation is

$$P(weather\ event|climate\ change)$$

for the probability that a weather event occurred 'given' (the | symbol) that climate change is present and

$$P(weather\ event|no\ climate\ change)$$

for the probability for this even if no climate change has taken place. The difference in probability can be attributed to climate change, provided no other factors have changed.

Climate models contain physical laws, which have been measured and validated with great confidence, such as the interaction of solar radiation with CO_2, radiative forcing, absorption and emission rates from different surfaces, and related thermo-dynamics. They also contain estimations from less well-understood climate effects, such as El Niño-events, ocean currents and complex interactions with the biosphere. The counterfactuals therefore always carry a degree of uncertainty.

Energy models, and household energy models specifically, have had much less research effort devoted to them than climate models. They, too, consist of well understood physics, such as the energy ratings of appliances, the laws of thermo-dynamics for converting different sources of energy into various energy uses and associated losses. They do, however, also contain a component that is far more difficult to quantify: humans.

It is a popular platitude that people are at the heart of the energy transition. They are also at the heart of many of the changes that need to happen. They may - in the terms of this paper - have to play a causal role in bringing about change.

Building a reliable counterfactual model for what people would 'do' has thus far only been dared by agent-based modellers and economists, in both cases with modest success when it comes to predictions. Here we present a framework that could help with the establishment of more evidence-based counterfactuals for energy demand. It builds in large part on the work of Judea Pearl, who has developed the framework for causal inference in the field of artificial intelligence (Pearl and Mackenzie 2018). This paper applies this framework to the field of energy demand and proposes that it can be a useful tool to guide research design and help understand causation.

Causal model

The first step in establishing causation is to create a causal model. This sets out what the hypothesised causal pathways are and what the confounding variables are. It is a hypothesis about what causes what.

The simplicity of a basic causal model should not detract from the breakthrough they constitute for the understanding of causality. Prior to such models, statistics simply lacked the language or symbolism to speak about causation.

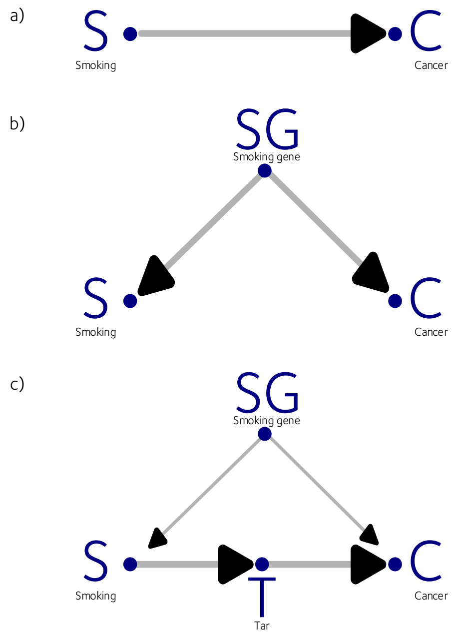

Figure 1: Causal model for smoking causing cancer.

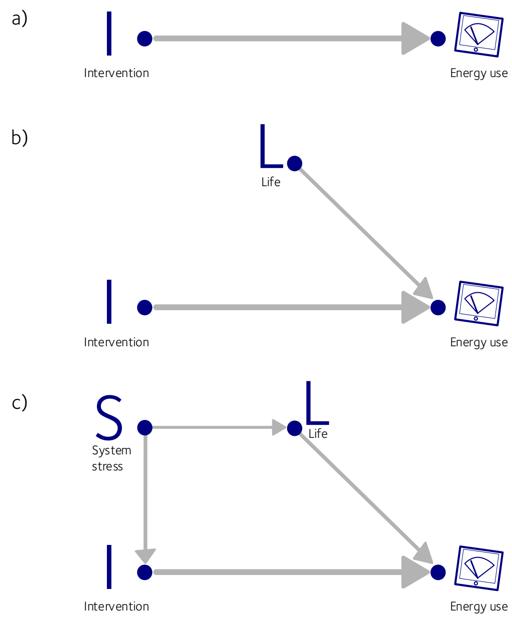

Figure 2: Causal models for demand response to an intervention. a) the direct impact of interventions b) life (and everything) as a confounding factor c) information leakage about system stress

A simple example from Pearl's Book of Why (Pearl and Mackenzie 2018) is the causal model for the effect of smoking on cancer. In Figure 1 (a) the arrow depicts the causal direction. The arrows in the diagram can be read as 'S causes C' in the direction of the arrow. Smoking causes Cancer.

It can also be helpful to think of the implicit logic as 'C listens to S'. The latter is especially useful in cases where multiple sources 'speak to C'.

However, the notion that smoking causes cancer was rebutted by the tobacco industry with the invention of a smoking gene. The argument was that smoking and lung cancer are only strongly correlated, but since correlation is not causation, the cause of *both* might be this smoking gene. People with such a gene - so it was suggested - are prone to take up smoking and are also prone to develop lung cancer. It is not smoking that is causal, it is this gene. Whether the existence of such a gene could be proven is immaterial. The mere possibility that it exists was sufficient to give the tobacco industry a narrative to claim that it cannot be proven that smoking *causes* cancer.

The causal model came too late for many lung cancer victims, but Figure 1 (b) shows the causal model as proposed by the tobacco industry. The smoking gene is a confounding variable. It causes both Smoking and Cancer. No arrow between S and C is required in this assumed model.

The genius of causal models becomes apparent when examining potential causal pathways. If a connection could be found that links Smoking and Cancer, independent of any smoking gene, then the causal case would be made. The unethical and impractical approach in this context is the randomised controlled trial (RCT). Given that only a proportion of the population supposedly have the smoking gene, a random sample of people could be forced to smoke. The same proportion as the general population should develop lung cancer. That would confirm the smoking gene theory. With our present knowledge, we would expect that the share of cancer in a random sample, which is forced to smoke, would be higher than the general population of smokers and non-smokers (gene carriers and non-gene carriers). Such an RCT, had it been permissible, would have proven the smoking gene theory wrong.

What can be done in the absence of such an experiment? The missing link turns out to be tar. Victims of lung cancer can be inspected for tar in their lungs and this rate is compared with a sample of the healthy population, irrespective of whether they smoke of not. The new causal (sub-)model is 'Tar causes Cancer'. If the statistics confirm this link (which they do), then only the origin of tar needs to be attributed to close the causal connection. Even the tobacco industry would struggle to argue that the smoking gene can deposit tar in people's lungs. QED. Additional data can substitute for an RCT, if accompanied by a causal model.

The causal model has further uses, which would go beyond the scope of this paper. For 'back door adjustment' and other procedures to deal with confounding variables, see Pearl's Book of Why. (Pearl and Mackenzie 2018)

Translating this example to the field of energy demand, an often-debated issue are the co-benefits of efficiency measures. Improving the building fabric of homes, especially for vulnerable and fuel poor households, could bring about health benefits. However, the effect sizes and incubation times make it difficult and expensive to test the causal effect with statistical validity. Instead, one can look for the 'tar'. It is sufficient to observe improvements in indoor air quality following well defined interventions and compare these to a comparable sample in an equivalent population (without that intervention). The medical impact of air quality on health is already established in the literature, such that this study is sufficient to 'close the causal link' between retrofit measures and health outcomes.

Controlling confounding variables

The tar example above is a fortunate case, where the necessary data may already be available and causality can be established without getting up from one's desk, simply by applying it to the causal model. This is not always the case. Sometimes the counterfactual can only be established with a controlled experiment.

The importance of 'controlled' relates to the treatment of confounding variables. In the tar example, the confounding variable was the smoking gene, which was not measurable (the fact that it does not exist does not help with its measurability). Because the presence of tar is arguably independent of the gene, it was possible to exclude the gene from the causal model.

In energy research confounding variables are everywhere. Energy demand may change because of changes in:

- Weather
- Population size
- Behaviour
- Energy system
- Economy
- Information
- Attitudes
- Transport
- Cost of energy
- Technology
- Practices
- Food systems
- Health
- Policy
- Built
- TV or sport

Some of these are measurable and can be controlled for. Others are not. But even when confounders are known and measured, causal models can help to understand if and how they should be correcting for.

For instance, we want to establish if environmental attitudes are causal in people's capacity to provide demand response. But we already know that pro-environmental attitudes correlate with higher incomes. To correct the confounding effects of income, we rebalance the sample to give us a more even income distribution for each of the attitude categories. The result is that we introduce more low-income households into the pro-environmental group and their demand response potential appears reduced. Was this an improvement of the finding?
No. Because higher income groups tend to have higher baseline energy demand, they also tend to be more able to provide demand response. Correcting a perceived confounding variable has made the results less accurate than a random selection would have done to begin with.

Thus, for this type of comparison to yield robust results the observation period needs to span several heating seasons with varying degrees of cold. And once several heating seasons are being considered anyway, one can do away with weather correction altogether and rely on the (counter-intuitively) more reliable randomness of weather.

Longitudinal studies are therefore particularly valuable in this context. They also allow for change events that can be traced back in time. If a sample underwent the same intervention at different times and an effect can be observed for each subject to occur at the associated time, then such data can be supportive of the causal model, without the intervention having to be administered externally and randomly. Confounders that are functions of time, such as weather, technology progress or energy costs, can be eliminated this way.

Interventions and the 'do' operator

Interventions can correct for confounders –known or unknown– provided they are administered independently of any possible confounders. A way to ensure independence with unknown variables is genuine randomness.

If for a random sample of sufficient size and over a sufficient amount of time, the probability of energy use observations is given by:

$$P(energyuse)$$

then any change in this probability observed in a randomly selected sub-sample with an intervention, expressed as:

$$P\big(energyuse|do(intervention)\big)$$

is *caused* by that intervention. Note, the use of the word 'cause', rather than correlation. Given that all other possible causes are randomly distributed in the control group and the intervention group (the one with the 'do' operation), the difference in observation can only be attributed to the *do* operator.

The need for randomness (and dictatorial enforcement of the do-operator)

Maintaining randomness is harder in practice than it sounds in theory. Sample bias still allows for valid results, provided the bias is understood and results can be qualified accordingly ("only valid for households with a smart meter, which means an under-representation of tenanted properties…"). Nevertheless, any non-randomness in the application of the *do* operator is more problematic, because it could undermine the independence of the intervention with the confounders.

For example, a study seeks to establish whether the installation of a smart meter causes demand reduction. Energy use is compared between a group that is randomly selected to be forced to have a smart meter and households that are not. If the randomly selected households are not forced to have a smart meter, and some of them can refuse, the causal model cannot be tested, because a possible confounder was able to sneak in: 'attitudes to smart meters'. A reluctance to have a smart meter may correlate with other possible causes of demand reduction, which are not the smart meter per se. Therefore, the demand reduction of the remaining sample of 'willing smart meter adopters' cannot be used as evidence for smart meters causing demand reduction. The *do* operator needs to be enforced with dictatorial authority to yield valid results. This may not always be possible, desirable or ethical. Additional data with a suitable causal model, as in the tar or indoor air quality example, would be desirable to establish causality in such cases.

A framework for causal research design

Based on the above observations and reflections, a simple framework for causal research design can be proposed. It consists of three steps:

1. Causal model - states what causes are assumed and what confounding variables are known or suspected
2. Observation - a counterfactual observation to provide the baseline for the intervention
3. Intervention - a randomised intervention to test the causal model

Additional criteria that can assist with the validity of the results are:

- randomness of sample (within known selection criteria)
- additional explanatory variables
- scale to suit the effect size
- longitudinal observation to correct for temporal confounders

In the following Section, this framework is reviewed in the context of three areas of energy demand research: demand response, heat pumps and price elasticity.

Causality in Energy

Demand response

A classic example of the need for a controlled experiment with a strong counterfactual is demand response. Much has been speculated about things people may or may not do under different conditions - often with more conviction than empirical evidence. Time use profiles have been studies for potential synchronicity of activities, which could lead to inferences about potential flexibility of practices. Variations in load profiles have been equally assessed for potential clues to flexibility indicators. Economists postulate that price elasticity alone could help to predict the future responses to price changes.

Causal model

Causal models with different assumptions about confounders are shown in Figure 2. In each case an intervention, such as a price signal or other information, is assumed to be the cause of changed in demand.

But changes in demand also occur without interventions. The response is subject to a range of confounding variables, including all the vagaries of everyday life, as shown in Figure 2 (b).

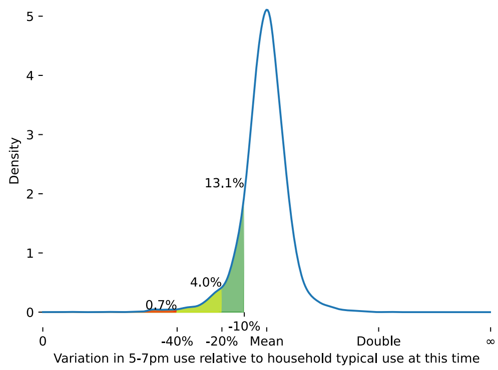

Figure 3: Probability distribution of load reduction 'by chance'

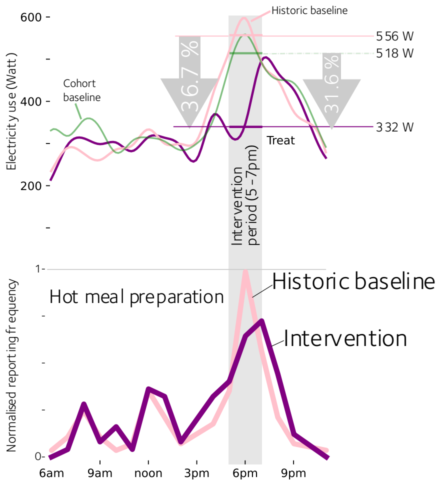

Figure 4: Demand response and activity response to intervention. The effect size differs for cohort and historic baselines.

Figure 3 shows the variability of demand at the same time of day (5pm to 7pm) for UK households. Without intervening there is a 13.1% chance that a household uses 10% less than their average electricity use on any given day, simply as a result of variability.

When attempting to separate the uncontrolled flexibility from the deliberate response as a result of an intervention, the intervention needs to result in a sufficiently large shift of the distribution across the sample.

To establish causality between intervention and response, the above framework can be applied using the causal model in Figure 2. We do this here merely to illustrate the real-world challenges with a seemingly straight-forward framework.

A possible intervention is to subject a group of households to time-varying prices and compare their response to a control group, which is not subjected to the same prices, but is in all other regards the same. This is why data from people who opt into a time-varying tariff does not yield valid results. They are not random and exhibit features (such as the willingness to participate in such a tariff) that may be related to the causal pathway under investigation.

A way to randomise such a sample is to take all the people who have opted into a time-varying tariff and randomly deny some of them their chosen tariff. Now the two groups can be compared and the observed differences in demand can be attributed to the tariff, at least for this subset of the population. In practice (and in research ethics) this approach may be problematic.

Pragmatically, one may have to accept that findings are not generalisable to the whole population, but only to the subset of the population willing to participate in such a tariff. This is a limitation of the research design, but not of the causal model. The causal model is still valid, but the findings are only applicable to a subset of the population.

Observation - establishing the baseline

Comparing the effect of the intervention to a counterfactual can be done in two ways. Historic comparisons compare each participant with themselves. What did their load profile look like on comparable days? For instance, a response on a Thursday could be compared with previous Thursdays in similar seasons. A common approach is to compare with 10 recent working days. If the historic data dates back to before the intervention is announced, this approach also safeguards against gaming, where a participant may deliberately inflate the baseline, to make a load reduction appear larger.

The second approach is a cohort comparison, where load is compared against non-participants at the same time. This approach is especially helpful to mitigate any temporal confounders, but could introduce errors arising from 'information leakage' (Figure 2 (c)). If non-participants become aware of the need to reduce demand at certain times, they may do so, regardless of rewards or being part of a trial. People might talk to each other, which is desirable in general, but difficult to 'control' in an experimental setting, or national media informs the wider population about the need to reduce demand at certain times.

While these types of information could affect the results, they do not undermine the validity of the causal inference, so long as the model accounts for them. In this case the causal assumption is not for the load reduction in general, but the *additional* load reduction, above and beyond that caused by the public information, correctly eliminated by being present in both intervention and control group.

Intervention – eliminating confounders

In the case of demand response, the intervention is sensitive to timing in ways that are difficult to randomise. In practice, demand responses are needed most at times of high demand. It is therefore reasonable that an intervention focusses on such times, which in the UK tends to be cold winter evenings between 5pm and 7pm. The UK system operator conducted a large-scale trial, offering rewards (via suppliers) for households that reduce demand during critical times. Households opt into trials to qualify for the rewards. But people who do not take up the offer may still be aware of the 'need to help out' and take actions (consciously or unconsciously) that result in demand reductions.

This causal pathway is depicted in Figure 2 (c). The intervention is no longer truly random, but 'listens' to system stress events. These same events may be 'heard' by participants inside or outside the intervention group. This could confound the resulting demand observation. When treated properly in the causal model, such effects can also yield additional insights into demand response mechanisms and improve the baseline for the pure intervention response.

Grunewald (2023b) conducted trials in which participants were asked to reduce demand between 5pm and 7pm on a specified day with a week's notice. On the intervention day and on a control day, participants also had to complete a diary of activities, such that the causes of any load shift or reduction could be better understood. The results are shown in Figure 4. The load reduction is evident against both the historic and the cohort baseline metrics. Furthermore, the activity records allow to link the load reduction to specific activities, such as hot meal preparation. This is indicative, but not yet a causal connection with cooking appliances, such as ovens. The explicit inclusion of individual appliances would illuminate the causal pathways of load reduction further.

Heat pumps

Heat pumps are expected to play a major part in the decarbonisation of the heat sector in the UK. Unlike many other interventions, installing a heat pump is not a drop-in replacement, but may necessitate wider changes to the building's heating system and fabric.

Many variables could affect the performance of the heating system and a causal model for the impact of a heat pump on energy use is illustrated in Figure 5. Some confounders, like the weather, are unambiguously independent of the heat pump installation (aside from their potential impact on mitigating climate change). Others are part of the 'heat pump process' and may help to understand the causality in changes to energy use, resulting from the presence of a heat pump, compared to alternative heating solutions. These could be broken down in great detail and the label 'Heating practices' serves as a broad umbrella term a broad range of factors.

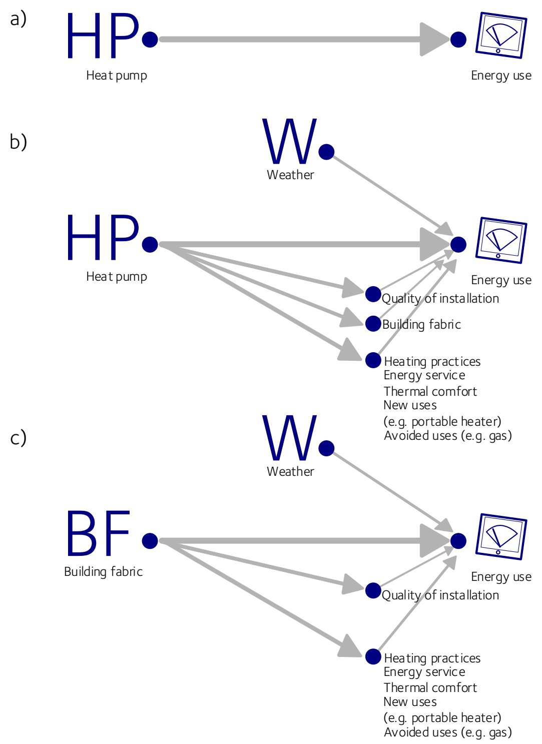

Figure 5: Causal models for the impact of heat pumps on energy use

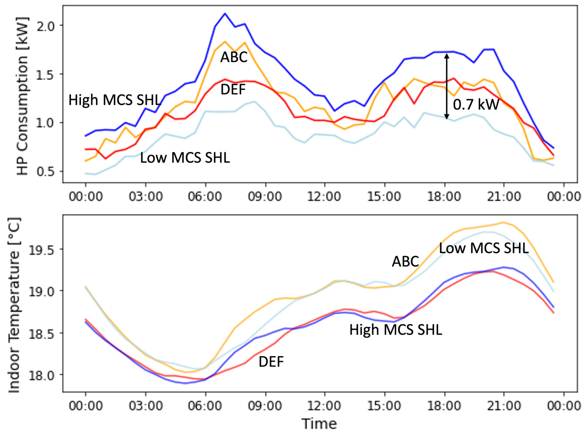

Figure 6: Causal (?) difference between homes with high and low heat loss. Source: Perelli-Rocco (2023)

Figure 6 shows the results from Perelli-Rocco (2023), were data was collected prior to and independent of the research question being formulated: do building characteristics affect heat pump consumption?

Here the heat pump is not the cause of the difference, but a given. All sample members have a heat pump, such that a comparison can be drawn between them, based on the variable ‘building fabric’, which in this case was approximated as the heat loss value.

The heat loss value is an observation, rather than an intervention. If households could be assumed to be randomly assigned buildings with different heat loss values, which depend on no other variables, then any difference observed in energy use between homes with different heat loss values could support the causal model. However, as illustrated in Figure 6, the building fabric could affect heating practices, which themselves affect the energy use.

Higher heat loss results in lower indoor temperatures as shown in the bottom of Figure 6. The fact that both EPC ratings and heat loss result in similar temperatures, but different energy use, suggests that the heat loss values do have a causal effect on energy use for homes with heat pumps, even though other confounders may not have been corrected for, due to the lack of a deliberate and controlled intervention.

Price elasticity

Macro-economically, the price elasticity of energy demand is not contested. When global energy prices rise, many businesses take economically rational decisions to scale back operations, relocate or change processes. Espey and Espey (2004) estimate the short-term elasticity of electricity demand to be -0.35 and -0.85 in the long term.

Such figures are widely used in models and simulations of demand response (very short term) and electricity system models (long term) (Bradley, Leach, and Torriti 2013; Roscoe and Ault 2010). Numerous studies state significant short-term responses to price-based demand response trials (Schofield et al. 2014; Torriti and Yunusov 2020). However, Zhu et al. (2018) conclude after extensive meta-analysis of international reviews that residential electricity demand is almost inelastic in the short term.

Household demand underlies fundamentally different dynamics to commercial and industrial loads. Households cannot suspend or relocate operations. Life has to go on, even when energy is more expensive. The causal model for the effect of high electricity prices on household demand may therefore be different.

To some extent, households may be able to temporarily scale back less essential operations or sacrifice certain comforts, such as heating (see Fawcett, McKenna, and Grunewald (2024)). Other loads, such as fridges, offer less opportunity to scale back, especially for households with limited disposable income, for whom the purchase of a new, more efficient fridge is not a viable short-term response to save money.

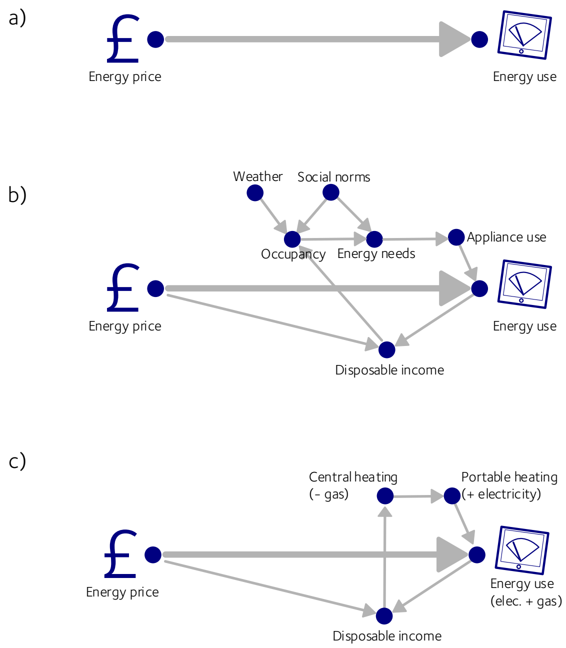

Figure 7: Causal models for energy use in response to price change. a) the direct economic causality, b) Giffen good response, c) counterproductive substitution.

To test the causality of price changes affecting energy demand, a causal model as shown in Figure 7 (a) needs to be tested. To correct for numerous confounding variables (Figure 7 (b)), both sample and intervention need to be randomly assigned. The intervention being the price of electricity, this poses regulatory challenges for experimental research.

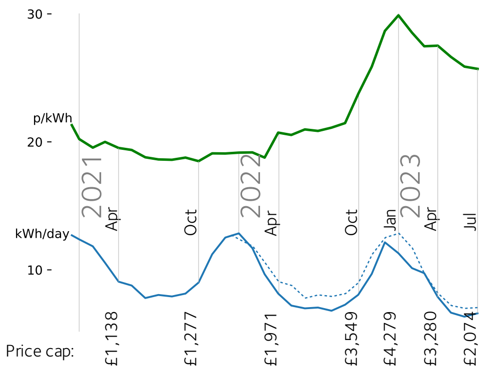

Figure 8: Demand and prices over time. Dotted line is a repeat of the previous year for reference.

In the UK, prices are set by suppliers and customers have free choice over suppliers and tariffs. To test the effect of price changes on demand in a controlled trial, this mechanism would need to be suspended for the experiment, which is not practical for research settings. However, this is in effect what happened during the price crisis in 2022–23. All users experienced price rises, followed by a price cap. The variation in price and demand is shown in Figure 8 as a time series, which suggests that demand fell, while prices rose. A classic case of correlation, which would seem to confirm negative price elasticity. However, electricity demand in the UK has fallen for decades, irrespective of price changes. This can be attributed to efficiency gains, policies, behavioural changes or other unknown factors, all of which need to be considered as potential confounders.

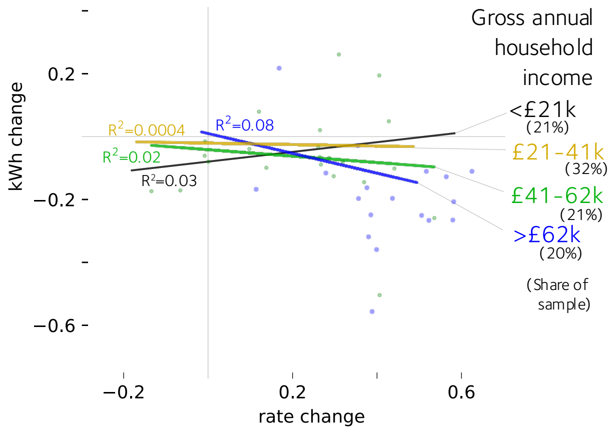

Figure 9: Price elasticity of energy demand is not borne out by 2022-2023 data. A new causal model is needed.

Everyone was affected by price rises to some extent. This ‘experiment’ did not have a randomly selected control group of households that were exempt from the price rises. And yet, the data in Figure 9 shows that households were affected to different degrees and at different times. A more detailed analysis of household-by-household elasticities shows a more mixed and inconclusive picture of price elasticity, which may affect different income groups differently (Grunewald 2023a).

These findings allow to refine the causal model, such that future observations can test new, plausible relationships. Two suggested dynamics are shown in Figure 7 (b) and (c). The first is a Giffen good, where the price rise leads to a reduction in disposable income, which may reduce the ability to go out (an unknown relationship). As home occupancy increases, so does energy use (a known causal relationship). To test the Giffen good hypothesis, occupancy and ideally individual appliance use would need to be included in the observation.

The second proposed dynamic is a counter-productive substitution. Fawcett, McKenna, and Grunewald (2024) observed significant reduction in gas use for pre-payment meter customers. It is suspected that some of them may have switched to electric portable heaters. The causal model in Figure 7 (c) could be tested if gas use and appliance level electricity, ideally in conjunction with indoor temperatures, were observed.

Discussion

The framework for causal research design, proposed above, has been illustrated on three areas of energy demand research: demand response, heat pumps and price elasticity. Table 1 summarises the differences between these three cases.

	DSR	Price elasticity	Heat pump
Research design	Proactive	Responsive	Ex-ante 3rd party data accessed
Intervention	Direct participant engagement	External events (price)	None
Sampling	Random with (un)known biases	As DSR	Heat pump owners only
Causal model	A priori and confirmed	Challenged and refined	Validated and refined
Explanations	Activities	Income	Heat loss
Observation	35% reduction	weak elasticity	700W peak reduction
Control	Self (longitudinal) and cohort	Self (longitudinal)	None
Model refinement	Load disaggregation, agency	Occupancy, energy service	Energy service change

Table 1: Case study comparison (DSR: demand side response).

Out of the three trials, the demand response trial is most consistent with the proposed framework. The causal model had been defined a-priori, hypothesising that a demand reduction request would result in a measurable demand reduction. The explanatory variables from survey and activity records allow to test the causal model with more confidence. Additional scale and longitudinality would further strengthen the generalisability.

The price elasticity trial is a convenience trial, using the same data source as the DSR trial. The energy price crisis was unanticipated and the sample was not designed specifically with the intention to test changes to major changes to energy prices. In the event, the data was suitable for analysis, as it contained some, but not all, of the desirable explanatory variables, such as half-hourly energy used, half-hourly tariffs, survey information, including income and ability to pay bills, but no explanatory data that would allow conclusions about behavioural responses during the high-price period. Nevertheless, the analysis has been sufficient to challenge the original causal model and helped to formulate refined causal models, which can be tested in more dedicated future studies.

The causality test of heat loss on heat pump energy performance, did not have a dedicated research design at the outset. The data was used as published and the analysis relied on the configuration of the living lab. The careful design of their study and open access to the observational and contextual data has been sufficient to test a causal model, but with obvious limitations. Numerous potential confounders could affect the heat pump performance alongside the heat loss of the building itself. The analysis performed is therefore a necessary, but not a sufficient basis on which to conclude causality. As with the price elasticity, the process of devising and testing a causal model can inform the next steps and additional variables needed to reach more conclusive causal results.

Conclusions

Establishing causality would be a major advance for many areas in energy research where thus far it was only possible to observe correlations. A better understanding of causation would strengthen the hand of policy makers and lead to better targeted interventions. The proposed research design is more challenging and costly, as it requires scale and longitudinal observations. In some cases it may even be ethically or practically impossible, but even the development of causal models itself can assist in identifying the unknowns and confounding variables.

A framework to establish causation in energy demand research has been proposed and illustrated with three case studies, which met the requirements to differing degrees.

The key research design components for causality are: 1) A causal model 2) a sufficiently random sample to establish a reliable baseline and 3) an intervention or observation that is administered independently from any potential confounding variable.

Additional contextual variables can help validate the causal model against residual confounders. What all cases had in common is that sufficient scale and longitudinal observation are important for the reliability of causal inferences.

The case studies have shown that deliberate research design is needed for causality to be established. In some cases, it may be possible to infer causality from existing data, provided sufficient explanatory variables have been collected alongside the observation of interest.

The requirements for experimental designs with sufficient scale and longitudinally are challenging for many budget-constrained research projects. Scalable, low-cost solutions, which allow to continuously observe and store diverse types of variables, are needed to build reliable baselines against which interventions and change can be observed.

The Energy Demand Observatory and Laboratory (EDOL) is currently developing such approaches to build a public observatory of energy demand data, which can be used for research in its own right, or act as the counterfactual for smaller studies that focus on the intervention part and need controls to assess the effect size against a random ‘unintervened’ sample.

References

Bradley, Peter, Matthew Leach, and Jacopo Torriti. 2013. "A review of the costs and benefits of demand response for electricity in the UK." *Energy Policy* 52 (0): 312–27.

Espey, James A, and Molly Espey. 2004. "Turning on the Lights: A Meta-Analysis of Residential Electricity Demand Elasticities." *Journal of Agricultural and Applied Economics* 36 (1): 65–81.

Fawcett, Tina, Eoghan McKenna, and Phil Grunewald. 2024. "Crisis Ready - How Longitudinal Data Helps to Make Sense of Crises and How to Prepare for the Next One." *ECEEE Summer Study Proceedings* 4-146-24.

Grunewald, Phil. 2023a. "Energy use in crisis - lessons for net-zero?" BIEE 14th Academic Conference. https://edol.uk/News/Output/23_BIEE/biee_paper_phil.pdf.

Grunewald, Phil. 2023b. "Peak demand reduction – who is flexibility, when and how?" *BEHAVE 2023 - the 7th European Conference on Behaviour Change for Energy Efficiency*, 403–17. https://enr-network.org/wp-content/uploads/Proceedings_BEHAVEconference_2023_def-1.pdf.

Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. 1st ed. USA: Basic Books, Inc.

Perelli-Rocco, Sofia. 2023. "How building characteristics affect heat pump consumption." MSc Thesis, https://edol.uk/Research/Labs/Heat/Sofia_MSc_Thesis.html: University of Oxford.

Roscoe, A. J., and G. Ault. 2010. "Supporting high penetrations of renewable generation via implementation of real-time electricity pricing and demand response." *Renewable Power Generation, IET* 4 (4): 369–82.

Schofield, J., R. Carmichael, S. Tindemans, M. Woolf, M. Bilton, and G. Strbac. 2014. "Residential consumer responsiveness to time-varying pricing." Report A3 for the ``Low Carbon London'' LCNF project. Imperial College London.

Torriti, Jacopo, and Timur Yunusov. 2020. "It's only a matter of time: Flexibility, activities and time of use tariffs in the United Kingdom." *Energy Research & Social Science* 69: 101697. https://doi.org/10.1016/j.erss.2020.101697.

Zhu, Xing, Lanlan Li, Kaile Zhou, Xiaoling Zhang, and Shanlin Yang. 2018. "A Meta-Analysis on the Price Elasticity and Income Elasticity of Residential Electricity Demand." *Journal of Cleaner Production* 201: 169–77. https://doi.org/10.1016/j.jclepro.2018.08.027.