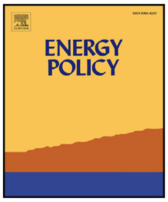
Energy Policy 178 (2023) 113581

Contents lists available at ScienceDirect

Energy Policy

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



Energy Performance Certificate renewal — An analysis of reliability of simple non-domestic buildings ' EPC ratings and pragmatic improving strategies in the UK

Mingda Yuan[a,*], Ruchi Choudhary[a,b]

[a] *Energy Efficient Cities initiative, University of Cambridge, Engineering Dept, Trumpington St, Cambridge, CB2 1PZ, UK*
[b] *Data-centric Engineering, Alan Turing Institute, 2QR, John Dodson House, 96 Euston Rd, London, NW1 2DB, UK*

ARTICLE INFO

Keywords:
EPC
Energy efficiency policy
Built environment
Clustering
Data mining

ABSTRACT

Energy Performance Certificates (EPCs) are designed to provide consistent and comparable energy efficiency ratings. Since 2008, any building constructed, rented or sold is required to obtain an EPC in the UK. 4% of simple office buildings have renewed their EPC to date in England and Wales. In this paper, by comparing the original and renewed EPCs of these buildings, we examine the reliability of non-domestic EPC ratings and discover realistic energy efficiency improvement strategies reflected in the changes. We have found that buildings with their conditions remaining unchanged can be better rated by approximately 10 points on average when their initial rating is above the minimum requirement in the regulation. The ones initially rated below the minimum requirement are expected to have a greater scale of improvement on ratings while making no efforts which brings doubts on the effect of the "Minimum Energy Efficiency Standard" policy. Evidenced human factors are also jeopardising the fairness of EPC ratings. Based on the changes in the EPCs, we have found 4 clusters of improved buildings and 5 clusters of deteriorated buildings. We derived practical energy efficiency improvement strategies that people have already naturally adopted.

1. Introduction

European Union member countries adopted the European Performance Buildings' Directive (EPBD) in 2003 (EPBD, 2010). As a result of the directive, Energy Performance Certificates (EPCs) have become the main instrument across Europe for rating the energy efficiency of existing buildings and providing relevant recommendations for improvements. In the UK, any building built, rented, or sold since 2008 is required to obtain an EPC valid for 10 years[1] (BEIS, 2017). Asset Ratings, the primary marker of English and Welsh non-domestic EPCs, are designed to show the energy efficiency rating across all types of non-domestic buildings.

In England and Wales, non-domestic EPCs are calculated using the Simplified Building Energy Model (SBEM) that embeds the UK National Calculation Methodology (NCM) (Communities and Government, 2020). Different from domestic EPCs generated from the Standard Assessment Procedure (SAP) based only on the inspected building, the non-domestic EPC system uses a "reference building" which is set to be the same for all buildings of a given type to ensure comparability. The reference building is a simplified equivalent virtual building (i.e. a building of the same size, shape and use as the actual building inspected) constructed in the Asset Rating calculation process. Reference buildings are all designed to be naturally ventilated, gas-fueled, servicing at a fixed space strategy with the insulation level and all other main parameters of the buildings identical to the 2006 Building Regulations Part L.[2]

The final Asset Rating is calculated by comparing the CO_2 emissions calculated for the actual building, termed Building Emission Rate (BER), and the emission calculated based on the reference building, termed Standard Emission Rate (SER), both in $kgCO_2/m^2$/year. Unlike domestic EPC data, the published non-domestic data contains very

---

* Corresponding author.
E-mail addresses: my356@cam.ac.uk (M. Yuan), rc488@cam.ac.uk (R. Choudhary).

[1] For non-domestic buildings, an EPC is required if the building's floor area is greater than 50 m^2.

[2] The Building Regulations set out requirements for specific aspects of building design and construction. Part L sets out the conservation of fuel and power, controls the insulation values of buildings elements, the allowable area of windows, doors and other openings, the air permeability of the structure, the heating efficiency of boilers, hot water storage and lighting. It also controls mechanical ventilation and air conditioning systems, space heating controls, airtightness testing of larger buildings, solar emission, and the certification, testing and commissioning of heating and ventilation systems, and requirements for energy meters. (Regulations, 2016).

https://doi.org/10.1016/j.enpol.2023.113581
Received 2 November 2022; Received in revised form 24 March 2023; Accepted 11 April 2023

Table 1
Comparison of information available in the published domestic and non-domestic EPC data based on a selection of variables.

Variable	Domestic EPC	Non-domestic EPC
Total Floor Area (m^2)	✓	✓
Main Heating Control Type	✓	X
Multi-Glaze Proportion (%)	✓	X
Glazed Type	✓	X
Glazed Area (m^2)	✓	X
Hot water Type	✓	X
Floor Type	✓	X
Windows Type	✓	X
Walls Type	✓	X
Roof Type	✓	X
Low Energy Lighting (%)	✓	X
Main Heating Fuel Type	✓	✓
Lighting Type	✓	X
Floor Height (m^2)	✓	X
Construction Age Band	✓	X
Fixed Lighting Outlets Count	✓	X
Low Energy Fixed Lighting Outlets Count	✓	X
Aircondition Present	X	✓[a]
Aircondition Present KW Rating	X	✓[a]
Primary Energy (kWh/m^2 per year)	X	✓[a]

[a] These variables contains higher than 60% missing value in the data, not included in the analysis in our research.

limited information on the building's physical conditions (e.g. no information on wall types, window types etc., as illustrated in Table 1). In addition to the calculated Asset Rating, the retrofit recommendations part is an important section of the EPC report. Although the building's physical conditions, like U-values of envelopes and roof types, are not directly available, they can be inferred from the recommendation report.

EPC is also an important instrument for other built environment energy efficiency policies. To motivate the improvement of buildings with the worst energy efficiency, the UK government has set a "Minimum Energy Efficiency Standards (MEES)" for non-domestic buildings in regulations approved by the parliament on 26 March 2015 (BEIS, 2017). The Regulation states:

(a) From 1 April 2018, landlords of non-domestic private rented properties (including public sector landlords) may not grant a tenancy to new or existing tenants if their property has an EPC rating of band F or G (shown on a valid EPC for the property).
(b) From 1 April 2023, landlords must not continue letting a non-domestic property which is already let, if that property has an EPC rating of band F or G.

Hence, the success of MEES depends on the reliability of EPCs.

The British government compiles the information related to the creation of EPCs in a central register that is publicly available. After 13 years in practice, it is now a good time to look back at the reliability of non-domestic EPCs and learn from the information they provide. Furthermore, a significant share of buildings has renewed their EPCs owing to market activities or after the expiration of their initial one. Analysis of the renewed EPCs against the original EPCs could potentially reveal the patterns of changes in the building stock.

This paper examines the reliability of non-domestic EPC ratings by comparing updated EPCs against the original generated for the same buildings and discovers patterns arising in the EPC renewal. The analysis is carried out in two parts:

- The analysis of EPC Asset Ratings based on the rating calculation process. (Section 4)
- The analysis changes in EPC Recommendations that potentially expose sources of Asset Rating changes and show the improvements commonly made against the recommendations made in the original EPCs, which would reveal the types of practical, preferred, and economically feasible renovations. (Section 5)

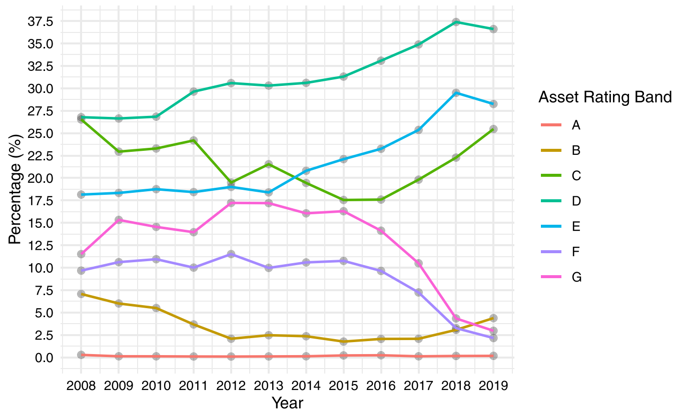
The detailed introduction of non-domestic EPC data and the analysis structure is presented in Section 3. Although the EPC Asset Ratings analysis and the EPC recommendations analysis are done through different methods and use different subsets of the EPC data, the results of the two analyses are expected to support each other. The results obtained by analysing the variables that are directly involved in the Asset Rating calculation would help better explain the patterns found in recommendation changes, while irregular patterns discovered in recommendation changes can confirm alarming findings revealed in the Asset Rating analysis. To our best knowledge, this is the first paper that studies the reliability of non-domestic EPCs ratings generated from a "reference building" calculation method. This is also the first paper to investigate patterns arising from non-domestic EPC recommendations.

2. Analyses of EPCs in research

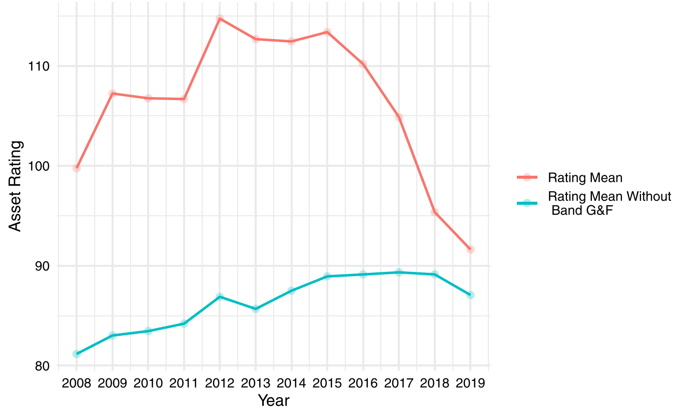
Across the academic literature, EPCs have been studied from the following perspectives: (a) correctness and reliability, (b) the value of the data associated with the inputs used for generating the Asset Rating, and (c) their role as policy instruments. It should be noted that, although all the EPC rating schemes are designed under the guidance of the EU Energy Performance in Building Directive (EPBD) (EPBD, 2010), the methodologies adopted in each country are not identical (Arcipowska et al., 2014; Semple and Jenkins, 2020). Hence, the studies based on different countries' EPC data are good references, but the conclusions may not be universal.

With respect to correctness, a number of studies indicate that variations arising from the assessment process alone can be significant. von Platten et al. (2019) studied the renewed EPCs of Swedish multifamily buildings and found that new EPCs tend to overestimate the energy performance improvement since the last assessment. Hårsman et al. (2016) found that the differences across assessors can result in 20% variation in the calculation of Swedish domestic buildings. Tronchin and Fabbri (2012) reveal that only 70% of ratings were in the correct range when 162 technicians independently evaluated the EPC rating of the same building in Italy. Hardy and Glew (2019) compared domestic EPCs generated for the same properties and estimated that between 36 and 62% UK domestic EPCs exhibit at least one major inconsistency. Inconsistencies found in domestic EPCs typically stem from disagreements across assessors on physical elements like floor or wall type, the energy efficiency of products like wall insulation or glazing performance, and lodgement errors (Burman et al., 2014). Moreover, Crawley et al. (2019) have found that one standard deviation measurement error decreases with domestic EPC rating in England and Wales, and the predicted error is higher than the UK guidance except for the most efficient buildings. An EPC "Mystery Shopper" study, where multiple domestic EPC assessors assessed the same properties (Jenkins et al., 2017), found that the EPC generating process contributed to variation in asset rating by an average of 11 points per dwelling on average, with one property recording a variation of over 30 points. Consistent with other studies, they found that differences in assessors' interpretation of critical parameters of buildings, such as the age band, wall type, etc., are the main reason for these large ranges. One study in Ireland Ahern et al. (2016) found that the use of default values for the thermal transmittance coefficients of the building envelope decreased their credibility.

Pasichnyi et al. (2019) have summarised 79 papers analysing EPC data across the EU and the UK. These studies use not just the Asset Rating but also the inputs associated with calculating it and the recommendations chosen by the auditors for each building. The main applications of EPC data are the calculation of building energy performance for energy retrofitting (Droutsa et al., 2016; Gouveia and Palma, 2019; Christensen et al., 2014) and the assessment of impacts of EPC on the real estate market (Hyland et al., 2013; Murphy, 2014; Fuerst and McAllister, 2011; Amecke, 2012; Fuerst et al., 2016). Gelegenis et al.



(a) Asset Rating yearly trend by band.



(b) Comparison of general Asset Rating trend and trend without buildings in band F and G.

Fig. 1. Asset Ratings of existing Level 3, B1 buildings from 2008 to 2019.

(2014) analysed the recommendations contained within Greek domestic EPCs. The authors have found that the constraint of a maximum of three recommendations per building is a limitation to identifying all required interventions that could improve the buildings' energy efficiency level. The most frequently recommended interventions are also of high cost, while other simpler and more cost-effective options are not recommended. They have also analysed the EPCs issued under the funding of "Saving at Home" program. Each building involved in the program is inspected twice — before and after the intervention. The comparison of the two EPCs for the same domestic buildings shows that the types of recommendations most suggested and adopted in Greece's domestic sector, like aluminium frames and solar water heating collectors, are more related to living styles and matters of security, comfort and house appearance rather than the energy efficiency objectives. Streicher et al. (2017) analysed 6000 domestic EPC in Switzerland and the retrofit recommendations included within them to estimate theoretical energy savings achievable for the Swiss residential building stock. Gupta and Gregg (2018) proposed retrofit packages for UK domestic buildings based on the carbon emissions of retrofit options reflected in the domestic EPCs.

Policy-wise, Davis et al. (2017) uses the EPC scores to model the effects of a tax-based switch to energy efficiency and to understand the tax incidence effects of such a policy. Yuan and Choudhary (2020) uses EPC information together with domestic gas consumption and social factors to provide a locally tailored design of residential heating policies.

As we are approaching the analysis of the reliability of the non-domestic EPC in the UK from all the published EPC data, we have greater coverage and completeness compared to most studies. For example, Amirkhani et al. (2021) also looked into the uncertainties in non-domestic buildings in the UK. However, they focused on three hotel buildings only, and the generalisability of the result may be limited. Our study also extended Hardy and Glew (2019)'s discussion of domestic EPCs to non-domestic EPCs and further discussed policy implications from the analysis.

3. Current asset ratings of level 3, B1 buildings and analysis structure

In the non-domestic EPC system, a building's complexity is designated across three levels: levels 3, 4 and 5. This paper focuses on office buildings of complexity level 3, i.e. small existing office buildings with heating systems less than 100 kW and cooling systems less than 12 kW.[3] This means that the bias introduced in non-domestic EPCs due to the complexity of large HVAC systems is not a factor in the analysis. Non-domestic buildings are classified into different categories based on their use. Businesses that can be carried out in a residential area without detriment to its amenity are categorised as B1 class buildings.[4] In the published non-domestic EPC data to date, the Level 3 B1 buildings take up 22.48% of the whole non-domestic building stock.

The Asset Rating resulting from a non-domestic EPC inspection is formally defined as:

$$AssetRating = \frac{BuildingEmissionRate}{StandardEmissionRate} \times 50 \tag{1}$$

the final value is rounded to an integer. To categorise the energy efficiency of buildings, the Asset Rating is linearly converted into energy bands from A to G with an interval of 25 between each band border. Band A buildings, with an Asset Rating ranging from 0 to 25, are the most energy efficient[5] and band G buildings, with Asset Ratings greater than 150, is the least energy efficient. A building with an Asset Rating of 50, right at the border of bands B and C, is just compliant with the Building Regulations Part L.

From Fig. 1(a), we can see that the share of bands F and G across Level 3 B1 buildings started to shrink after the approval of the MEES in 2015 and further sharply dropped after the regulations came into effect in 2018. As only 0.59% of Level 3 B1 non-domestic EPCs are issued to the newly built buildings, the shrink of the F and G categories are then considered mainly contributed by existing buildings. The decline of F and G-rated EPCs indicates that the regulations have effectively motivated buildings in bands F and G to improve their energy efficiency. These improvements provide an opportunity to study the actual updating patterns of these buildings at scale. That is, what are the energy improvements property owners are voluntarily adopting in reality to improve the buildings? Fig. 1(b) shows that, if band F and G buildings are not included, the average Asset Rating of simple office buildings increases across band A–D buildings, implying that there is a need to encourage owners of buildings that are above the regulation limit to further improve their energy efficiency.

In the past 12 years, 4% of Level 3, B1 buildings (hereafter referred to as B1-3 buildings) have had their EPC updated. We further separate these buildings into three categories, based on the extent of difference between the updated and the original Asset Rating:

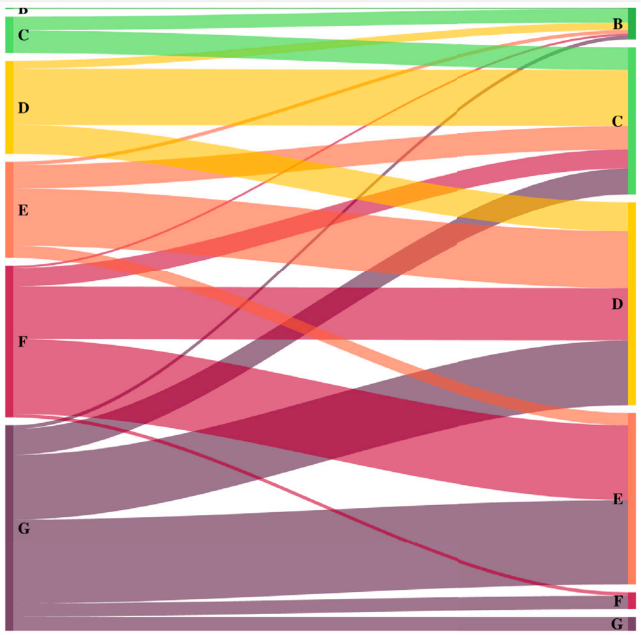
- Improved Buildings :
  (Updated Asset Rating - Original Asset Rating) < -5
- Deteriorated Buildings :
  (Updated Asset Rating - Original Asset Rating) > 5
- Unchanged Buildings :
  $-5 \leq$ (Updated Asset Rating - Original Asset Rating) ≥ 5

---

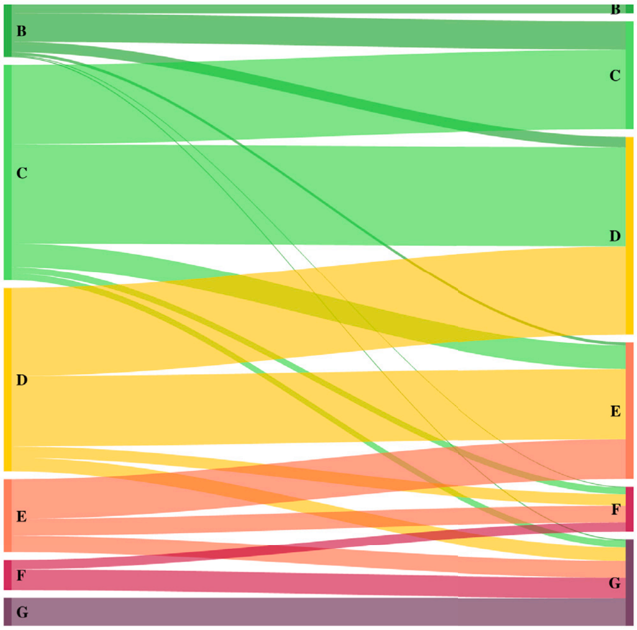
[3] From the published EPC data, they have a primary energy consumption of 336 kWh/m^2 per year on average and have a mean floor area of 267 m^2.

[4] From 1 September 2020, this classification is effectively replaced with the new Class E(g), but does not affect our study.

[5] A+ buildings with negative Asset Ratings are not considered in our study.



(a) Improved Buildings: 66.17% (3478 in count)



(b) Deteriorated Buildings: 19.38% (1242 in count)

Fig. 2. Asset Rating changes of renwed EPCs.

In this study, the concept of improvement and deterioration is based on the changed Asset Rating. The quality assurance procedure of EPC in England and Wales requires that a sample of the EPCs created by each assessor undergoes a desk-based audit and that 90% of the audited EPCs must be within ±5 points of the original value for non-domestic buildings. Hence, we set the boundaries of unchanged buildings to have Asset Rating changes between −5 and +5 to allow insignificant fluctuation. Fig. 2 shows that while 66% of updated buildings have improved their Asset Ratings, 19% of updated buildings have worse Asset Ratings on their new EPC. Many buildings with good initial Asset Ratings dropped their level in the new EPC. Hence, in addition to analysing the patterns of improvements in buildings, it is also essential to find out how and why some buildings have their energy efficiency level dropped.

Although we cannot directly acquire knowledge of the building's physical conditions, like U-values of envelopes and roof types, we can infer the conditions by the recommendations in the recommendation report. The recommendations that are triggered by the conditions of the building are listed in Table 2. The assessor who carries out the inspection can add or delete recommendations according to the cost-effectiveness and suitability of a recommendation with respect to a given building. One can assume that the EPC recommendations comprehensively represent building conditions due to the triggering mechanism and that changes in recommendations reflect changes in the actual building conditions.

There are three endogenous assumptions according to the design of the non-domestic EPC-generating process:

- EPC Asset Ratings are consistent and comparable across buildings and buildings' updates.
- EPC Recommendations reflect the building conditions due to the recommendation triggering mechanism, and changes in recommendations in the EPC update process reflect changes in building condition.
- The assessors do not bias the Asset Rating and recommendations in a significant manner.

We have identified and verified that 5400 out of 130,213 B1-3 buildings had updated their EPC[6] once[7] between 01/04/2010 and 30/09/2020. In this work, we will test the validity of these assumptions along with the analysis of patterns of changes in these buildings. Signs of the violation of assumptions would suggest the EPC generating process is not entirely operating as the expectation by the design of the system.

The reliability of Asset Ratings across buildings is hard to test as buildings differ from one another in many aspects. Comparing the EPCs before and after the renewal of the same building is a viable alternative in order to examine the reliability of their Asset Ratings. We do so in the following two manners:

(1) Examining the Asset Rating calculation. As introduced in Section 1, the Asset Rating is defined as proportional to the ratio of the Building Emission Rate (BER) of the inspected building and the Standard Emission Rate (SER) of a reference building. For the same building, the change of Asset Rating upon EPC renewal can result from changes in the BER, the SER or both. Asset Rating reliability can be assessed by analysing the relationship between Asset Rating, SER, and BER. The fact that the SER is based on 2006 Building Regulations Part L compliance standards and fixed heating and cooling conditions helps maintain Asset Ratings' comparability across all buildings. SER should theoretically remain the same for the same building's EPCs, with Asset Rating changes resulting from BER alterations. We will investigate if SER changes affect updated Asset Ratings.

(2) Considering changes in building conditions. Non-domestic EPC data contains very limited information about actual building conditions. We use EPC recommendations to help further investigate calculated Asset Ratings. If the recommendations of the same building are unchanged, we assume the building conditions triggering the recommendations have also remained the same. By comparing the Asset Ratings of the same unchanged building along with its recommendations, we can test if there are any systematic changes in Asset Ratings in the renewal process.

---

[6] The EPC data is accessed from https://epc.opendatacommunities.org/.

[7] There are also 148 B1-3 buildings have updated their EPC more than once. We do not include them in the study to ensure comparability.

Table 2
List of all non-domestic EPC recommendations.

Code	Trigger	Description	Category	Payback
EPC-C1	Using default cooling efficiency	The default chiller efficiency is chosen. It is recommended that the chiller system be investigated to gain an understanding of its efficiency and possible improvements.	COOLING	3
EPC-C2	Poor/fair cold generator efficiency	Chiller efficiency is low. Consider upgrading chiller plant.	COOLING	3.5
EPC-C3	Poor/fair duct leakage	Ductwork leakage is high. Inspect and seal ductwork	COOLING	7.5
EPC-W1	Hot water is not provided by the space heating heat generator and poor/fair heat generator efficiency	Install more efficient water heater	HOT-WATER	4.15
EPC-W3	Storage heat loss > (default value × 0.9)	Improve insulation on DHW storage	HOT-WATER	3.8
EPC-W4	There is secondary HWS circulation and there is no time control	Add time control to DHW secondary circulation	HOT-WATER	4.5
EPC-W2	HWS efficiency is poor	Consider replacing DHW system with point of use system	HOT-WATER	8
EPC-E1	Any floors have U-value > 1.0	Some floors are poorly insulated - introduce/improve insulation. Add insulation to the exposed surfaces of floors adjacent to underground, unheated spaces or exterior.	ENVELOPE	15
EPC-E2	Any flat roofs have U-value > 1.0	Roof is poorly insulated. Install/improve insulation of roof.	ENVELOPE	25
EPC-E3	Any solid walls have U-value > 1.0	Some solid walls are poorly insulated - introduce/improve internal wall insulation.	ENVELOPE	6.5
EPC-E4	Any cavity walls have U-value > 1.0	Some walls have uninsulated cavities - introduce cavity wall insulation.	ENVELOPE	3.7
EPC-E5	Any glazing have U-value > 3.5	Some windows have high U-values - consider installing secondary glazing	ENVELOPE	4.6
EPC-E6	Pitched roofs with lofts have U-value > 1.0	Some loft spaces are poorly insulated - install/improve insulation.	ENVELOPE	5.6
EPC-E7	permeability > 14	Carry out a pressure test, identify and treat identified air leakage. Enter result in EPC calculation	ENVELOPE	7
EPC-E8	Any glazing have U-value > 3.5	Some glazing is poorly insulated. Replace/improve glazing and/or frames.	ENVELOPE	9.3
EPC-F1	Oil or LPG as fuel	Consider switching from oil or LPG to natural gas	FUEL-SWITCHING	1.08
EPC-F2	Coal as fuel	Consider converting the existing boiler from coal to natural gas	FUEL-SWITCHING	3.75
EPC-F3	Coal as fuel	Consider switching from oil or LPG to biomass	FUEL-SWITCHING	3.81
EPC-F4	Oil or LPG as fuel	Oil or LPG to biomass (heating)	FUEL-SWITCHING	6.7
EPC-F5	Gas as fuel	Consider switching from gas to biomass	FUEL-SWITCHING	6.72
EPC-F6	Coal as fuel	Consider switching from coal to oil	FUEL-SWITCHING	8.4
EPC-H2	Heating system does not have centralised time control	Add time control to heating system	HEATING	1.8
EPC-H5	Heating system does not have room by room time control	Add local time control to heating system	HEATING	5.8
EPC-H6	Heating system does not have room by room temperature control	Add local temperature control to the heating system	HEATING	4.8
EPC-H7	Heating system does not have optimum start and stop control	Add optimum start/stop to the heating system	HEATING	2.5
EPC-H8	Heating system have weather compensation controls	Add weather compensation controls to heating system	HEATING	5
EPC-H1	Poor heat generator efficiency	Consider replacing heating boiler plant with high efficiency type	HEATING	2.3
EPC-H3	Poor or fair heat generator efficiency and fuel is gas, oil or LPG	Consider replacing heating boiler plant with a condensing type	HEATING	6.6
EPC-H4	Using default heating efficiency	The default heat generator efficiency is chosen. It is recommended that the heat generator system be investigated to gain an understanding of its efficiency and possible improvements.	HEATING	3
EPC-L1	Have T12 lamps	Replace 38 mm diameter (T12) fluorescent tubes on failure with 26 mm (T8) tubes	LIGHTING	0.6
EPC-L2	Have GLS lamps	Replace tungsten GLS lamps with CFLs: Payback period dependent on hours of use	LIGHTING	0.85
EPC-L3	Have high-pressure mercury discharge lamps	Replace high-pressure mercury discharge lamps with plug-in SON replacements	LIGHTING	1.8
EPC-L4	Have GLS lamps	Replace tungsten GLS spotlights with low-voltage tungsten halogen: Payback period dependent on hours of use	LIGHTING	2.5
EPC-L5	Have T8 lamps	Consider replacing T8 lamps with retrofit T5 conversion kit.	LIGHTING	2.8
EPC-L6	Have high-pressure mercury discharge lamps	Replace high-pressure mercury discharge lamps with complete new lamp/gear SON (DL)	LIGHTING	3.5
EPC-L7	Fluorescent lamps have mains frequency ballasts	Introduce HF (high frequency) ballasts for fluorescent tubes: Reduced number of fittings required	LIGHTING	5.7
EPC-V1	Solar gain limit defined in the NCM is exceeded in any zone in the building	In some spaces, the solar gain limit defined in the NCM is exceeded, which might cause overheating. Consider solar control measures such as the application of reflective coating or shading devices to windows.	OVERHEATING	1.7

(*continued on next page*)

Table 2 (*continued*).

Code	Trigger	Description	Category	Payback
EPC-R1	Heating fuel is electricity, and heat generator efficiency <2	Consider installing a ground source heat pump	RENEWABLES	11.7
EPC-R2	Wind turbine not installed	Consider installing building mounted wind turbine(s)	RENEWABLES	15.9
EPC-R3	Solar thermal water heating not installed	Consider installing solar water heating	RENEWABLES	20.2
EPC-R4	Photovoltaic system not installed	Consider installing PV	RENEWABLES	44.7
EPC-R5	Heating fuel is electricity, and heat generator efficiency <2	Consider installing an air source heat pump	RENEWABLES	9.8

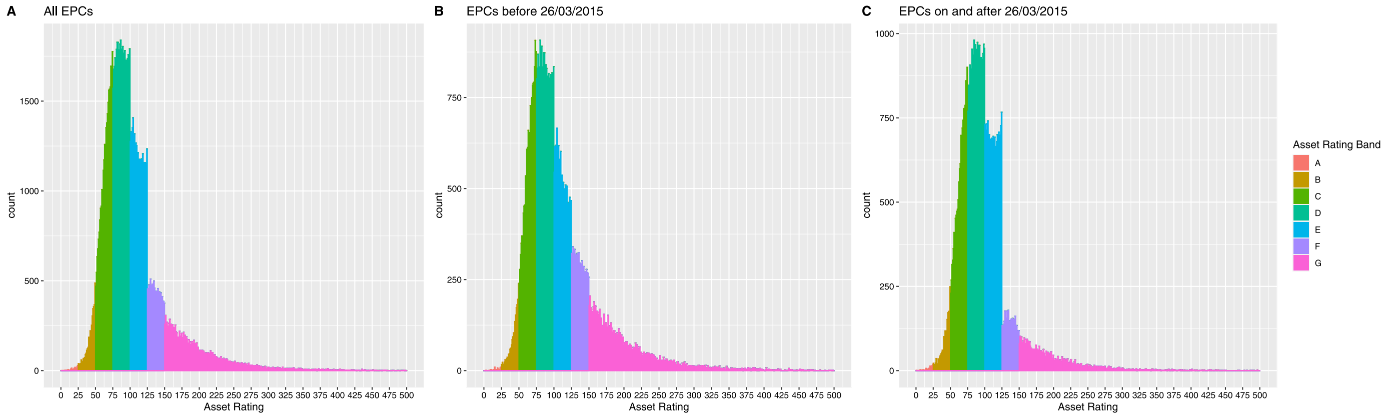


Fig. 3. Asset Rating distributions.

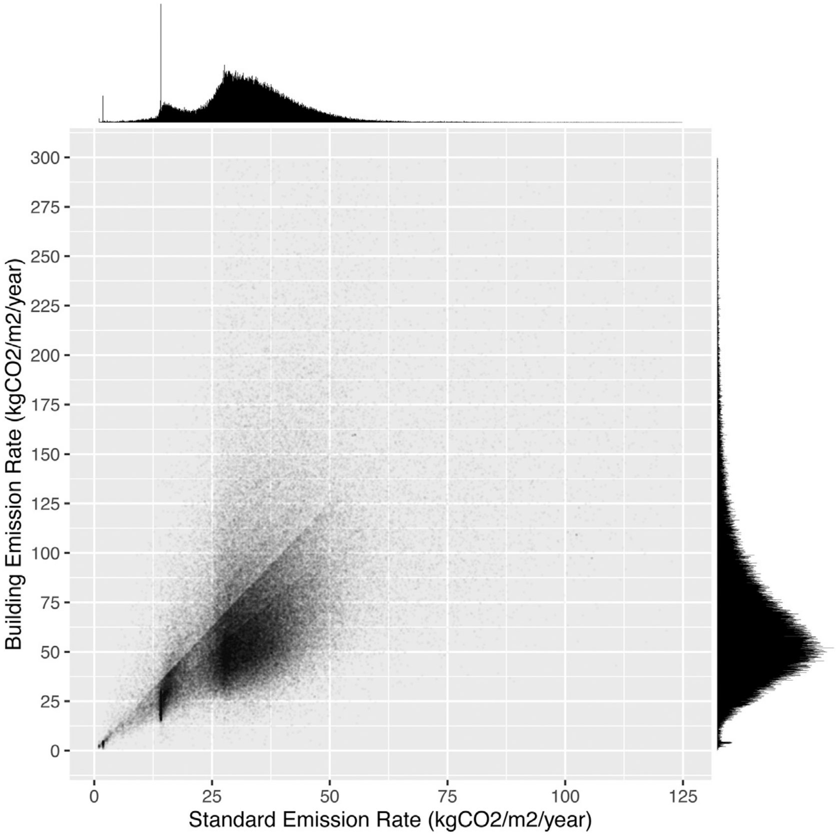


Fig. 4. SER and BER distributions.

4. Reliability of non-domestic EPC ratings

4.1. Distribution properties

We first explore the distribution properties of Asset Rating, SER and BER of all B1-3 buildings used in this study. All three distributions are right-skewed with a long right tail as shown in Figs. 3 and 4. In the presence of long-tailed distributions, the task of outlier detection is extremely difficult. Without the data like the U-value of the wall used in the EPC generation, we cannot distinguish the extreme values in the natural distribution from outliers. Considering the sensitivity of parametric statistics (e.g.mean) towards outliers, we omit the extremal 2.5% of the data from both tails, retaining the middle 95% of the data in all the parametric statistics to avoid the influence of possible outliers.

In Fig. 3A, the Asset Rating distribution shows a relatively smooth curve from the 0–75 (band A–C) interval but exhibits sharp drops and separations at the boundary of each band thereafter. An evident drive for the sharp decrease within bands F and G is the renewal ban of buildings in these two bands from 26/03/2015 by the aforementioned regulation. In addition to the significant reduction of frequency in bands F and G in Fig. 3C, the effect of the policy also leads to an upward curve in the band B and E intervals (100–125). However, Fig. 3B indicates that a non-smooth transition between bands after an AR of 75 existed prior to the policy's legislation. It should be noted that drops between building bands would not have occurred without external influence. When buildings are improved to achieve a better rating, the improvement would still result in a smooth Asset Rating distribution, as building owners would be unaware of how much work is required to achieve a specific rating. Hence, these sharp changes are considered not "naturally" generated and indicate a human element, such as assessors rating buildings close to a border higher so that they receive a better band rating.

The SER distribution and BER distribution are both multimodal (having multiple distinct peaks), as shown in Fig. 4. The SER distribution is trimodal with two spikes at 1.88 and 14.17, while the BER distribution has a big main body and a small spike around zero. Without the original data used to calculate SER, we cannot infer the underlying cause of the two distinct spikes, but it is reasonable to assume these two specific numbers are generated using the same set of default values in the calculation. Typically, the multimodality of data implies that the observed distribution can be categorised into multiple types. However, for the same reason, it is not possible to further decompose the SER distribution into different sub-distributions based solely on their generation process differences. The lack of original SER calculation data and the differences between SER and BER density distributions do not directly lead to the conclusion of typology. In such cases, anti-modes (the least frequent value between modes) found via kernel density estimates, shown for this case in Fig. 5, can be used as natural boundaries of different distribution parts.

4.2. Influence of BER and SER on Asset Rating

All three variables (Asset Rating, BER and SER) have complex distributions. The bands in the Asset Rating further complicate the analysis as we have shown non-smooth transitions between bands. Hence, we expect the relationship between these variables also to be in complex forms (e.g. trends may vary by interval selection). The Pearson correlation coefficient is most widely used to assess the global statistical association of two variables, but it is incapable of capturing local variations of relationships as needed in this case. Local Weighted Polynomial Regression, equivalent to a Savitzky–Golay filter, is a non-parametric regression method that effectively models local trends in the data. This is a standard statistical method applied when the two variables analysed have an unknown relationship and are not expected to have a deterministic relationship that can be described simply via a mathematical function such as linear regression. At each data point, a low-degree polynomial (fitted using weighted least squares) is modelled for a subset of data around the point. The points closer to the initial data point are assigned a higher weight, while those further away are weighted less. As the subset range moves along the data, the algorithm results in a smoothed fitting curve, referred to as LOWESS (locally weighted scatterplot smoothing) curve. The LOWESS curve can be seen as the conditional means of the response variable, and a confidence interval of the curve can also be calculated. As the resulting curve models complex local relationships, it cannot be expressed via any single mathematical formula, making it difficult to summarise. An alternative to model local relationships is Gaussian Process. Different from Local Weighted Polynomial Regression, Gaussian Process models the whole sample but not local subsets. In our case, with all variables having heavy tails, structural change of data in the tail part would bring difficulties in Gaussian Process modelling. We expect SER to remain at the same level at the point of renewal of the EPC for the same building. However, the data is only partially aligned with this expectation. Fig. 6 A, B, and C show the distributions of value changes in Asset Rating, SER and BER in the renewed EPC compared to their original values. Just over half (50.22%) of the updated buildings exhibit an unchanged SER, as indicated by the spike at zero in Fig. 6B. It is necessary to investigate how the SER changes in the other half of the building affect the Asset Ratings. Both the means of Asset Rating and BER decreased in the renewed EPC, as shown in Fig. 6 A and C, but with long tails and skewed distributions, indicating a need for detailed analysis.

The LOWESS curves of each variable and their changes in the EPC update are shown in Fig. 6 D, E and F.[8] For example, the LOWESS curve in Fig. 6 D plots the conditional mean and its confidence intervals of the net change in Asset Rating (y-axis) against the given original Asset Rating (x-axis). The global correlation between the original Asset Rating and the difference between the original and new Asset Rating (Δ Asset Rating) is −0.75 and would be more significant if not considering the sparse right tail. This means the decrease in Asset Rating is more significant when the original Asset Rating is higher, except for the extreme cases in tails when the original Asset Rating is higher than 225. Decomposing the Asset Rating trend into SER and BER trends yields further understanding: Fig. 6E shows the difference between the original and new SER against the original SER has a negative correlation except for sparse tail areas. The curve intercepts with zero ΔSER at 35.5 of original SER, close to the mean original SER of 33.02. Buildings with lower than average SER in the initial inspection tend to increase in the update, while buildings with higher initial SER tend to decrease. This is usually considered a "regression effect" where the remeasured data tend to move towards the mean without any intervention. In our case, clear groups of updated buildings have their new SER shifted between different sub-distributions from the original value as shown in Fig. 6E, but the negative correlation holds within each sub-distribution change group. As discussed above, changes in building conditions should not result in a change in the building SER.

4.3. Systematic decrease of Asset Rating in the EPC update

Similar to the SER, the LOWESS curves in Fig. 6 D and F indicate strong negative correlations between the change in the updated value and the original value of Asset Rating and BER, except for the right sparse tail. The higher the original Asset Rating/BER value, the bigger the decrease in the updated value, and vice versa for the increase in the lower original values. One needs to note that these correlations reflect the general pattern of the updated buildings but do not necessarily hold valid for every specific building. However, different from SER, we cannot conclude that this is due to the regression effect for the Asset Rating and BER. Instead, the changes in Asset Rating and BER in the

---

[8] The details of the fitting process of Local Weighted Polynomial Regression models are described in Appendix.

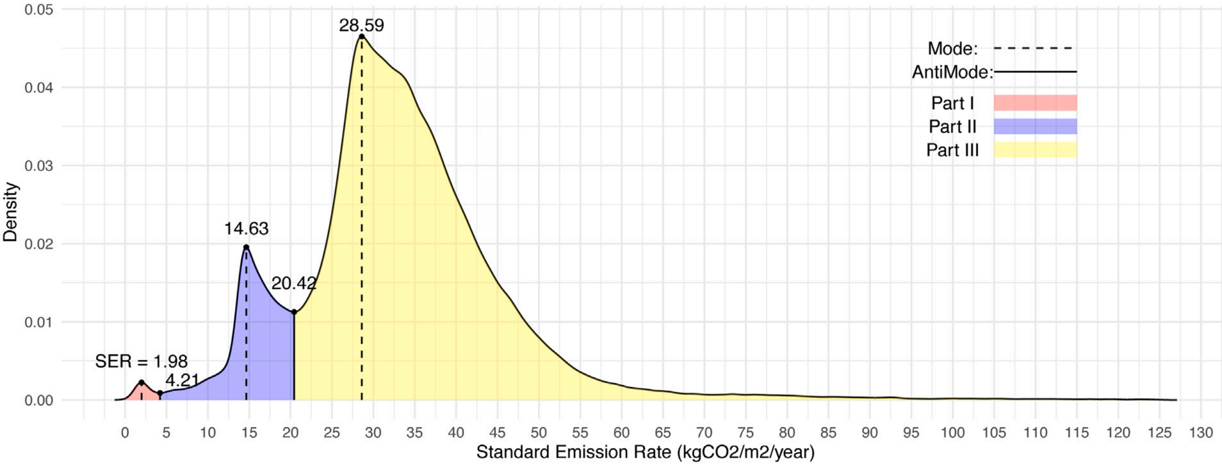


Fig. 5. SER density.

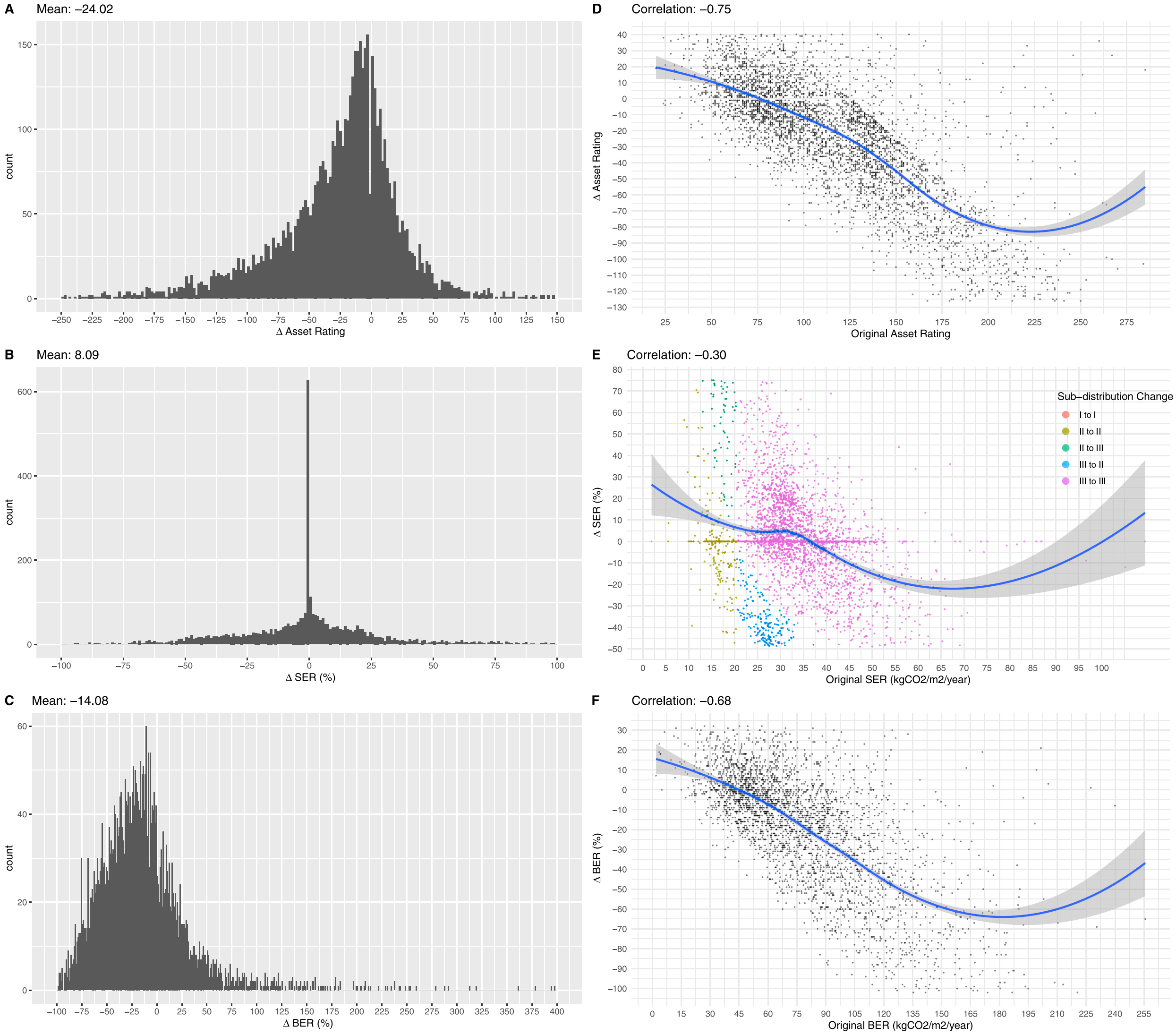


Fig. 6. Changes (Δ) in Asset Rating, SER and BER in the EPC updates.(A,B,C as distributions and D,E,F as LOWESS curve against the original value with 95% confidence interval).

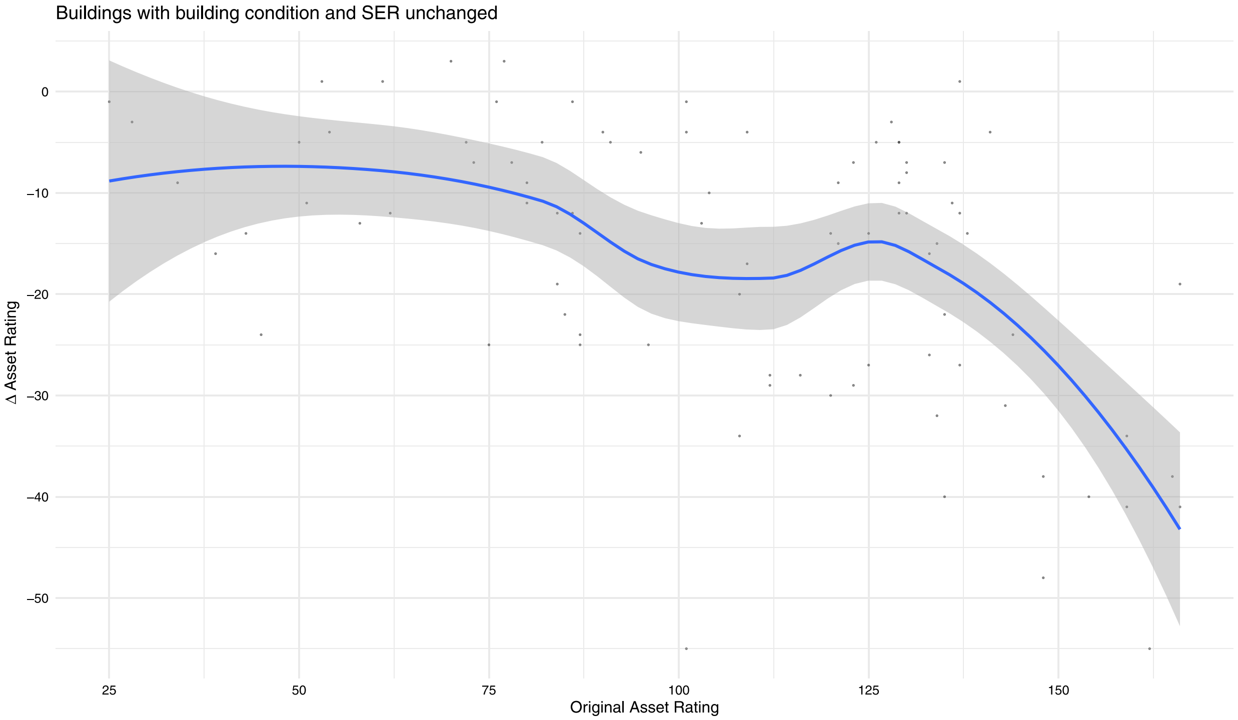


Fig. 7. LOWESS curve of changes (Δ) in Asset Rating's against the original Asset Rating with 95% confidence interval.

second inspection may be caused by either self-correction in the second inspection or by actual changes in the building. We will further discuss the Regression Effect of AR and BER of updated buildings with cluster analysis results later in Section 5.

We have found 116 buildings with identical information recorded in the updating process, including recommendations that indicate the comprehensive building conditions and SER remained unchanged. BERs are the only differences between the original and updated EPCs. As shown in Fig. 7, we have found that, with everything else being the same, Asset Rating (affected solely by changes in BER in this case) has significantly and systematically decreased in the EPC update on average. The changes to the EPC calculation framework over the last 13 years may partially explain this decrease. The scale of the decrease is enlarged when the original Asset Rating is greater than 125 (band F boundary). This indicates some buildings are rated above “The Minimum Level of Energy Efficiency” in the new EPC while staying unchanged. We cannot separate systematic improvements from improvements caused by actual changes in the buildings other than, but the existence of systematic improvement is alarming as it casts doubts on the reliability of all EPC ratings. These 116 buildings lodged their EPCs between 2010 and 2020 but had an average time gap between the two EPCs issued for 2 months. Hence, the motivation for updating these EPCs may be partially due to property owners not satisfied with their first rating, and this also brings the possibility that the second-time assessors may try to help to improve the EPC ratings.

4.4. SER and BER changes of improved and deteriorated buildings

For buildings whose Asset Ratings have improved, we can see from the LOWESS curve in Fig. 8A that the SER change has a weak negative correlation with the original Asset Rating. The 95% Confidence Interval of the conditional mean contains zeros except for the interval of 96–146 of the original Asset Rating. This means, on average, the SER tends to increase significantly for buildings with an original Asset Rating from 96 to 146. The improvement of Asset Ratings of buildings that were originally rated in the interval 96–146 is potentially amplified by the increase of SER. Fig. 9A exhibits that the improved buildings have a BER significantly decreased on average, especially the ones that have an original Asset Rating greater than 125.

Conversely, buildings whose Asset Rating deteriorated show a weak positive correlation between SER change and the original Asset Rating. The confidence interval of the conditional mean of SER change has its upper bound below zero between the interval of 20–72 in Fig. 8B. This means the change of SER is significant in the interval. As shown in Fig. 9B, while the conditional mean of the change in BER is significantly increased on average, the buildings with an original Asset Rating of less than 70 have a higher increase. Together with the findings on SER change above, the buildings with an original Asset Rating between 20–72 may have their new Asset Rating further increased in the EPC update, not because of any degradation in the conditions of the building but due to the potential decrease of SER. Buildings that maintained their original Asset Rating (SER change and BER change shown in Figs. 8C and 9C) exhibit no significant changes on average in the EPC update.

Analysing the improved and deteriorated buildings separately, we have shown the mean of SER changes is significantly greater or smaller than zero when the original Asset Rating is in certain intervals. These significant SER changes are undesired as SER is designed to stay unchanged for the same building to maintain comparability and consistency. Together with the findings from the 116 unchanged buildings and distributions of Asset Ratings, we have justified that the current non-domestic EPC system is not providing reliable Asset Ratings at all times. Possible human factors may jeopardise the fairness of the non-domestic EPC ratings.

5. Patterns in buildings’ updates

We have analysed the reliability of non-domestic EPC ratings solely based on the calculation of the ratings. As discussed in Section 3, the recommendations of EPCs are representative of the conditions of the assessed buildings due to the triggering mechanism. Hence, the changes in the recommendations would reflect the changes in building conditions. We now apply pattern recognition algorithms to discover patterns in the changes (added or dropped in the new EPC) of EPC

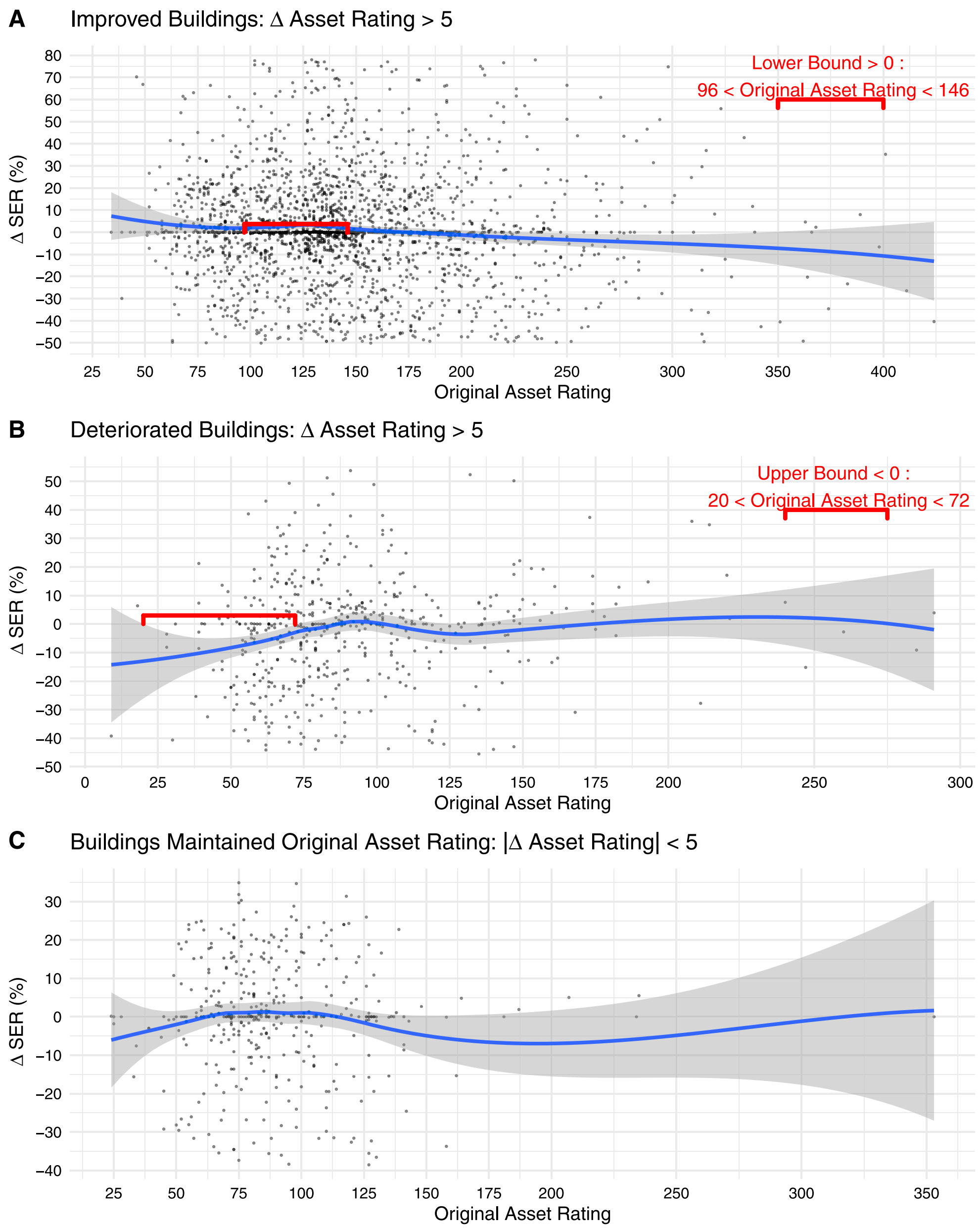


Fig. 8. LOWESS curve of changes (Δ) in SER against the original Asset Rating with 95% confidence interval.

recommendations issued for the same building. If the discovered patterns contradict with expected outcomes when the EPCs are issued in a reliable way, then along with our understanding of SER and BER changes, they would further validate our findings in Section 4. The patterns that do not show irregular properties would reveal the changes in the buildings naturally adopted by people and thus inspire future retrofits and energy efficiency policies.

5.1. Cluster analysis and association rule mining

Two levels of building condition changes' patterns emerge in the EPC renewal: the building-level pattern representing the integrated changes of the whole building and the action-level pattern describing the relationship among various possible actions that change the buildings' energy efficiency.

The discovery of building-level patterns takes all the variables that describe the specific changes in each building and then finds buildings that are changing similarly. Such a task is usually referred to as cluster analysis. When the data is of large volume and high dimensions, as they are in this case, the Gaussian mixture model (GMM)[9] is a good algorithm for cluster analysis. GMM is a model-based clustering algorithm that generates clusters based on the probabilistic distribution of the data. GMM assumes all variables included in the modelling come from a multivariate Gaussian density with a finite number of mixture components (clusters). GMM involves low level of subjective human judgment, such as selecting the optimum number of clusters, and produces robust results (Bishop and Nasrabadi, 2006). The optimum number of clusters is decided by information criteria during the fitting process rather than subjectively decided by the modeller. One disadvantage of GMM is that the interpretation of the results is less straightforward compared to simpler ones like hierarchical clustering due to the model complexity. But simpler models like hierarchical clustering cannot handle large volume and high dimension data as good as GMM. Furthermore, GMM requires the input data to be continuous. When the data is of high dimensional binary form, like discussed later

---

[9] The GMM model is realised via R package "mclust" (Scrucca et al., 2016).

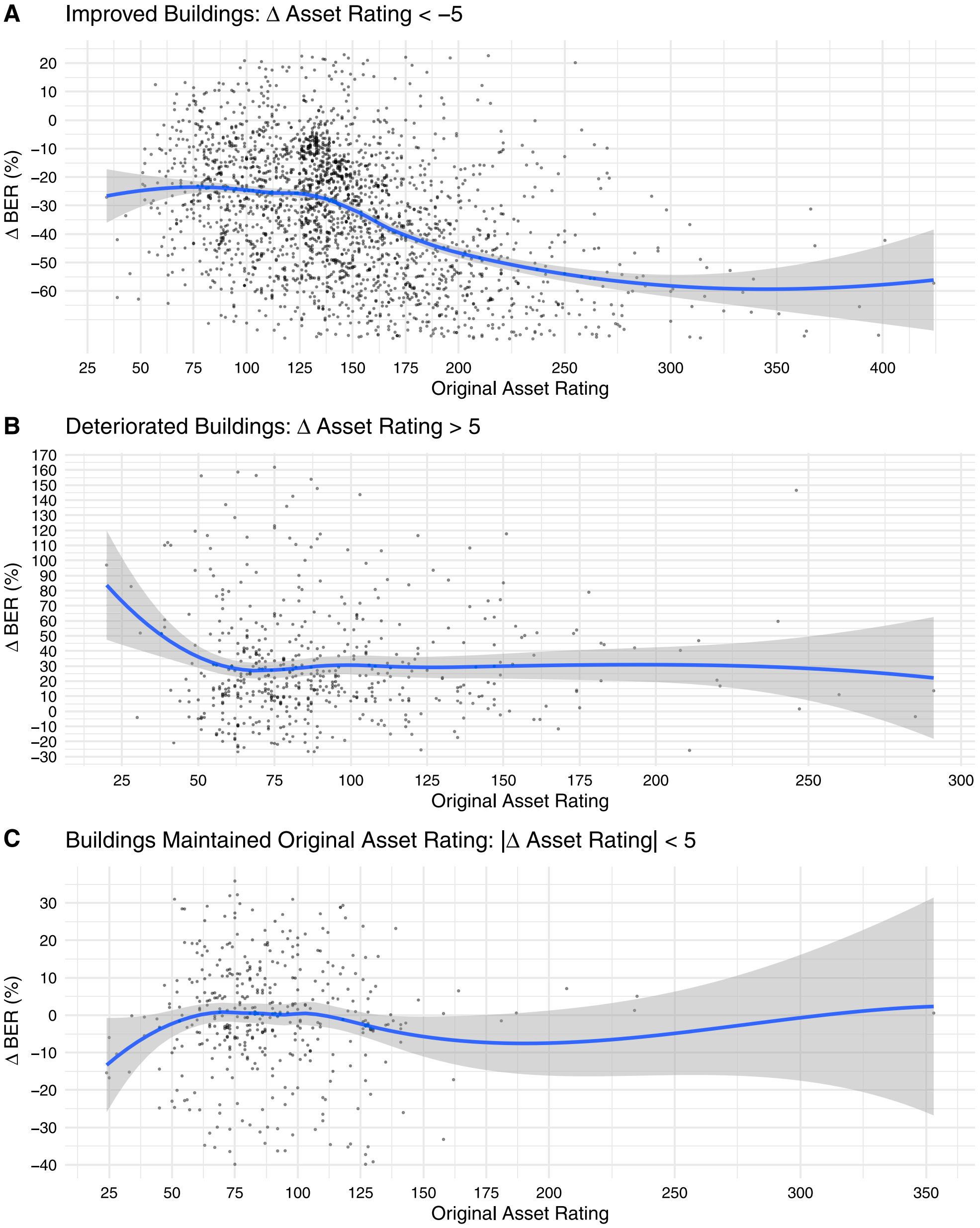


Fig. 9. LOWESS curve of changes (Δ) in BER against the original Asset Rating with 95% confidence interval.

in Section 5.2, logistic principal component analysis (LPCA) can help transform the binary data into continuous data and reduce dimensions to focus the most important information in the data and avoid noise by selecting a small number of principal components that preserve the majority of the variance in the original data.

The objective of the action-level analysis is to investigate if one variable representing a single change in the building condition is linked to any other variables. This is carried out via association rule mining. Analysing the action-level patterns of different clusters discovered at the building level helps us understand how people choose to change the building conditions. Association rule mining was initially introduced to discover product purchase patterns in supermarkets' large-scale transaction data. For example, a strong rule found in the supermarket transaction data such as $\{\text{Bread}\} \Rightarrow \{\text{Milk}\}$ implies that if a customer purchases bread, it is likely that they also buy milk. Similarly, we would like to discover association rules in the changes of buildings to see if people's choice to change one specific aspect of the building would regularly lead to some other changes. Our study limits the association rule mining to have only one item from both the left-hand side and the right-hand side. To select the rules of interest, we need to apply various constraints based on significance measures. The common choices of the measures are Support and Confidence, where Support shows how frequently the two items appear in the dataset together, while Confidence indicates how often the rules are found to be true.

5.2. *Pattern discovery results*

We assume the possible changes in building conditions during the EPC updates are comprehensively reflected in the changes of triggered recommendations. Hence, we have 84 indicator binary variables (either 0 or 1) to represent the changes in the 42 recommendations. Half of them indicate whether the recommendation has newly appeared in the new EPC, while the other half indicates whether the recommendation that existed in the old EPC is dropped in the update. As we also have the primary heating fuel information of buildings, we introduce an additional indicator variable (1-Heating Fuel changed; 0-Heating Fuel

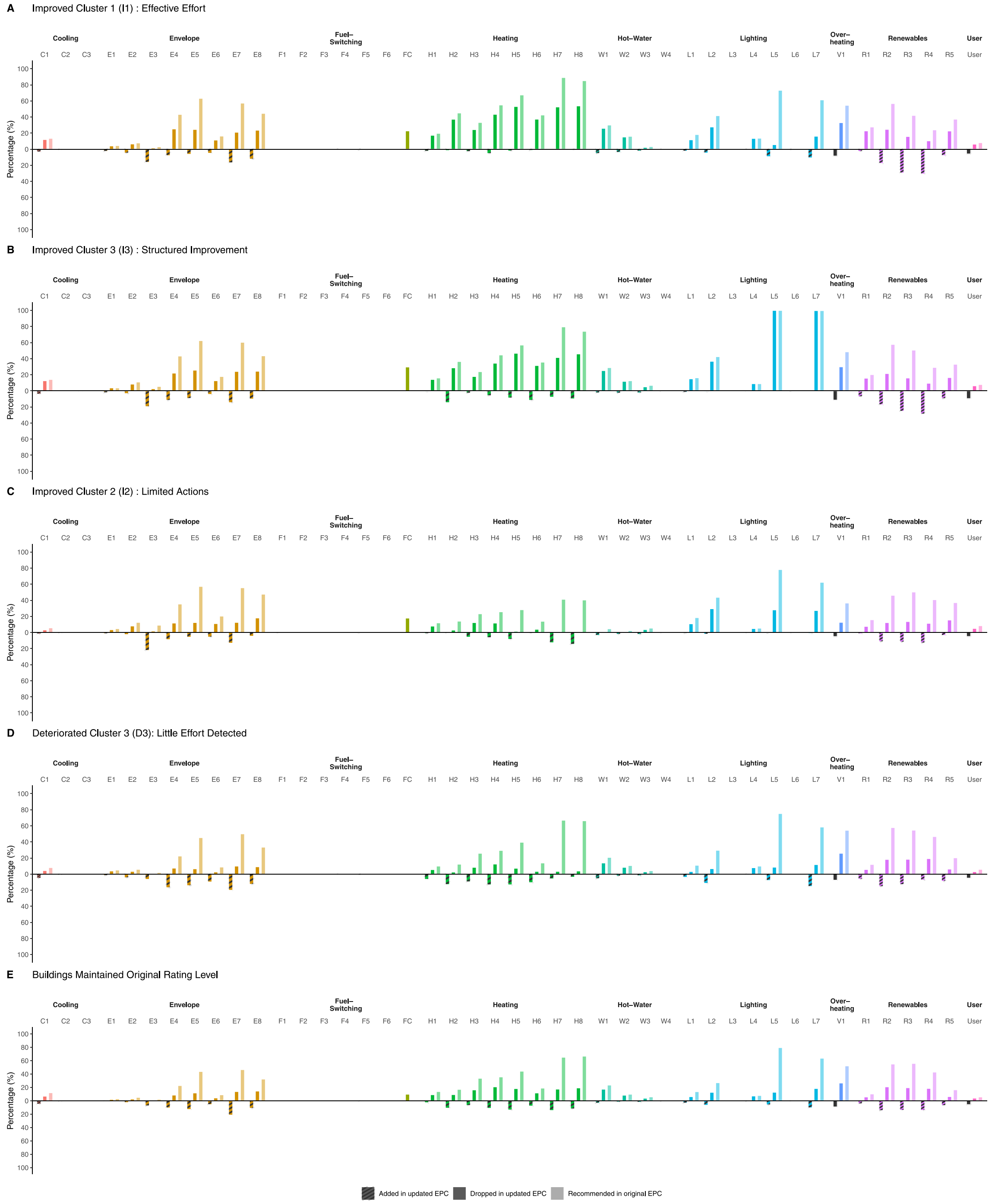


Fig. 10. Patterns of recommendation changes in clusters Part I.

unchanged) to describe if the building has different primary heating fuel in the new EPC.

Logistic principal component analysis of the total 85 indicator binary variables helps us preserve the most important variations within the data in fewer dimensions and eliminate noises. Based on the first 30 logistic PCAs, which preserve 80% of variation of the original data, we find 4 clusters of improved buildings (I1, I2, I3, I4) and 5 clusters of deteriorated buildings (D1, D2, D3, D4, D5) through GMM clustering. To compare the differences of each cluster, we have summarised the variable differences, namely recommendation change and heating fuel change of buildings, for each cluster in Figs. 10 and 11. The solid shaded bars above the x-axis represent the percentage of buildings in the cluster with each recommendation dropped in the EPC renewal, while the stripped shaded bars below the x-axis represent the percentage of buildings with a given recommendation added in the renewal. The percentage of the existence of recommendations in the original EPCs of buildings is also included in these figures as lightly shaded bars above the x-axis to better illustrate the changes. By comparing the adjacent bars above the x-axis, we can easily see the improvement of buildings in different clusters.

In addition to the comparison of variables included in the clustering process, we have also summarised the information in the EPC that is not included in the clustering process and does not directly reflect a building condition change or with missing values (e.g. SER before

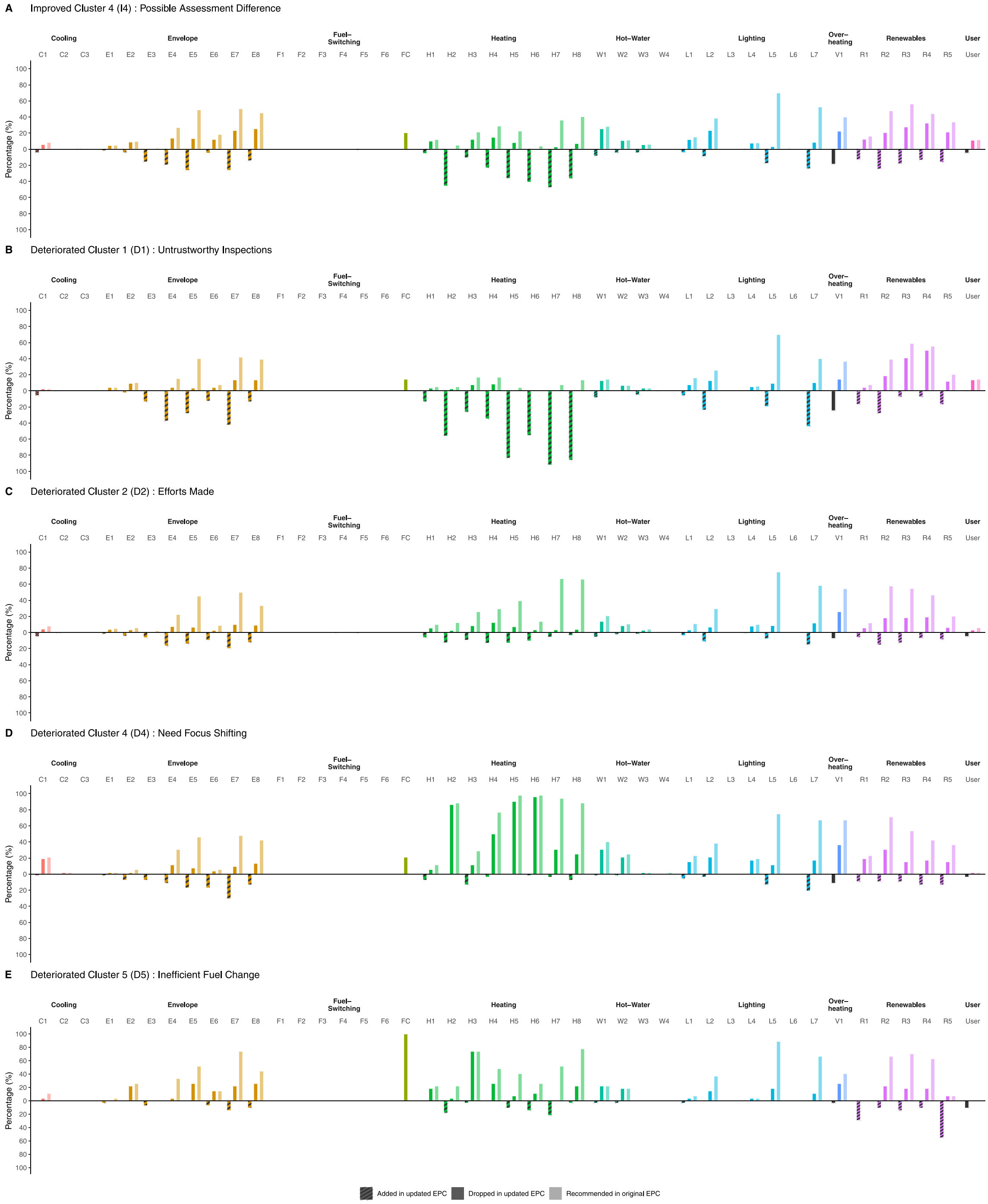


Fig. 11. Patterns of recommendation changes in clusters Part II.

April 2010). This summary is shown in Table 3 to assist the analysis of clusters' differences. It is worth noting again that all parametric statistics (e.g. mean) in our study are calculated based on the middle 95% of the data to avoid the possible influence of outliers.

While the clustering analysis aims to find the patterns at the building level, to further discover the relationship among the changes, we apply association rule analysis to find if some changes are associated with other changes in each cluster discovered. The Apriori algorithm used to discover association rules is set to find rules with the support of 5% and confidence of 80% to assure the quality of discovered rules.

The nine discovered clusters and the group of buildings that maintained original Asset Ratings together provide us with two main types of patterns: (a) retrofit strategy patterns and EPC system improvement possibilities, and (b) signs of the unreliability of the current non-domestic EPC system. These two pattern types are now explored in more detail.

5.2.1. Clusters show retrofit strategy patterns and EPC system improvement possibilities

Cluster I1 & I3: Importance of structured move Buildings in Cluster I3 and I1 have the greatest and the second greatest improvement in their Asset Ratings, respectively. They exhibit similar patterns in Fig. 10 B and A. Most original recommendations dropped in the renewal section, with few new additions. These buildings can still improve envelope

Table 3
Cluster differences based on EPC information.

Cluster		Share (%)	Count	Average Asset Rating change	Median Asset Rating change	Average Original Rating	Average Time Gap (Year)	Average Recommendation Added	Average Recommendation Dropped	Average Original Standard Emission Rate (kg $CO_2/m^2/year$)	Average Standard Emission Rate Change (%)	Average Original Building Emission Rate (kg $CO_2/m^2/year$)	Average Building Emission Rate Change (%)
Clusters of Improved buildings	All Improved Buildings	66.17	3573	−47.26	−38	139.31	3.84	2.57	4.94	31.64	1.22	90.40	−32.56
	Improved Cluster 1 (I1): Effective Effort	20.11	1085	−48.67	−39	139.44	4.44	2.27	6.94	31.44	−1.85	91.51	−35.72
	Improved Cluster 2 (I2): Limited Action	25.73	1389	−44.75	−36	138.21	2.71	1.65	2.93	31.51	1.91	88.09	−29.96
	Improved Cluster 3 (I3): Structured Improvement	8.09	437	−55.35	−48	141.40	4.77	2.48	8.06	31.64	0.89	92.78	−38.28
	Improved Cluster 4 (I4): Possible Assessment Difference	12.25	662	−45.34	−35	140.55	4.62	5.35	4.04	32.28	5.38	93.11	−29.59
Clusters of Deteriorated buildings	All Deteriorated Buildings	19.38	1046	26.22	20	82.49	3.76	3.28	3.44	34.01	−2.67	58.79	30.04
	Deteriorated Cluster 1 (D1): Untrustworthy Inspections	2.48	134	33.07	27	81.57	4.59	8.19	3.03	36.08	5.6	61.53	51.25
	Deteriorated Cluster 2 (D2): Efforts Made	2.67	144	23.79	20	78.38	5.18	2.73	7.48	36.85	−7.2	55.09	25.70
	Deteriorated Cluster 3 (D3): Little Effort Detected	12.54	677	24.54	19	83.62	3.52	2.84	2.57	33.82	1.47	58.60	27.99
	Deteriorated Cluster 4 (D4): Need Focus Shifting	1.11	60	38.14	25	83.04	3.81	2.67	7.4	39.66	−5.33	65.47	37.69
	Deteriorated Cluster 5 (D5): Inefficient Fuel Change	0.57	31	35.60	30	84.00	4.47	2.66	4.66	37.45	−5.72	66.23	34.18
Buildings Maintained Original Rating		14.43	421	−0.35	−1	86.88	3.83	2.67	3.77	36.13	1.59	64.11	1.07
All Updated Buildings		100	5400	−24.05	−17	118.30	3.74	2.69	4.29	32.29	0.81	79.59	−16.97

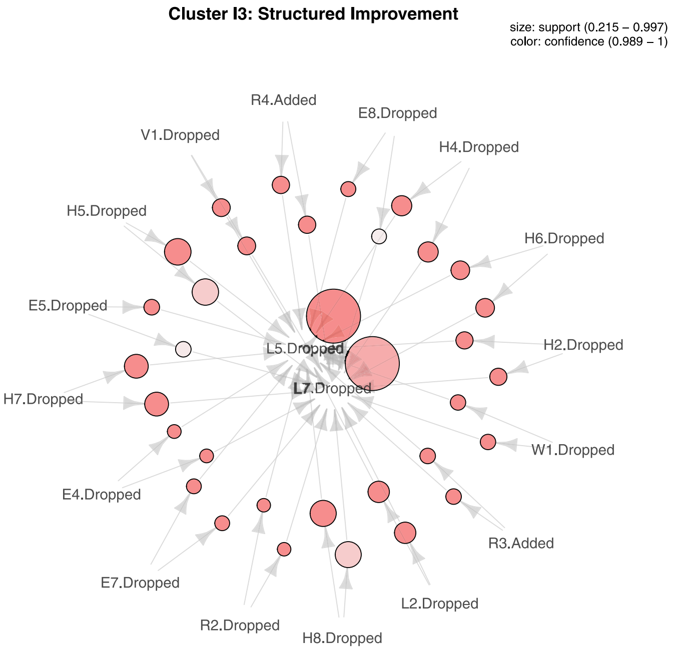


Fig. 12. Association rules of Cluster I3.

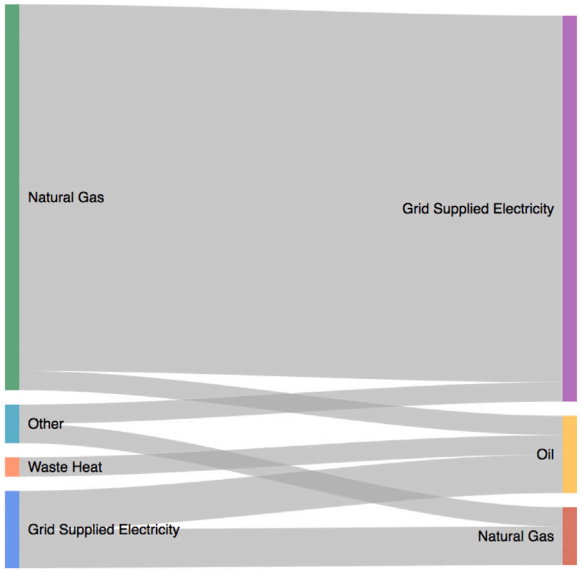


Fig. 13. Heating fuel switch pattern in Cluster D5 (31 buildings).

conditions. Yet, a relatively high percentage of newly recommended solar and wind renewable energy options show rather unrealistic improvement suggestions. The main difference lies in the lighting section: Cluster I3 buildings dropped L5 and L7 (changing inefficient lamps) recommendations, while Cluster I1 buildings barely improved.

The importance of structured moves is also reflected in the result of the association rule analysis. No association rules are found in the changes of recommendation in Cluster I1 buildings, while a centred rule structure is found for Cluster I3, as illustrated in Fig. 12. Each circle represents an association rule, and the arrows directing in and out of the circle show the left and right sides of the rule. For example, the rule {E7.Dropped} ⇒ {L5.Dropped} has E7.Dropped arrow towards the rule circle and followed by another arrow pointing out of the circle pointing L5.Dropped. This means that in Cluster I3, when a building has recommendation E7 dropped in the renewed EPC, recommendation L5 also tends to be dropped. The structure of association rules in Cluster I3 means that L5 and L7 are improved when people choose to improve any other aspect of the building. The absence of rules found for Cluster I1 implies that all changes in the recommendations occur independently. Hence all the reflected changes made to improve buildings are also done independently; no two retrofit options are constantly selected together. Both L5 and L7 are simple improvements that involve only changes in lighting efficiency. We can see that the simple move of further improving the lighting energy efficiency has effectively made Cluster I3 buildings better rated than Cluster I1 buildings and best rated among all clusters. Hence, the simple but effective structure of combining lighting improvement with other planned changes should be encouraged in future building energy efficiency retrofit policies or initiatives. In future versions of EPC recommendation systems, it is necessary to include more structured recommendation bundles instead of single ones and not have a one-size-fits-all approach.

Cluster D5: Inefficient fuel change Cluster D5 has the smallest share of 0.57% but has the most distinct pattern. All the buildings in this cluster

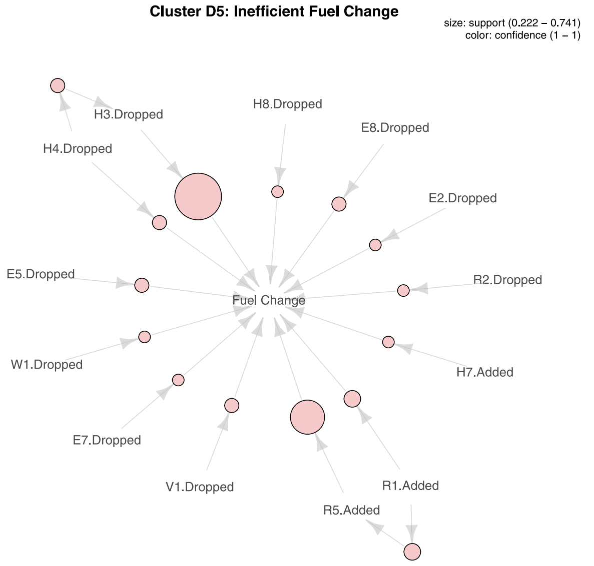


Fig. 14. Association rules of Cluster D5.

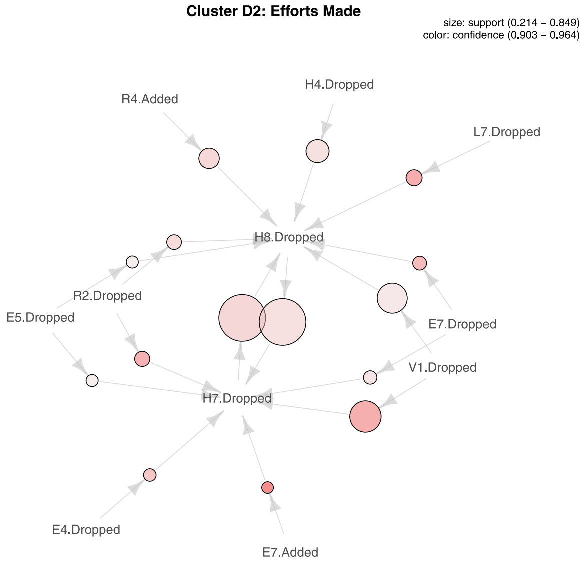


Fig. 15. Association rules of Cluster D2.

have changed their primary heating fuel between the two EPC inspections, as illustrated in Fig. 13. As suggested by the two association rules with the highest support in Fig. 14, namely {H3.Dropped} ⇒ {Fuel Change} and {R5.Added} ⇒ {Fuel Change}, recommendation H3 (property having poor or fair heat generator efficiency and fuel is gas, oil or LPG, as shown in Table 2) dropped in the update as most buildings have changed fuel from Natural Gas to Grid Supplied Electricity. Accordingly, the recommendation system suggests deploying heat pumps (R5/R1) for their next change. Although they show evident efforts in envelope change, together with a greater than average decrease in SER, the non-efficient fuel change did not improve the Asset Ratings of the buildings but led to the second most significant increase (worsened) of Asset Rating. The recommendation system should provide more detailed motivations to directly change to the heat pump. Most fuel changes in these buildings shifted from natural gas to grid-supplied electricity. The EPC system, favouring natural gas heating due to past low prices, should review its evaluation methods and adapt to energy price fluctuations and technological advancements.

5.2.2. Clusters show unreliability of the current system

Cluster I2: Limited actions, Cluster D3: No effort detected & buildings maintained original rating Together with the low number of recommendations added or dropped, Fig. 10 C, D, and E reveal limited changes in EPC updates for Clusters I2, D3, and buildings maintaining original Asset Ratings. Short time gaps between assessments indicate these buildings are more active on the market and have fewer opportunities for significant changes. Unchanged recommendations imply unchanged building conditions. In this case, we cannot summarise primary sources of the Asset Rating improvement (though minimum among improved clusters) of Cluster I2 buildings, nor can we explain Cluster D3 buildings. No association rule was discovered for these groups, showing independent changes in EPC recommendations.

With limited building condition changes and similar patterns across clusters, buildings with good initial ratings are more likely to worsen or maintain ratings, while poorly rated buildings can improve. This aligns with previous analysis findings.

These clusters represent 52.7% of updated buildings. The average SER changes of these buildings are at a reasonably low level means the main source of Asset Rating difference would be the changes in BER. The main difference among these three clusters is the level of the initial Asset Rating. With limited possible building condition changes and similar patterns detected in these buildings across three clusters, it shows that for the similar limited actions taken to change the building conditions, buildings with good initial ratings are more likely to have worse ratings or maintain their initial rating in the EPC update. In contrast, buildings rated poorly initially have the potential to increase their Asset Ratings. This is aligned with our findings in the previous analysis based only on the ratings. We have discussed the possible regression effect of updated Asset Ratings shown in Fig. 6. The negative correlation between the original Asset Rating and changes in the update suggests a possible self-correcting process, indicating EPC generation system inconsistency.

Cluster I4: Possible assessment difference & Cluster D1: Untrustworthy inspections Cluster I4 buildings have twice the recommendations added and the highest average SER increase. The increase in the SER would amplify the decrease in the ratio of SER and BER, resulting in a lower (better) Asset Rating. The great increase in the SER of the buildings hints that the new rating would be based on a very different reference building. It is reasonable to assume a significant share of Asset Ratings' improvement in this cluster originated from the differences in the assessment procedure, hence triggering different recommendations. The differences in the two reference buildings may come from human factors, such as the variations of the inputs from different assessors.

In Cluster I4, the highest percentage of envelope recommendations newly triggered among all improved clusters indicates assessor inspection differences as U-values of building envelopes are not expected to change significantly during a short ageing period. When the SER calculated from the reference building varies significantly, the comparability of the Asset Rating is compromised. The conclusions drawn from the analysis would also be meaningless based on an inconsistent description of the building.

Following the same logic, the increase of SER in cluster D1, which is the highest in the deteriorated clusters, would lead to a decrease in Asset Rating if the BER of the building stays the same. However, the increase of the Asset Rating in the cluster is among the highest ones, which means the BER has worsened by a greater amount as the SER has increased, shown in Table 3. With the fewest original recommendations and most added in the update, either the original or updated EPC may not reflect actual conditions. Though only 2.48% of updated buildings are in this cluster, their EPCs could be misleading and discourage improvements.

Cluster D2: Efforts made In 2.57% of buildings, recommendation changes contradict decreased Asset Ratings. Cluster D2 has the lowest original average Asset Rating and smallest rating change among deteriorated clusters. Fig. 11C shows the highest number of recommendations dropped, while Fig. 15 displays systematic improvement with heating system controls (H7 and H8) dropped. Added envelope recommendations in the EPC update are the main improvement pathway, similar to Clusters I1 and I3.

SER and BER analysis partially explains the contradiction. Low original Asset Ratings and BERs imply a potential for BER increase in EPC updates. The greatest average SER decrease among clusters amplifies BER increases. Hence, the possible efforts made to improve the buildings are offset by the systematic changes in SER and BER.

Cluster D4: Need focus shifting Cluster D4, forming 1.11% of updated buildings, has the most significant average Asset Rating increase. Recommendations associated with the building envelope are issues for these buildings. Despite many recommendations dropped in other sections, envelope recommendations show little reduction and new ones are added (Fig. 11D). Inconsistency between assessors, especially in the envelope sector, is evident. Although the percentage of BER change is lower than Cluster D1, the SER decrease results in the highest Asset Rating increase for Cluster D4. If the building owners have improved other sections, as reflected in the recommendation drop, the envelope inspection inconsistency is discouraging when they a get worse EPC rating in the end.

6. Conclusion and policy implications

We have analysed the EPC updates of simple office buildings. This analysis was formed by both an examination of the reliability of the non-domestic EPC ratings themselves and an exploration of patterns of the EPC updates. The EPC generating process is designed to be a consistent and comparable system with non-significant human factors involved such that the generated Asset Ratings are unbiased. The analysis shows that, in practice, the comparability, consistency and human-factor-free assumptions in the EPC-generating process do not always hold. Half of the buildings analysed have their Standard Emission Rate(SER) unchanged following the design of the system, while the other half have exhibited varying SERs across the updates. We have found evidence showing that even buildings with unchanged conditions can obtain a decreased Asset Rating in the EPC update, with buildings below minimum efficiency level ratings decreasing their Asset Rating even more. Furthermore, for buildings having improved Asset Ratings, their SER tends to increase on average during the update, particularly when the building was initially poorly rated. Conversely, for buildings with deteriorated Asset Ratings, the SER is projected to decrease on average when the building has a low initial Asset Rating. Such properties mean that, systematically, it is easier for buildings with bad initial Asset Ratings to improve and harder for buildings with low initial Asset Ratings to maintain their level. This would jeopardise the fairness of the EPC system and would mislead people’s understanding of buildings’ energy efficiency levels. Most importantly, EPC-based policies like MEES are cast into doubt. On the other hand, policies like MEES with a hard threshold may also affect the reliability of EPCs as they potentially increase the incentives for people to game the EPC system. Alternatives of such policies or modifications with adding retrofit encouraging elements should be considered. Moreover, the EPC system should try to further minimise human factors in the procedure to reduce the bias introduced by possible assessor actions.

With association rule mining, the clustering analysis based on recommendation changes in the EPC update process has discovered patterns of building condition changes and further evidence of EPC inconsistencies. Some clusters show signs of significant human factors involved in the EPC process. Different assessors may produce EPCs that vary greatly and would not be able to give consistent Asset Ratings to the same building. The renovation strategies found in the EPC updates are shown to be reasonable and viable actions as they are people’s natural choices and are already happening in reality. We find policies that encourage people to improve buildings more systematically, like improving lighting conditions along with other planned moves, is a practical next step to further upgrade the energy efficiency of non-domestic buildings as buildings that improved the most are found to be already doing this. Meanwhile, the EPC system should also provide more structured retrofit recommendations to lead building owners to go beyond the lighting section when they try to carry out structured moves. For buildings that have already improved, solar and wind renewable energy have shown to be the primary next step. The future policy should also encourage people who have decided to change their main heating fuel to move directly to the use of heat pumps when conditions permit to avoid inefficient fuel changes. The EPC system may also need to redesign its calculation for electric heating systems.

We have done our analysis based on the best available non-domestic EPC information. This is the first paper that analyses the reliability of non-domestic EPC ratings calculated based on the “reference building” method. The analyses of patterns that emerged from EPC recommendations are also informative, especially when there is little information about the building’s physical conditions in the published non-domestic EPC data. Suppose more details of non-domestic EPCs are published in the future, it is important to justify our findings further with more concrete data, expand the study to other non-domestic buildings and investigate the mechanisms behind the observations, such as the multi-modal phenomenon of SER distribution.

CRediT authorship contribution statement

Mingda Yuan: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Visualization. **Ruchi Choudhary:** Conceptualization, Validation, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

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

Data availability

The data used in this paper is published by the UK government and is publicly available, the link is provided in the article as well

Acknowledgements

This work was supported by AI for Science and Government (ASG), UKRI’s Strategic Priorities Fund awarded to the Alan Turing Institute, UK (EP/T001569/1)

Appendix. Local Weighted Polynomial Regression models

The Local Weighted Polynomial Regression Models used in this study is first proposed by Cleveland (1979), and is realised by applying the “Stats” package in R (R. Core Team, 2022).

The key parameters for the models are Span, Degree and Weights. Span determines the width of the sliding window, when less than 1, it is the proportion of the total dataset that is used in each local fit, and when it is equal or greater than 1, the whole dataset is used. We set the Span to be 0.75. Degree is the overall degree of the locally-fitted polynomial. We set it to 2 which is locally-quadratic fitting to allow flexibility in the fitting process. Weights is an optional expression for weights assigned to individual observations in the sum of squared residuals that forms the local fitting criterion. We use the default option which means an unweighted fit is carried out as our data do not present need for a weighted fit.

References

Ahern, C., Norton, B., Enright, B., 2016. The statistical relevance and effect of assuming pessimistic default overall thermal transmittance coefficients on dwelling energy performance certification quality in Ireland. Energy Build. 127, 268–278.

Amecke, H., 2012. The impact of energy performance certificates: A survey of German home owners. Energy Policy 46, 4–14.

Amirkhani, S., Bahadori-Jahromi, A., Mylona, A., Godfrey, P., Cook, D., Tahayori, H., Zhang, H., 2021. Uncertainties in non-domestic energy performance certificate generating in the UK. Sustainability 13 (14), 7607.

Arcipowska, A., Anagnostopoulos, F., Mariottini, F., Kunkel, S., 2014. Energy performance certificates across the EU–a mapping of national approaches. In: Buildings Performace Institute Europe. BPIE, Brussels.

BEIS, 2017. The non-domestic private rented property minimum standard—Landlord guidance. Department for Business, Energy and Industrial Strategy London, UK, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/824018/Non-Dom_Private_Rented_Property_Minimum_Standard_-_Landlord_Guidance.pdf.

Bishop, C.M., Nasrabadi, N.M., 2006. Pattern Recognition and Machine Learning. Springer.

Burman, E., Mumovic, D., Kimpian, J., 2014. Towards measurement and verification of energy performance under the framework of the European directive for energy performance of buildings. Energy 77, 153–163.

Christensen, T.H., Gram-Hanssen, K., de Best-Waldhober, M., Adjei, A., 2014. Energy retrofits of danish homes: Is the energy performance certificate useful? Build. Res. Inform. 42 (4), 489–500.

Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Amer. Statist. Assoc. 74 (368), 829–836.

Communities, Government, L., 2020. A Technical Manual for SBEM: UK Volume. Communities and Local Government.

Crawley, J., Biddulph, P., Northrop, P.J., Wingfield, J., Oreszczyn, T., Elwell, C., 2019. Quantifying the measurement error on England and Wales EPC ratings. Energies 12 (18), 3523.

Davis, P., McCord, M.J., McCluskey, W., Montgomery, E., Haran, M., McCord, J., 2017. Is energy performance too taxing? J. Eur. Real Estate Res..

Droutsa, K.G., Kontoyiannidis, S., Dascalaki, E.G., Balaras, C.A., 2016. Mapping the energy performance of hellenic residential buildings from EPC (energy performance certificate) data. Energy 98, 284–295.

EPBD, 2010. Directive 2010/31/EU of the European parliament and of the council of 19 may 2010 on the energy performance of buildings (recast). Off. J. Eur. Union 18 (06), 2010.

Fuerst, F., McAllister, P., 2011. The impact of energy performance certificates on the rental and capital values of commercial property assets. Energy Policy 39 (10), 6608–6614.

Fuerst, F., McAllister, P., Nanda, A., Wyatt, P., 2016. Energy performance ratings and house prices in Wales: An empirical study. Energy Policy 92, 20–33.

Gelegenis, J., Diakoulaki, D., Lampropoulou, H., Giannakidis, G., Samarakou, M., Plytas, N., 2014. Perspectives of energy efficient technologies penetration in the greek domestic sector, through the analysis of energy performance certificates. Energy Policy 67, 56–67.

Gouveia, J.P., Palma, P., 2019. Harvesting big data from residential building energy performance certificates: Retrofitting and climate change mitigation insights at a regional scale. Environ. Res. Lett. 14 (9), 095007.

Gupta, R., Gregg, M., 2018. Targeting and modelling urban energy retrofits using a city-scale energy mapping approach. J. Clean. Prod. 174, 401–412.

Hardy, A., Glew, D., 2019. An analysis of errors in the energy performance certificate database. Energy Policy 129, 1168–1178.

Hårsman, B., Daghbashyan, Z., Chaudhary, P., 2016. On the quality and impact of residential energy performance certificates. Energy Build. 133, 711–723.

Hyland, M., Lyons, R.C., Lyons, S., 2013. The value of domestic building energy efficiency—evidence from Ireland. Energy Econ. 40, 943–952.

Jenkins, D., Simpson, S., Peacock, A., 2017. Investigating the consistency and quality of EPC ratings and assessments. Energy 138, 480–489.

Murphy, L., 2014. The influence of the energy performance certificate: The Dutch case. Energy Policy 67, 664–672.

Pasichnyi, O., Wallin, J., Levihn, F., Shahrokni, H., Kordas, O., 2019. Energy performance certificates—New opportunities for data-enabled urban energy policy instruments? Energy Policy 127, 486–499.

R. Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

Regulations, B., 2016. Building Regulations, Approved Document L: Conservation of Fuel and Power. Department for Communities and Local Government (United Kingdom).

Scrucca, L., Fop, M., Murphy, T.B., Raftery, A.E., 2016. Mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. R J. 8 (1), 289.

Semple, S., Jenkins, D., 2020. Variation of energy performance certificate assessments in the European union. Energy Policy 137, 111127.

Streicher, K.N., Parra, D., Buerer, M.C., Patel, M.K., 2017. Techno-economic potential of large-scale energy retrofit in the Swiss residential building stock. Energy Procedia 122, 121–126.

Tronchin, L., Fabbri, K., 2012. Energy performance certificate of building and confidence interval in assessment: An Italian case study. Energy Policy 48, 176–184.

von Platten, J., Holmberg, C., Mangold, M., Johansson, T., Mjörnell, K., 2019. The renewing of energy performance certificates—Reaching comparability between decade-apart energy records. Appl. Energy 255, 113902.

Yuan, M., Choudhary, R., 2020. A two-step clustering framework for locally tailored design of residential heating policies. Sustainable Cities Soc. 63, 102431.