Title
The effectiveness of US energy efficiency building labels (vol 2, 17033, 2017)

Journal
NATURE ENERGY, 2(4)

ISSN
2058-7546

Authors
Asensio, OI
Delmas, MA

Publication Date
2017-04-01

Peer reviewed

nature energy

ARTICLES

PUBLISHED: 27 MARCH 2017 | VOLUME: 2 | ARTICLE NUMBER: 17033

The effectiveness of US energy efficiency building labels

Omar Isaac Asensio[1,2] and Magali A. Delmas[3*]

Information programs are promising strategies to encourage investments in energy efficiency in commercial buildings. However, the realized effectiveness of these programs has not yet been estimated on a large scale. Here we take advantage of a large sample of monthly electricity consumption data for 178,777 commercial buildings in Los Angeles to analyse energy savings and emissions reductions from three major programs designed to encourage efficiency: the US Department of Energy's Better Buildings Challenge, the US Environmental Protection Agency's Energy Star program and the US Green Building Council's Leadership in Energy and Environmental Design (LEED) program. Using matching techniques, we find energy savings that range from 18% to 30%, depending on the program. These savings represent a reduction of 210 million kilowatt-hours or 145 kilotons of CO_2 equivalent emissions per year. However, we also find that these programs do not substantially reduce emissions in small and medium sized buildings, which represent about two-thirds of commercial sector building emissions.

Energy supplied in buildings accounts for an estimated 8.8 gigatons of CO_2 emissions globally or about one-third of total energy use and carbon emissions[1]. The United Nations Environment Program (UNEP) and many energy experts argue that the buildings sector has the largest potential for delivering long-term and cost-effective emissions reductions in both developing and developed countries[2]. A recent analysis by the National Research Council contends that the full development of cost-effective energy efficiency technologies in buildings could eliminate the need to construct new electricity-generating plants in the United States[3]. A critical question is what kind of programs can catalyse reductions in emissions. This question is especially important given the current lack of global carbon regulation. In the United States, there are three major voluntary information programs aimed at reducing building emissions: The US Environmental Protection Agency's (EPA) Energy Star Program, the US Green Building Council's (USGBC) Leadership in Energy and Environmental Design (LEED), and the US Department of Energy's (DOE) Better Buildings Challenge. Participation in these programs has increased rapidly over the past 10 years and has reached 21 billion square feet (sqft) of floor space (see refs 4–6 and Supplementary Note 1). These programs aim to encourage private investment in energy efficiency. Examples of such investments include structural upgrades for indoor heating, ventilating and air conditioning (HVAC); smart energy management systems; and efficient lighting, sensors and other controls.

Information programs reduce barriers to investment and encourage energy efficiency through two main mechanisms. The first mechanism involves lowering search and information costs for energy planning decisions. This often includes subsidized building audits that provide tailored information about potential savings through available technologies, and benchmarking of best practices through a network of peers. This is a main focus of the DOE Better Buildings Challenge, which provides energy audits to support US commercial and industrial building owners who commit to reducing energy and water consumption in existing buildings by 20 percent or more over 10 years[7]. The program provides public recognition for performance but it does not offer a separate certification label. A recent meta-analysis of peer-reviewed studies in energy conservation found that technical audits, such as those provided to many Better Buildings Challenge partners, were effective to reduce energy consumption in the residential sector[8].

A second mechanism by which information programs can promote voluntary energy efficiency adoption involves market signalling through a prominently displayed energy efficiency label. Labelling is the focus of both the LEED and Energy Star programs, which provide third-party certification for efficient buildings based on a comparative 1–100 Energy Star score. Only those buildings that receive an Energy Star score of 75 (75th percentile or better) compared to similar buildings nationwide are eligible to apply for the Energy Star or LEED certification label in a given year. Unlike Energy Star, which is a government supported label for energy efficiency, LEED is privately supported. The USGBC rates LEED buildings based on a tiered rating scheme, which includes reductions in energy use, but also focuses on improvements such as water use, materials and resources, indoor environmental quality, and sustainable design.

Each program has unique institutional features (see Supplementary Note 1), but largely attracts different segments of the commercial real estate market, with minor overlap in participation. For those buildings not eligible to participate in either Energy Star or LEED certification programs, the Better Buildings initiative provides a market entry point for energy efficiency investment and participation in existing buildings.

These programs are often described as green clubs, in which voluntary participation provides reputation benefits to its members[9]. Building owners and managers who participate in these programs often gain recognition for their more efficient buildings through market mechanisms that sometimes include premiums such as increased asset prices and tenant rents[10–13]. These

[1]Georgia Institute of Technology, School of Public Policy, 685 Cherry Street NW, Atlanta, Georgia 30332, USA. [2]UCLA Institute of the Environment & Sustainability and Anderson School of Management Ziman Center for Real Estate, La Kretz Hall, Suite 300, Los Angeles, California 90095-1496, USA. [3]UCLA Institute of the Environment & Sustainability and Anderson School of Management, La Kretz Hall, Suite 300, Los Angeles, California 90095-1496, USA. *e-mail: delmas@ucla.edu

Table 1 | Annual building emissions by building class (2005–2012).

	Number of buildings	Square footage (in million sqft)	Metric tons of CO_2	Percentage of building emissions	Square footage in Energy Star, LEED or LABBC (in million sqft)	% of square footage in Energy Star, LEED or LABBC
Class A	456	107.7	585,410.71	36.4	66.98	91.04
Class B	3,452	125.5	437,936.05	27.3	4.77	6.48
Class C	14,698	238.3	582,999.95	36.3	1.82	2.47
Total	18,606	471.5	1,606,346.71	100	73.57	100

Buildings located in Los Angeles Department of Water and Power territory.

economic returns reflect expectations of lower energy costs for building occupants. However, while the evidence on green premiums has received increased attention, the realized environmental performance of these investments is as yet largely undetermined. While there have been program estimates associated with buildings participating in these programs, these estimates have not been able to clearly isolate the effects of each program.

One principal limitation of such analyses has been the lack of access to longitudinal high-resolution building energy performance data[14,15]. Another difficulty for program evaluation arises in the fact that participating firms in voluntary programs seldom constitute random samples. We also never directly observe the alternative states among participants in which an investment or participation decision is not made, making it difficult to construct valid control groups for program evaluation[16]. Further, cross-sectional analyses of programs can be misleading because of endogeneity issues, which can be due to a number of reasons, including technology adoption, pricing and consumer preferences, all of which potentially limit our ability to make causal inferences.

Recently, scholars have called for better research designs and rigorous observational studies that utilize both pre- and post-performance data[14,17]. They have argued that the most rigorous approaches in program evaluation are to use randomized controlled trials[17–19] and quasi-experimental research designs[20,21]. Experiments are useful, but are often cost prohibitive or infeasible[22,23]. In such cases, matching strategies, particularly those used in combination with other methods, can be used to mitigate covariate imbalances in models based on observational data[24–26].

In this Article, we perform a comparative analysis of the effectiveness of energy efficiency labelling strategies in the Better Buildings Challenge, Energy Star and LEED programs in the commercial buildings sector, enabled by access to monthly electricity consumption data for all commercial buildings in Los Angeles from 2005–2012[27]. We find that all these programs deliver high magnitudes of energy savings that range from 18% to 30%, depending on the program. These savings represent a reduction of 210 million kilowatt-hours (kWh) or 145 kilotons of carbon dioxide equivalent (CO_2e) emissions per year. Due to the long lifespan of buildings and retrofits, these savings are likely to persist, particularly in larger, more energy-intensive buildings. However, due to eligibility rules and participant self-selection, we find that current information programs do not substantially target emissions reductions in small and medium sized buildings, particularly in the 75th percentile and below by consumption, which represents up to two-thirds of commercial sector building emissions and the long tail for greenhouse gas mitigation efforts from building efficiency improvements.

Program evaluation overview

Our data set includes 178,777 buildings with 16.5 million panel observations. Because participation is not randomly assigned, we use matching strategies to compare the performance of participating buildings against similar buildings that are not part of these programs. In this way, we control for overt sources of bias due to systematic differences between participating and non-participating buildings, which affect evaluation outcomes, and then we test the sensitivity of our estimates to hidden bias. Matching strategies mimic randomization by identifying a comparison group of buildings that is statistically similar to treated buildings, based on observable characteristics. We use computational advances in matching algorithms to match buildings on a more comprehensive set of characteristics than previous literature. In doing so, we quantify an important source of evaluation error when estimating emissions reductions.

Los Angeles is an ideal setting to study energy efficiency for three reasons. First, unlike many other cities that may be at earlier stages of adoption, we can already observe significant participation in these three programs simultaneously during this period for a total of 192 million sqft of commercial floor space in 254 buildings. Second, Los Angeles is the largest market in the US for green building investments, and is often considered a model for other cities[28]. Third, we have access to high-resolution data at the building level, which allows us to go beyond simulations or predictive modelling to assess emissions reductions[29].

Characteristics of building participants

Participating buildings for Better Buildings Challenge, Energy Star and LEED certified buildings in Los Angeles are generally larger and more energy intensive (in kWh per month and kWh per sqft) than non-participating buildings. Participating buildings are also more likely to have been renovated, which is to be expected as building owners and managers often consider capital investments for energy efficiency during periods of renovation. These differences are significant both in means and distributions from the general population (Supplementary Table 1). Thus, a simple comparison of mean outcomes for participating and non-participating buildings is unlikely to yield accurate estimates of the causal effect of program participation. For example, during the period from 2005–2009, participating buildings are significantly more energy intensive (1.134 kWh per sqft) than an average non-participating building (0.893 kWh per sqft). Participating buildings are also commonly designated as Class A buildings, the more coveted and higher quality real estate assets, and to a lesser extent, Class B or Class C buildings, which indicates positive selection. Buildings may be classified as A, B, or C in descending quality based on such parameters as desirability of location, age of building, building infrastructure and maintenance. Building class designations are subjective ratings used by real estate professionals to gauge building quality, and may vary from market to market. For example, Class C buildings are only about 2% by square footage in participating buildings (Table 1). However, Class C buildings, which most often represent smaller, ageing buildings, still account for a substantial 36.3% of commercial sector emissions or 583 kilotons of CO_2 emissions in Los Angeles. These baseline differences suggest that counterfactual strategies based on a comparison group of average non-participating buildings would be ineffective reference groups versus more rigorously matched controls.

Descriptive statistics also reveal significant differences in building characteristics between programs (see Supplementary Table 2). For example, building construction year, renovations and quality

Table 2 | List of balancing characteristics used in matching.

Observable building characteristic	Data source
Physical building characteristics	
Year built	CoStar/public record
Year renovated	CoStar
Building location/climate	
Climate zone	Public utility/NOAA
Occupancy characteristics	
Rentable building area (square footage)	CoStar/public record
Property type	CoStar
Occupancy rate (percentage leased)	CoStar
Building quality	
Building class	CoStar
CoStar rating*	CoStar
Industry characteristics	
SIC industry code	Public record
Utility customer class	Public utility
Building operating expenses	
Average rent	CoStar
Taxes per sqft	CoStar

*The CoStar building rating system is a national rating system for commercial buildings, which captures a number of characteristics including architectural attributes, structural and systems specifications, amenities, site and landscaping treatments and detailed property type specifics. Ratings reflect commercial real estate quality as valued by investors.

ratings all differ substantially between programs. These differences in participant profiles before matching reveal both different targeting strategies by program managers and administrators, and self-selection into the respective programs. See Supplementary Note 2 for a more detailed discussion on observable bias.

Matching algorithms

A few studies evaluating building performance data have used matching procedures based on propensity scores to control for overt bias, or the fact that the treatment and control groups differ in ways that matter for the outcomes under study[10,30]. These studies typically use a single covariate based on building location (for example, proximity, or linear distance) to enable comparisons between buildings[30]. The main identification assumption, although largely unproven, is that buildings close to each other are more similar to buildings that are far away. However, matching buildings on a single distance measure does not address two important potential sources of selection bias. The first is due to remaining imbalances in other relevant covariates, which can bias estimates; and the second is due to the lack of sampling density in the region of the common support, which is often prevalent in finite samples[16]. In our review of the literature, few published studies report the degree of covariate imbalance in matching studies with observational data, and none that we are aware of in the energy efficiency literature.

In our analysis, we match on an expanded set of covariates compared to those previously available in the literature. We use several matching strategies to enable performance comparisons—including genetic matching, which uses a search algorithm to automatically find the optimal covariate balance in the reference group[31–34]. See Methods for more details. Our reference group consists of the universe of all commercial buildings in the service territory of the Los Angeles Department of Water and Power (LADWP), the nation's largest municipal utility. This includes 56 neighbouring cities and 1.4 million customers. We match buildings on 12 characteristics found in the literature to affect building energy consumption (Table 2). These include: location (climate zone); physical building characteristics (square footage, year built, year renovated); occupancy (percentage leased, tenant type); building use type (property type); industry characteristics

Table 3 | Program energy savings.

Program	Average treatment effect	Std. Err. (Abadie–Imbens)	*P* value	N matched observations
LABBC	−18.69	10.95	0.09	35,939
LEED certified	−29.99	12.06	0.06	35,439
Energy Star	−19.31	5.81	0.02	35,416

Estimates using matching procedures with weather and time controls.

(Standard Industrial Classification (SIC) industry code, utility customer class); building operating expenses (average rents, taxes per sqft); and building quality (building class). We also include the CoStar analyst ratings (scored from 1–5) to mitigate hidden bias and capture other unobserved characteristics quantitatively. The CoStar rating is a national rating scheme for commercial buildings that considers a combination of factors typically unobserved by evaluators, such as building amenities, construction quality, architectural attributes, management, location/accessibility, systems standards and specifications, detailed property specifics and market factors. Using this approach, we believe that we substantially reduce observable bias arising from participant selection (see Fig. 1).

Energy savings of information programs

To evaluate the impact of participation in information programs on building energy savings (measured as the percentage energy change in kilowatt-hours (kWh) per sqft), we implement matching procedures and then conduct post-matching regressions to adjust for time variation on building energy use. In post-matching regressions, we include important seasonality and time controls, such as heating and cooling degree-days, to adjust for weather variation and any calendar shocks on consumption. See the Methods section for details. In Table 3, we report the final estimates of energy savings for each program in the City of Los Angeles. The estimates are robust to several matching procedures and specifications, which yield quantitatively similar results, and we report the most conservative estimates. The energy savings from the Los Angeles Better Buildings Challenge (LABBC) program is −18.69%, significant at the 10% level. These savings are the result of building technology upgrades identified through LABBC audits in 91 participating buildings totalling 35 million sqft of floor space. The most common building upgrades include HVAC systems (72%), lighting and controls (14%), and improvements in building envelope (6%). Other upgrades (8%) include deep renovations in pumping, ventilation and sensor technology. These building efficiency upgrades are primarily structural, although a few implemented projects include behaviourally informed changes such as data server optimization and computer power management. The savings for Energy Star and LEED programs are −19.31% ($p < 0.02$) and −29.99% ($p < 0.06$), respectively, over the period 2005–2012. We find that building efficiency investments across all three programs show significant progress towards long-run environmental policy goals of 20% savings over ten years.

Across 125.9 million sqft of total participating floor space in the three programs, this is an annual reduction of 210.2 million kWh of city energy use. Using EPA (eGrid2012) emissions factors based on LADWP's local electricity mix, the savings amount to 145 kilotons of non-baseload CO_2 emissions per year. To put these numbers in context, the savings from Los Angeles commercial sector building improvements are the equivalent of burning 70.6 kilotons of coal each year. We contrast the magnitudes of these savings from capital upgrades versus behavioural intervention programs commonly employed in the residential sector, which yield significantly lower percentage savings by an order of magnitude, ranging from 2–3% for the highest-quality studies[8].

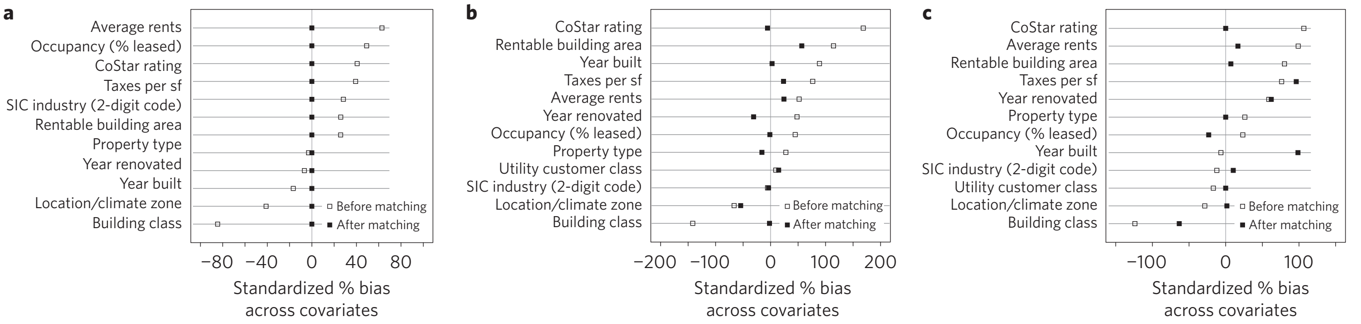

Figure 1 | Bias reduction in matched samples for the three energy programs. Reducing observable bias by nearest-neighbour matching with replacement for LABBC Buildings (**a**), Energy Star Buildings (**b**) and LEED buildings (**c**).

Reducing evaluation error

We also evaluated energy savings using more conventional regression-based methods (Supplementary Table 6, row 1). We find that without adjusting for covariate imbalances, the estimates of energy savings are overstated by at least 7 percentage points (for example, shifting point estimates of building energy savings from −18% to −25%). For a 50 million sqft city energy efficiency program for example, a 7 percentage point evaluation error due to an improperly specified reference group would be the equivalent of 47 million additional kWh per year or approximately US$5.2 million in utility bills incorrectly attributed to energy savings. These results confirm the potential benefits of matching in combination with other methods, especially as a means to mitigate bias resulting from how the evaluation is done and what buildings data are collected and observed.

We find that participating green buildings are not a random draw from the general population, but rather a self-selected set of buildings with observable characteristics, which clearly indicate higher quality (for example, positive selection). This is an important finding, and so by matching on a rich set of covariates, we reduce selection bias based on multiple criteria (see Fig. 1). We note that while matching is not perfect, we believe that we eliminate or substantially reduce bias, particularly for LABBC and Energy Star buildings, as the bias reduction by Building Class after matching is close to or equal to 100% (Supplementary Tables 3–5). See the Methods section for further discussion.

Our study is not without limitations. First, our estimates are internally valid for the sample period and geography analysed. Los Angeles is the largest market for green building investments and remains the second largest metropolitan area in the US by population. Our estimates may be indicative of anticipated performance in other cities and regions, but the actual returns will depend on specific market conditions, investment levels and building use profiles. We know, for example, that resource efficiency investments can be negatively affected during economic downturns[35]. Second, it is possible that a building makes efficiency investments without participating in one of these programs. Also, some efficient buildings could get certified without additional investments. However, this is unlikely because the programs have very specific requirements that would not normally be included during construction or renovation. Third, while electricity is the major energy source for commercial buildings in the US, natural gas is also an important source of energy consumption. It will be useful to evaluate impacts on natural gas consumption in future work.

Program cost-effectiveness

We are able to calculate program cost-effectiveness for LABBC participants, for which we have financial data reported to us by program administrators. We find a program cost of 5.54 cents per reduced kilowatt-hour (kWh), which includes both public and private expenditures. This cost-effectiveness ratio compares favourably with prior estimates of returns to demand-side management programs[36–38] commonly used for government policy analysis, in which private spending is typically unobserved. This figure, however, does not include benefits in the form of higher property values and tenant rents. Total public expenditures of US$4.2 million for the LABBC program through 2012, include: $3.5 million in direct costs for conducting the audits and approximately US$700,000 in administrative costs. Private expenditures include an estimated US$74 million in building efficiency investments by building owners and managers. In qualitative interviews with commercial building owners and managers, the most cited reasons for participating are: savings with utilities, lower operating and maintenance costs, recognition from tenants, access to technology providers and local support. Unfortunately, financial operating data for specific properties participating in Energy Star and LEED programs are not disclosed as part of the certification process. As project implementation costs are proprietary and kept confidential by individual owners and managers, we are not able to generate cost-effectiveness estimates for these programs in the current study. From an evaluator's perspective, this is important future work. The estimated mitigation cost of 5.54 cents per reduced kWh in commercial buildings is comparable to the 5 cents per kWh previously estimated for behavioural energy conservation research and development programs most commonly employed in the residential sector[18], keeping in mind, however, that capital upgrades are subject to much larger investment hurdles and criteria.

External validity

We compared the performance of our sample of 178,777 Los Angeles commercial buildings against other commercial buildings in the United States. We converted the annual building consumption in electric energy use intensity (EUI) in kBtu/sqft/year and compared this with an external sample of commercial buildings from the Building Performance Database (BPD) maintained by the Lawrence Berkeley National Laboratory. The BPD is currently the largest publicly available national data set on building energy use in the United States[14]. As of May 2016, the BPD contains annual electricity data for 128,876 commercial buildings in all 50 states. These data were aggregated from smaller data sets and collected by various organizations, cities, utilities, publicly funded energy efficiency programs and building portfolio owners. In Fig. 2, we show that the distribution of LA commercial buildings follows a similar electric EUI profile as compared to commercial buildings in the rest of the country. This suggests that, at similar investment levels, the expected reductions in site EUI values, which determine output emissions reductions, could follow a similar distributional profile in other cities and regions. Los Angeles remains a leading market for green building improvements. However, the actual distributional profile of efficiency investments in other cities and regions will depend on specific market conditions, investment levels and building use profiles.

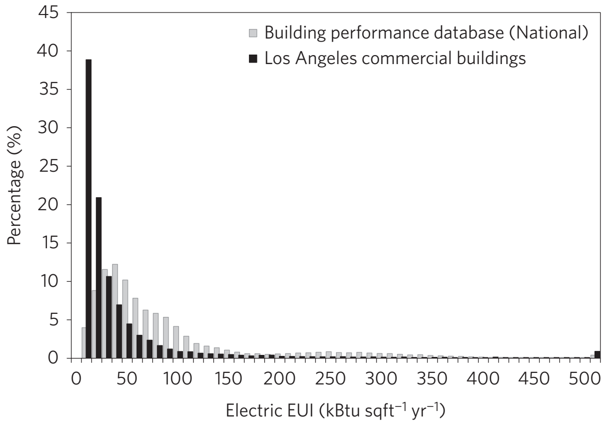

Figure 2 | Comparison of site Energy Use Intensity. EUI distribution of Los Angeles commercial buildings (black) versus a national data set from the Building Performance Database (BPD) assembled by the Lawrence Berkeley National Lab (grey).

Discussion and Policy recommendations

Commercial building owners and managers face steep investment hurdles. For the 91 initially enrolled buildings in the LABBC, total project costs for implementing the recommended energy conservation measures in 35 million sqft of floor space are an estimated US$82.81 million in 2012 dollars. The minimum investment levels per building range from US$136,000 up to about US$8.4 million for the largest buildings, net of available rebates and incentives. We observe a 30–40% project implementation rate in the LABBC program. This compliance rate following energy efficiency audits is consistent with previous studies[39]. Although the magnitude of these required investments may easily be justified for larger investors who own and operate larger Class A or Class B buildings, we note that even a 10% rental premium would be hard to justify financially in smaller, ageing infrastructure such as in Class C buildings. Investor strategies by asset class could partially explain the dominant participation among premium Class A buildings and the weaker participation among Class B or Class C buildings. However, weak participation at the lower end of the market is also structural. For instance, even for highly motivated Class C investors, only a fraction of buildings with net leases, for example, where tenants share in utility costs (as opposed to gross leases where tenants face zero marginal costs for utilities), have the ability to pass along investment costs to tenants. This suggests that a large share of the market becomes inaccessible to major private investment due to principal agent problems. Thus, the fact that participation and investments are primarily observed in larger commercial buildings (that is, 50,000 sqft and above) suggests that more effort might be required to attract smaller, capital-constrained investors.

Targeted information programs are needed to address both investment inefficiencies and energy use externalities[40–43]. Barriers to investments in energy efficiency still remain. For example, the evidence suggests that individual building owners and managers appear to be more sensitive to total implementation costs rather than to actual energy savings[39,44]. Research also shows that top management support[45] and the sequencing of recommendations can affect individual adoption decisions at a portfolio level[46]. When managers decide to invest, we show that structural upgrades are effective at reducing energy intensity in commercial buildings at an impressive performance level consistent with long-run emissions and energy reduction goals. These structural investments in building technologies are cost-effective versus demand-side management or new generation, but require major capital outlays, albeit at a lower level than investing in new capacity. For every public dollar invested in the community-based Los Angeles Better Buildings Challenge, this yielded an estimated return of US$17.6 in private infrastructure spending through 2012. Given the limits to public finance in funding capital upgrades in existing buildings and infrastructure, public–private partnerships aligned towards grand challenges may serve to extend the traditional boundaries of the public sector and increase directed innovation towards meeting societal goals.

Voluntary energy efficiency labelling programs are effectively targeting the most energy-intensive office buildings at the high end of the market. This is because existing programs and incentives currently result in positive selection—larger premium office space under professional management and owned by investors who seek rental and asset price premiums. From an emissions reduction point of view, the need for broader participation in energy efficiency is particularly relevant for building owners and managers in the least efficient three-quarters of buildings, particularly those buildings ineligible for energy efficiency labelling. These non-participating buildings tend to be smaller Class B and Class C buildings, but they are greater in number and in aggregate represent a significant two-thirds of greenhouse gas emissions inventories in the commercial building stock (Table 1).

In our participant interviews with major capital investors, we asked whether the future of investing in commercial energy efficiency would likely come from their portfolios of non-certified buildings—to which one investor replied: 'The current programs are not targeting poorer performing buildings.'

We argue that potential policy responses may be needed not only at city or regional level, but also at the state and federal level. For example, mandated information disclosure programs, which would require all commercial buildings to measure and disclose their energy use, might help to broaden participation and motivate poorer performers. First, they provide all performers with benchmarking information about relative consumption. Gathering building energy use data for the entire building population establishes a performance baseline that allows building owners to compare their buildings to similar buildings, but also to evaluate the magnitude of potential energy savings. Second, market pressure created by consumers and investors might create incentives for building owners to reduce their energy use when such information is shared throughout a city or industry. However, practical implementation may require significant investments to integrate information systems between utilities and jurisdictions for secure uploading and information management.

In summary, our study shows that increases in the availability of data can allow evaluators to become more accurate in societal accounting of energy and emissions reductions. Tracking these investments in the private sector presents challenges not just for evaluation efforts, but also for attributing its underlying causes. Without careful research design, when private investments in energy efficiency are made, we cannot be sure whether these investments are the result of strategic community policies, or whether they result merely from private considerations at the individual building or project level. The answer is that both of these considerations may be necessary to accelerate new investment. While energy savings are a primary outcome of building energy labels, we suggest further research into other outcomes, such as rental prices, vacancies and contracts. This will help clarify strategies that support long-run benefits, which could help broaden participation.

Methods

Identification strategy. We are interested in causal estimates of program participation for our three information programs (LEED, Energy Star and LABBC). This means we have three treatments and the counterfactual is a world in which these programs do not exist and there are no additional efficiency investments related to program participation. We estimate energy savings for each program in two steps. First, we match on observable building characteristics to maximize covariate balance between participating and non-participating buildings, independent of the outcome Y. Next, we conduct post-matching regressions that include statistical controls for weather and seasonality on building energy use. Using notation for the Neyman–Rubin causal model, the quantity of interest is the average treatment effect on the treated (ATT), commonly expressed as:

$$\text{ATT} = E\{E(Y_i X_i|, T_i = 1) - E(Y_i|X_i, T_i = 0|T_i = 1)\} \quad (1)$$

We select on observables and maintain two standard identifying assumptions. First, for each building i, the observed building covariates, X_i, are conditionally independent of the treatment (for example, conditional independence). Second, buildings with the same covariates have a positive probability of being participants or non-participants, $0 < \Pr(T = 1|X_i) < 1$ in the common support (for example, common overlap). The unit of analysis is at the building level.

Data. The data set consists of the universe of all commercial buildings in the Los Angeles Department of Water and Power (LADWP) service territory from 2005–2012. This includes 16,536,241 observations of 178,777 buildings in the City of Los Angeles and 56 neighbouring cities. We also acquired detailed building stock characteristics from CoStar, the premier commercial real estate database. CoStar provides building level information including physical building and occupancy characteristics and various measures of building quality. The 12 matching characteristics are listed in Table 2. To mitigate hidden bias, we include the CoStar rating (scored from 1–5), which captures several potentially unobserved factors quantitatively. The CoStar rating is a national rating system for commercial buildings that captures several characteristics commonly unobserved by evaluators, such as architectural attributes, structural and building system specifications, amenities, site and landscaping treatments, detailed property type specifics and other market factors. We also obtained local weather station data from the National Oceanic and Atmospheric Administration (NOAA) to calculate heating and cooling degree-days for each building to its nearest weather station by zip code.

Matching procedures. We evaluate several matching procedures and specifications. We evaluate matching procedures based on propensity scores and based on the genetic search algorithm. Supplementary Figs 1 and 2 shows the estimated propensity scores, for example, the conditional probability of treatment, using 1:1 nearest-neighbour matching (NNM), which is one of the most frequently employed matching procedures in the literature. The left-hand side in Supplementary Fig. 1 shows the density of estimated propensity scores for all non-participating commercial buildings and the right-hand side shows the density of estimated propensity scores for participating commercial buildings before matching. In Supplementary Fig. 2, we show the results after matching and sufficient sampling density in the region of common support, which is favourable in this application due to the high ratio of non-participating to participating buildings. We use observations on the common support and impose a calliper of 0.25 standard deviation to ensure that only the closest building matches are used to estimate treatment effects. We allow for replacement in the matching procedures, meaning that matched reference buildings can be re-used in the pool of available matches.

We also implement the genetic matching algorithm using procedures described in refs 34,47. Genetic matching uses a search algorithm that attempts to find the optimal matches by automatically determining the weight each covariate is given. Genetic matching solves the problem of 'researcher discretion' in model selection and avoids the manual process of iteratively checking covariate balance in post-matched samples and then re-estimating propensity scores. By using an automated process to search the available reference data, the search algorithm does not require the evaluator to make stronger functional form assumptions about the data generating process. We evaluate the performance of genetic matching against other commonly used matching strategies based on propensity scores, particularly in reducing median bias and other balance measures. In the energy efficiency application, we find that genetic matching performs at least as well as 1:1 nearest-neighbour matching with replacement in various measures of matching success. These include paired t-tests for balancing variables, p values from Kolmogorov–Smirnov (K–S) distributional tests, and empirical quantile–quantile (QQ) plots. For further comparisons, we also evaluate other matching procedures such as calliper, kernel and radius-matching to further analyse trade-offs between bias reduction and efficiency.

Assessing match quality. One common indicator to assess match quality for a given covariate is the standardized percent bias. The standardized percent bias (SB) is defined as[48]:

$$\text{SB} = \frac{100(\overline{X}_{\text{Treat}} - \overline{X}_{\text{Control}})}{\sqrt{S^2_{\text{Treat}} + S^2_{\text{Control}}/2}} \quad (2)$$

where $\overline{X}_{\text{Control}}$ is the mean of the control group and $\overline{X}_{\text{Treat}}$ is the mean of the treatment group; S^2_{Control} is the variance of the control group and S^2_{Treat} is the variance of the treatment group. In Fig. 1, we report the reductions in standardized bias across our covariates after matching. Other commonly used measures for evaluating match quality may include paired t-tests, K–S distributional tests and empirical quantile–quantile (QQ) plots. We show that matching procedures effectively mitigate systematic differences in covariate distributions and achieve a high degree of balancing across a broad set of observable building characteristics after matching. The performance of different matching algorithms varies case by case and depends largely on the data sample and distribution[24].

Supplementary Table 6 lists the detailed results of estimated program level energy savings using several matching algorithms including: NNM, radius, and kernel matching methods along with genetic matching where the objective function is either minimizing difference in p-values or minimizing distance in quantile–quantile (QQ) plots between treatment and matched controls. We report Abadie–Imbens standard errors[49]. See ref. 50 for a detailed review of the relative merits of these matching strategies in the empirical literature, particularly in evaluating trade-offs between bias reduction and efficiency of estimation. In Supplementary Table 6, the energy savings (average treatment effects on the treated) for participating buildings are in the range of 19–25%. Both NNM ($k = 1$) and genetic matching lead to optimal balancing characteristics (zero median bias) in estimates of program level energy savings without time-based weather controls, although standard errors are larger with the genetic matching algorithm. From the standpoint of bias reduction, genetic matching performs at least as well as 1:1 NNM with replacement, as both approaches converge to a central estimate of −19%. In Supplementary Table 7, we report the final matching results for each of the three programs by genetic matching. The estimates range from −19 to −30%. The median bias across covariates after matching is 0% for LABBC, 16.7% for LEED certified buildings and 5.6% for Energy Star certified buildings.

We eliminate or substantially reduce standardized bias via matching procedures based on propensity scores and/or the genetic search algorithm. We do observe a higher median bias after matching for the LEED buildings, due to sampling and size distribution of buildings in Los Angeles. For instance, LEED buildings are characteristically large Class A buildings with an average building area of 457,918 sqft and, notably, 15% of LEED buildings in the sample are 1 million sqft or larger. Consequently, there are fewer large non-certified Class A office buildings available in the reference set. Thus, 12% of Class C observations and 22% of Class B observations have been matched sub-optimally with Class A LEED buildings. This explains the higher conditional bias for LEED buildings after matching (Supplementary Table 5). Given our sample, we are able to reduce observable bias for LABBC and Energy Star buildings, close to or equal 100% (Supplementary Tables 3 and 4).

We may be additionally concerned about time-varying or seasonal effects on building energy consumption, particularly due to outside weather variation. For the final estimates of energy savings listed in Table 3, we first use NNM with replacement to extract a vector of covariate weights in the effective sample, and additionally include heating and cooling degree-days to report weather-adjusted treatment effects in post-matching regressions. The weather vector for heating degree-days (HDD) and cooling degree-days (CDD) in a given month is

$$\text{HDD}_{d,i} = \max\left\{0, \sum_{d=1}^{N} 65 - \theta_{\text{outside}}\right\}_i$$
$$\text{CDD}_{d,i} = \max\left\{0, \sum_{d=1}^{N} \theta_{\text{outside}} - 65\right\}_i \quad (3)$$

By US convention, the indoor base temperature is 65 °F (ref. 8). All buildings were matched to its nearest weather station by zip code. For a discussion of the importance of weather controls in evaluating energy information programs, see refs 8,51.

Robustness checks. We conducted several important robustness checks. We carefully considered the selection of covariates and matching specifications to reduce or eliminate observable bias. We explored alternative specifications such as the inclusions of higher order terms and interactions. We also evaluated the sensitivity of our estimates to hidden bias using Rosenbaum's bounds[52]. In Supplementary Table 8, we report results of our sensitivity analyses for hidden bias. In calculating Rosenbaum's bounds, we estimate thresholds for changes in statistical inference for different values of the sensitivity parameter Γ. This allows us to estimate changes in p values or significance levels in the presence of a potentially unobserved confounder. The critical gamma values are 1.21 for the LABBC program, 2.13 for the Energy Star program and 1.53 for the LEED program. For the Energy Star program, for example, the critical gamma value of 2.13 means that an unobserved covariate would have to affect the outcome kWh consumption by 2.1 times (roughly double the energy intensity in kWh/sqft) before we would change our inference at the 90% confidence level. To change our inference, an unobserved confounder for the LEED program should need to change the energy intensity of participating buildings by approximately 53% and for LABBC the figure is approximately 21%. Although we acknowledge that all studies are sensitive to sufficiently large biases, for our outcome of interest in kWh consumption per sqft, we believe these values may be sufficiently robust to small hidden bias, particularly for unobservable factors that may be uncorrelated to the covariates. Given our sample size, we are able to reduce conditional bias for LABBC and Energy Star buildings, close to or equal 100% (Supplementary Tables 3 and 4).

One possible source of hidden bias could be due to the influence of 'forward-looking' management. Previous research has suggested that 'forward-looking' management could play an indirect role in returns to green buildings through the hiring of more effective building managers[10], although this phenomenon has not been test empirically. We looked at the average consumption of professionally managed buildings versus non-professionally managed buildings, such as whenever a building had on-site property manager. We tested whether the presence of on-site management could be a proxy for more 'forward-looking' management under the hypothesis that on-site managers might be able to run the buildings more efficiently[53]. In our sample, we find no significant difference in the mean energy consumption of participating buildings with or without on-site management (0.64 kWh/sqft for participating buildings with on-site management versus 0.65 kWh/sqft for participating buildings without on-site management; t-test p-value 0.28). We therefore do not include the on-site manager dummy in the set of matching covariates. We confirmed this result by including the on-site manager dummy in the set of matching covariates and found no appreciable change in inference for our program estimates. Management quality, whether forward looking or not, is inherently unobservable with our data. As we include the 1–5 CoStar rating, which we expect to be weakly correlated with management quality, we acknowledge that we have only an indirect mitigation of this possible unobservable characteristic. Understanding the influence of management quality, particularly in an empirical setting with some sort of exogenous variation, would be great to explore as future research.

As further robustness tests, we also conduct placebo tests to validate our matching specifications. This is especially important given that we have no experimental benchmark to compare our causal estimates of program participation. Placebo tests are commonly underutilized in observational studies and are the conceptual equivalent of administering a sugar pill in a clinical trial[26]. In a placebo test, one attempts to find a stratum of data, and an outcome to test for the presence of treatment effects where none is logically possible. We implemented placebo tests in two ways. First, we obtained a list of 'prospect' buildings from LABBC program managers. These are buildings targeted by program managers for possible energy efficiency retrofits based on observable characteristics, but who have not yet been contacted nor are participating in the program. If our estimates are robust, then we expect to find zero treatment effect for 'prospect' buildings after matching on the same set of 12 building covariates and over the same period. In total, we matched 13,344 observations for 82 prospect buildings in our sample. The estimated treatment effect for prospect buildings over the same period is −2% versus matched controls, but not significant, meaning not statistically different from zero. We repeated placebo tests for prospect buildings versus reference groups of matched buildings that did not participate in Energy Star or LEED programs, and were also able to recover treatment effects not statistically different from zero after matching. We also conducted placebo tests 'in time' by evaluating the performance of participating buildings prior to retrofitting. Using this approach, we also found nonsignificant effects after matching on identical sets of covariates prior to program participation. These placebo tests give further credence to the robustness of our matching procedures and causal estimates. See Supplementary Note 3 and Supplementary Tables 9–11 for additional information on placebo tests in observational data.

Code availability. Computer code files and anonymized log files are available on Figshare (doi: http://dx.doi.org/10.6084/m9.figshare.4625119, ref. 54).

Data availability. Aggregated data that support the plots and findings in this paper are publicly available at http://energyatlas.ucla.edu. In California, access to individual electricity account information is protected as private information and may not be posted publicly. Financial data regarding participation and investment in the Los Angeles Better Buildings Challenge program are available upon request with the approval of the LABBC. Access to the external sample of commercial buildings from the Buildings Performance Database (BPD) is publicly available at: https://bpd.lbl.gov. Local weather station data used for calculating heating and cooling degree-days is available from the National Oceanic and Atmospheric Administration (NOAA) Quality Controlled Local Climatological Data (QCLCD) at https://www.ncdc.noaa.gov. Access to individual building and occupant characteristics is restricted under a commercial licence from CoStar.

Received 26 August 2016; accepted 17 February 2017; published 27 March 2017; corrected 11 April 2017

References

1. IPCC *Climate Change 2014: Mitigation of Climate Change* (eds Edenhofer, O. *et al.*) (Cambridge Univ. Press, 2014).
2. *Buildings and Climate Change: Summary for Decision-Makers* DTI/1240/PA (UNEP, 2009).
3. National Research Council *Real Prospects for Energy Efficiency in the United States* (National Academies, 2010).
4. *Energy Star Facts and Stats 2015* (EPA, 2015); www.energystar.gov/buildings/about-us/facts-and-stats
5. *Green Building Facts* (USGBC, 2015); http://stg.usgbc.org/articles/green-building-facts
6. *Better Buildings Challenge Overview* (DOE, 2016); http://betterbuildingssolutioncenter.energy.gov/sites/default/files/attachments/Better%20Buildings%20Challenge%20Overview.pdf
7. *Better Buildings 2012 Progress Report* (DOE, 2012); https://www4.eere.energy.gov/challenge/sites/default/files/uploaded-files/BB_ProgressRpt_508_121112.pdf
8. Delmas, M. A., Fischlein, M. & Asensio, O. I. Information strategies and energy conservation behavior: a meta-analysis of experimental studies from 1975–2012. *Energy Policy* **61,** 729–739 (2013).
9. Potoski, M. & Prakash, A. Green clubs: collective action and voluntary environmental programs. *Annu. Rev. Polit. Sci.* **16,** 399–419 (2013).
10. Eichholtz, P., Kok, N. & Quigley, J. M. The economics of green building. *Rev. Econ. Stat.* **95,** 50–63 (2013).
11. Eichholtz, P., Kok, N. & Quigley, J. M. Doing well by doing good? Green office buildings. *Am. Econ. Rev.* **100,** 2492–2509 (2010).
12. Fuerst, F. & McAllister, P. Green noise or green value? Measuring the effects of environmental certification on office values. *Real Estate Econ.* **39,** 45–69 (2011).
13. Kahn, M. E. & Kok, N. The capitalization of green labels in the California housing market. *Reg. Sci. Urban Econ.* **47,** 25–34 (2014).
14. Mathew, P. A. *et al.* Big-data for building energy performance: lessons from assembling a very large national database of building energy use. *Appl. Energy* **140,** 85–93 (2015).
15. Pincetl, S., Graham, R., Murphy, S. & Sivaraman, D. Analysis of high-resolution utility data for understanding energy use in urban systems: the case of Los Angeles, California. *J. Ind. Ecol.* **20,** 166–178 (2016).
16. Heckman, J. J., Ichimura, H., Smith, J. & Todd, P. Sources of selection bias in evaluating social programs: an interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method. *Proc. Natl Acad. Sci. USA* **93,** 13416–13420 (1996).
17. Allcott, H. & Greenstone, M. Is there an energy efficiency gap? *J. Econ. Perspect.* **26,** 3–28 (2012).
18. Allcott, H. & Mullainathan, S. Behavior and energy policy. *Science* **327,** 1204–1205 (2010).
19. Asensio, O. I. & Delmas, M. A. Nonprice incentives and energy conservation. *Proc. Natl Acad. Sci. USA* **112,** E510–E515 (2015).
20. Boomhower, J. & Davis, L. A credible approach for measuring inframarginal participation in energy efficiency programs. *J. Public Econ.* **113,** 67–79 (2014).
21. Ito, K. Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing. *Am. Econ. Rev.* **104,** 537–563 (2014).
22. Imai, K., King, G. & Stuart, E. A. Misunderstandings between experimentalists and observationalists about causal inference. *J. R. Stat. Soc. A* **171,** 481–502 (2008).
23. Keane, M. P. A structural perspective on the experimentalist school. *J. Econ. Perspect.* **24,** 47–58 (2010).
24. Ho, D. E., Imai, K., King, G. & Stuart, E. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Polit. Anal.* **15,** 199–236 (2007).
25. Imbens, G. W. & Wooldridge, J. M. Recent developments in the econometrics of program evaluation. *J. Econ. Lit.* **47,** 5–86 (2009).
26. Sekhon, J. S. Opiates for the matches: matching methods for causal inference. *Annu. Rev. Polit. Sci.* **12,** 487–508 (2009).
27. Pincetl, S. *Los Angeles Energy Atlas* (California Center for Sustainable Communities, 2016); www.energyatlas.ucla.edu
28. *Los Angeles Recognized for Leadership in Energy Efficiency, Better Buildings Challenge* (DOE, 2014); http://energy.gov/articles/los-angeles-recognized-leadership-energy-efficiency-better-buildings-challenge
29. Crawley, D. B., Hand, J. W., Kummert, M. & Griffith, B. T. Contrasting the capabilities of building energy performance simulation programs. *Build. Environ.* **43,** 661–673 (2008).
30. Deng, Y., Li, Z. & Quigley, J. M. Economic returns to energy-efficient investments in the housing market: evidence from Singapore. *Reg. Sci. Urban Econ.* **42,** 506–515 (2012).
31. Andam, K. S., Ferraro, P. J., Pfaff, A. & Sanchez-Azofeifa, G. A. Measuring the effectiveness of protected area networks in reducing deforestation. *Proc. Natl Acad. Sci. USA* **105,** 16089–16094 (2008).

32. Noah, M. A. *et al.* Referral to an extracorporeal membrane oxygenation center and mortality among patients with severe 2009 influenza A(H1N1). *J. Am. Med. Assoc.* **306,** 1659–1668 (2011).
33. Brady, H. E. & McNulty, J. E. Turning out to vote: the costs of finding and getting to the polling place. *Am. Polit. Sci. Rev.* **105,** 115–134 (2011).
34. Diamond, A. & Sekhon, J. S. Genetic matching for estimating causal effects: a general multivariate matching method for achieving balance in observational studies. *Rev. Econ. Stat.* **95,** 932–945 (2013).
35. Delmas, M. A. & Pekovic, S. Resource efficiency strategies and market conditions. *Long Range Plan.* **48,** 80–94 (2015).
36. Arimura, T. H., Li, S., Newell, R. G. & Palmer, K. Cost-effectiveness of electricity energy efficiency programs. *Energy J.* **33,** 63–99 (2012).
37. Loughran, D. S. & Kulick, J. Demand-side management and energy efficiency in the United States. *Energy J.* **25,** 19–43 (2004).
38. Auffhammer, M., Blumstein, C. & Fowlie, M. Demand-side management and energy efficiency revisited. *Energy J.* **29,** 91–104 (2008).
39. Anderson, S. T. & Newell, R. G. Information programs for technology adoption: the case of energy-efficiency audits. *Resour. Energy Econ.* **26,** 27–50 (2004).
40. Jaffe, A. B. & Stavins, R. N. The energy-efficiency gap what does it mean? *Energy Policy* **22,** 804–810 (1994).
41. Gillingham, K., Newell, R. G. & Palmer, K. Energy efficiency economics and policy. *Annu. Rev. Resour. Econ.* **1,** 597–620 (2009).
42. Dietz, T. Narrowing the US energy efficiency gap. *Proc. Natl Acad. Sci. USA* **107,** 16007–16008 (2010).
43. Matisoff, D. C., Noonan, D. S. & Flowers, M. E. Policy monitor—green buildings: economics and policies. *Rev. Environ. Econ. Policy* **10,** 329–346 (2016).
44. DeCanio, S. J. Barriers within firms to energy-efficient investments. *Energy Policy* **21,** 906–914 (1993).
45. Blass, V., Corbett, C. J., Delmas, M. A. & Muthulingam, S. Top management involvement and the adoption of energy efficiency practices: evidence from small and medium-sized manufacturing firms in the US. *Energy* **65,** 560–571 (2014).
46. Muthulingam, S., Corbett, C. J., Benartzi, S. & Oppenheim, B. Energy efficiency in small and medium-sized manufacturing firms: order effects and the adoption of process improvement recommendations. *Manuf. Serv. Oper. Manage.* **15,** 596–615 (2013).
47. Sekhon, J. S. Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *J. Stat. Softw.* **42,** 1–52 (2011).
48. Rosenbaum, P. R. & Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70,** 41–55 (1983).
49. Abadie, A. & Imbens, G. W. Bias-corrected matching estimators for average treatment effects. *J. Bus. Econ. Stat.* **29,** 1–11 (2011).
50. Caliendo, M. & Kopeinig, S. Some practical guidance for the implementation of propensity score matching. *J. Econ. Surv.* **22,** 31–72 (2008).
51. Asensio, O. I. & Delmas, M. A. The dynamics of behavior change: evidence from energy conservation. *J. Econ. Behav. Organ.* **126,** 196–212 (2016).
52. Rosenbaum, P. R. *Observational Studies* (Springer, 2002).
53. Kahn, M. E., Kok, N. & Quigley, J. M. Carbon emissions from the commercial buildings sector: the role of climate, quality and incentives. *J. Public Econ.* **113,** 1–12 (2014).
54. Asensio, O. Matching code for buildings (2017); http://dx.doi.org/10.6084/m9.figshare.4625119

Acknowledgements

This study would not be possible without S. Pincetl and her Energy Atlas team at the California Center for Sustainable Communities. For collaboration and access to program data, we thank D. Hodgins and J. Gould at the Los Angeles Better Buildings Challenge and B. Stapleton at the Los Angeles Cleantech Incubator. We also thank T. Vir Singh for high performance computing support on the Hoffman2 cluster. For helpful comments and suggestions, we thank H. Godwin, P. Kareiva, B. Lawrence, J. Lim, S. Neff, S. Muthulingam, S. Pincetl, D. Rajagopal, J. Sekhon and M. Tikoff Vargas. This research was supported by funding from the UCLA Rosalinde and Arthur Gilbert Program in Real Estate, Finance and Urban Economics; the Easton Technology Leadership Program at the UCLA Anderson School of Management; the University of California Center for Energy and Environmental Economics at Berkeley; the UCLA Institute for Digital Research and Education (IDRE) Postdoctoral Fellowship, and the Pritzker Distinguished Chair in Environment and Sustainability.

Author contributions

O.I.A. and M.A.D. contributed equally to this study.

Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.A.D.

How to cite this article: Asensio, O. I. & Delmas, M. A. The effectiveness of US energy efficiency building labels. *Nat. Energy* **2**, 17033 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Competing interests

The authors declare no competing financial interests.

Corrigendum: The effectiveness of US energy efficiency building labels

Omar Isaac Asensio and Magali A. Delmas

Nature Energy **2,** 17033 (2017); published 27 March 2017; corrected 11 April 2017.

In the version of this Article originally published, the value in the first sentence was misattributed to the United States and should have referred to global emissions. The sentence has been updated to 'Energy supplied in buildings accounts for an estimated 8.8 gigatons of CO_2 emissions globally or about one-third of total energy use and carbon emissions[1].'