sustainability

MDPI

Article

Measurement Timing and Endogeneity in EPC-Based Policy Classification: Regression Discontinuity Evidence from High-Floor Residential Properties in England and Wales

Chaiyathad Phutthadet [1] and Kongkoon Tochaiwat [2,*]

[1] School of Business and Communication Arts, University of Phayao, Phayao 56000, Thailand; chaiyathad.ph@up.ac.th
[2] Faculty of Architecture and Planning, Thammasat University, Pathum Thani 12121, Thailand
* Correspondence: kongkoon@tu.ac.th

Abstract

This study examines EPC Band C threshold status and subsequent recorded EPC-score changes in high-floor residential buildings in England and Wales (2013–2023), using Energy Performance Certificate (EPC) data. A previously unreported panel-construction problem, Unique Property Reference Numbers are occasionally recorded at a building rather than unit level, is identified and addressed using a conservative same-unit filtering procedure. A threshold-based eligibility construction is shown to be mechanically related to recorded EPC improvement when measured from a dwelling's post-improvement rather than baseline score: this timing choice alone accounts for the entire positive association found in a naive baseline specification. Measured instead from the baseline score, the association reverses its sign under every panel-construction and threshold combination tested, a reversal corroborated by a regression discontinuity design using a different estimation strategy on the same score variable, though the estimated magnitude is sensitive to bandwidth choice. Two secondary strategies examining whether interest rates moderate this relationship produced no credible estimates, so this question remains open. The findings indicate that eligibility indicators built from the same data used to measure outcomes require explicit attention to measurement timing, with implications for evaluations using EPC thresholds as proxies for financial or policy eligibility.

Keywords: green finance; energy retrofit; regression discontinuity design; measurement bias; energy performance certificates

Check for updates

Academic Editor: Chunlu Liu

Received: 25 July 2026
Revised: 25 August 2026
Accepted: 31 August 2026
Published: 2 September 2026

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Decarbonizing the built environment is a central component of global sustainability transitions. Buildings and construction account for a substantial share of global energy demand and energy-related emissions [1,2], and because much of the housing stock already exists, retrofit investment is widely recognized as essential for reducing emissions [1,2]; the methods and effects of such retrofits are examined in the growing literature [3–5].

Energy-performance labels and related classifications increasingly inform housing decarbonization policies and green-housing finance instruments [6]. More broadly, capital markets mobilize climate finance through a growing range of green financial instruments [7]. Underlying this expanding use is an assumption receiving limited direct scrutiny: that indicators classifying a building, loan, or borrower as "green" or "eligible" can be treated as reliable, exogenous markers of a pre-existing condition, rather than as constructions

sensitive to how and when they are measured. Certification-based studies suggest that green labels need not map one-to-one onto realized energy or carbon outcomes [8], while green-certified properties can also differ in mortgage default risk and loan terms [9].

This paper examines that assumption directly, using building-level EPC data for high-floor residential buildings in England and Wales (2013–2023). It is fundamentally a measurement and identification paper: the central question is whether, and how, the timing convention used to construct an EPC-based eligibility indicator affects its estimated relationship with the outcome it is used to explain, rather than an evaluation of any specific green-finance instrument's effectiveness.

As in the wider literature, the initial aim was to estimate the relationship between EPC Band C threshold status, monetary policy conditions, and recorded EPC-score changes. This was motivated by the broader literature asking whether the effectiveness of green-finance instruments depends on macro-financial conditions, since monetary policy affects borrowing costs and credit availability in capital-intensive sectors such as real estate [10–12], with the related literature on green central-bank policy examining how monetary policy affects the financing costs of low-carbon firms specifically [13]. Green finance more broadly is also associated with regional energy-transition outcomes [14], and is intended to mobilize capital toward environmentally beneficial activities including energy-efficiency projects [15]. The data-quality audit and identification strategies point elsewhere.

The relationship, reported using the study's initial post-outcome convention, appears to depend substantially on how eligibility is timed relative to the outcome it predicts, rather than solely on a causal effect of green-finance access. This finding forms the paper's main contribution; the role of monetary policy conditions is examined as a secondary, exploratory objective, discussed further in Section 3.3.2.

This paper makes two contributions, relevant to research on green finance, retrofit investment, and building decarbonization. The first concerns a previously unreported panel-construction problem in building-level EPC register analysis, addressed here using a conservative same-unit filtering procedure. Unique Property Reference Numbers (UPRNs), the standard linkage key for EPC-based property datasets, are occasionally recorded at the building rather than the unit level. This causes naïve panel construction to compare physically distinct units as repeated observations of one property. Recent EPC-based property modelling studies combine certificate, textual, and spatial information for energy-performance prediction, but longitudinal verification of whether repeated administrative records represent the same physical dwelling is not their principal methodological focus [16].

The second, more central contribution shows that a threshold-based eligibility construction can become close to definitionally related to recorded EPC improvements when eligibility is measured at the same inspection as the outcome: a dwelling with no improvement can never newly cross the threshold, by construction. This timing choice alone accounts for the entire positive eligibility–retrofit association produced by the study's initial post-outcome specification. Whether the same mechanism explains findings elsewhere in the literature cannot be confirmed until a study using the same construction is identified; measuring eligibility from the baseline score instead reverses the sign under every combination tested here.

This second contribution is corroborated by a regression discontinuity design, applied around the eligibility threshold using the baseline EPC score as the running variable, which addresses the concern that eligibility remains endogenous even once correctly timed. This provides evidence under a different estimation strategy: despite drawing on the same underlying score variable, it is consistent with the same reversal, though the magnitude is sensitive to bandwidth and should be read as suggestive rather than conclusive. Two

further strategies, a difference-in-differences design around the 2018 MEESs mandate and a predetermined exposure-by-rate interaction for interest-rate exposure, did not produce credible estimates; the moderating role of monetary policy is left open for future research.

These findings carry a methodological implication for the wider literature on green finance and building decarbonization: eligibility and compliance indicators derived from the same energy-performance data used to measure recorded EPC-score-change outcomes should be constructed and reported, with attention to measurement timing, since the direction of an unadjusted correlation can depend on this choice alone. They also carry a practical implication for policy evaluation: administrative green-finance eligibility flags cannot be assumed to be exogenous markers of a building's baseline condition without verification.

The remainder of the paper is structured as follows. Section 2 reviews the relevant literature on energy efficiency investment, green finance, measurement and classification challenges in sustainability metrics, and monetary transmission. Section 3 develops the conceptual framework and revised research hypotheses. Section 4 describes the data, the panel-construction audit, and the identification strategies. Section 5 presents the empirical results. Section 6 discusses the findings, their policy implications, and limitations. Section 7 concludes the manuscript.

2. Literature Review

2.1. Energy Efficiency Investment in Residential Buildings

Investment in energy efficiency improvements has been widely studied in the context of the "energy efficiency gap", the persistent underinvestment in energy-saving technologies despite potential economic benefits [17]. Barriers include information asymmetries, behavioural biases, and split incentives between landlords and tenants; studies of low-income Australian households identify these barriers alongside financial constraints in that specific context [18].

Among these explanations, reviews focusing on households in energy poverty identify financial and institutional constraints as major barriers to retrofit investment in that population specifically [19]: energy efficiency upgrades require substantial upfront capital expenditure, and the high upfront cost together with homeowners' willingness to pay are important determinants of retrofit investment [20], while financing constraints are frequently discussed as potential barriers in the literature reviewed above. Financing instruments, including grants and loans, are recurring elements of effective retrofit-policy design [21].

2.2. Green-Finance Instruments and Additionality

Green finance has increasingly been recognized as a mechanism intended to mobilize capital toward sustainable investments, with country-level empirical evidence linking green finance, financial globalization, and environmental sustainability [22]. Instruments such as green mortgages and preferential lending schemes support investment in energy-efficient buildings [6,23], and green finance more broadly is associated with regional green-finance development and broader energy-transition outcomes [14]; incentive-planning and optimization frameworks are designed to determine incentive levels intended to promote selected retrofit strategies in the building sector [24]. Preferential lending and related instruments are intended to reduce upfront financing barriers and improve the affordability of renovation investment [23].

A related but distinct question, more prominent in the corporate green-bond literature than in residential retrofit research, is whether green-finance instruments generate additional investment that would not otherwise have occurred, or whether they primarily

relabel investments that borrowers were already undertaking or intending to undertake: the question of additionality [25].

This question has particular force here: if eligibility is measured using a dwelling's energy performance after, rather than before, a retrofit event, a positive association can arise by construction, independent of whether the instrument induced any additional investment. Existing retrofit-finance reviews discuss financing conditions and programme qualification requirements mainly at the instrument or policy level [23], while regional green-finance studies use aggregate financial-development measures [14]. These strands do not directly establish how a property-level EPC threshold indicator should be timed relative to a recorded EPC-score-change outcome.

2.3. Energy Performance Certificates as Policy Instruments

Energy Performance Certificates (EPCs) have become a central policy instrument across UK building and finance policy, extending well beyond their original purpose of informing prospective buyers and tenants about a dwelling's energy performance at the point of sale or letting. EPC scores and bands are now embedded directly into regulatory and financial mechanisms with material consequences for property owners. Green mortgage products condition preferential lending terms on a property's EPC band. Individual-mortgage green-product eligibility at major UK high-street lenders is typically documented to require Band A or B (SAP $\geq$ 81), not Band C, though NatWest has separately adopted a portfolio-level ambition for 50% of its mortgage book to be at or above EPC Band C by 2030 [26]. Current NatWest and Barclays product terms both require a valid EPC rating of A or B (or an energy efficiency score of 81 or above) for their respective green mortgage products. NatWest's separate Band C target applies to the composition of its mortgage portfolio, not to individual-loan eligibility [26,27]. Product criteria of this kind are time-varying, so they are reported here with the relevant access date, not as a stable, generally applicable standard. The Minimum Energy Efficiency Standards (MEESs) regulation similarly uses EPC bands to define compliance obligations for privately rented properties. This was phased in over two stages: from April 2018 the standard applied to relevant new and renewed domestic tenancies, and from April 2020 it extended to continuing lettings, prohibiting landlords from letting or continuing to let F- or G-rated dwellings unless improved or covered by a registered exemption, with a minimum EPC E threshold [28]. EPC bands also structure two differently designed national programmes. The Energy Company Obligation is placed on energy suppliers under the Electricity and Gas (Energy Company Obligation) Order 2022, with local-authority flexible-eligibility routes; it can cover dwellings across EPC Bands D–G and sets upgrade targets by band [29,30]. The Warm Homes: Social Housing Fund is a government-funded programme aimed specifically at upgrading social housing below Band C to Band C [31].

This expanding policy role rests on an implicit assumption that the EPC score is a reliable, comparable measure of a dwelling's underlying energy performance, stable enough to support binary eligibility and compliance decisions with direct financial consequences. Yet EPCs for existing dwellings are generated through the Reduced Standard Assessment Procedure (RdSAP), a methodology that relies on assessor-recorded building characteristics rather than measured energy consumption using a reduced-data methodology that infers missing inputs via standardized assumptions rather than direct measurement [32], and whose calculation methodology has been revised on several occasions over the period covered by most retrofit studies, one source of measurement variation among several [33]. This is not a theoretical concern. A national-scale analysis of 1.6 million repeated EPC assessments in England and Wales estimated a one-standard-deviation error that declines with EPC performance, from approximately $\pm$8 points near a rating of 35 to approximately

±2.4 points near a rating of 85—large enough in the lower range to shift a dwelling across an official EPC band [33]. A separate audit of the open EPC database found that 27% of records were flagged by specific data-quality rules, from which an estimated underlying error rate of between 36% and 62% was derived. This was disproportionately concentrated in flats and maisonettes, the property type examined in this study [34]. Whether an indicator subject to this documented level of assessor-driven measurement error can bear the weight now placed on it as an eligibility and compliance instrument, particularly when the same indicator is later used to measure the outcome it was meant to help finance, has received comparatively little direct scrutiny in the literature reviewed above.

2.4. Measurement Validity and Classification Challenges in Sustainability Metrics

The broader measurement literature, spanning building physics as well as econometrics, offers conceptual grounding for this concern. Indicator validity, the extent to which an observed proxy captures the underlying construct it is intended to represent, is a first-order data-quality concern whenever administrative or regulatory records are repurposed for research or for automated eligibility determination. This concern is closely related to the measurement uncertainty routinely quantified in building energy modelling; calibration studies explicitly recognize parameter and model uncertainty in retrofit-oriented building-energy analysis [35]; input-data quality has separately been shown to distort evaluation error in a low-income household retrofit programme [36]. Administrative records of the kind analyzed here are typically collected to support an operational function, not a research design, and are consequently susceptible to classification error introduced by inconsistent recording practice, assessor discretion, or revisions to the underlying methodology over time [33,34], rather than by the sampling error that dominates concerns about survey-based measurement. Reproducibility of findings derived from such records depends on documenting these data-quality characteristics, rather than treating administrative fields as ground truth by default.

A sharper concern arises when the timing of the measurement interacts with the causal or classificatory structure of the analysis. Variables measured after treatment can introduce post-treatment bias when used as controls, depending on their causal role and the estimand of interest [37,38]; this is not a blanket prohibition on post-treatment variables, and the same literature is discussed further in Section 2.6. The present study examines a related but distinct timing problem in which a post-outcome measure is used to construct the eligibility indicator itself. This timing-based concern is conceptually distinct from the more familiar selection-on-observables form of endogeneity: it does not require any confounding relationship between eligibility and the determinants of the outcome, since the bias instead arises structurally from the measurement convention itself, independent of any economic mechanism connecting eligibility to the outcome it is used to explain.

ESG ratings, one class of sustainability classification, have been shown to diverge systematically across providers because of differences in scope, measurement, and weighting [39].

Within the specific context of UK residential energy performance data, the Energy Performance Certificate score used throughout this study, and in the wider literature on UK housing retrofit, is generated by assessors applying the Reduced Standard Assessment Procedure (RdSAP), a methodology that has been revised on several occasions during the period covered by this study [33]. Related EPC-based research has also documented social and spatial patterns of residential energy efficiency in England [40].

Direct evidence of this is presented in Section 4: the construction-age-band field disagrees between successive inspections of the same dwelling at a rate rising smoothly from 16% (2008) to 58% (2020 onward), consistent with methodology revision rather than genuine identity change, and the energy-efficiency score shows heaping consistent with assessor

rounding rather than a smooth distribution. The literature search undertaken for this study did not identify a prior study documenting this specific interaction between EPC measurement characteristics and eligibility-construction timing, or showing that an undocumented measurement-timing convention can reproduce the entire positive eligibility–retrofit association produced by this study's own initial post-outcome specification; whether prior analyses share this construction is not established here.

2.5. Administrative Data and Reproducibility

Separate research, situated primarily in applied microeconometrics and official statistics rather than in energy or housing research specifically, addresses the challenges of constructing research panels from administrative record-linkage systems. Administrative registers are typically designed to support a specific operational function, such as property transaction registration or, in the case of the EPC register, regulatory disclosure, rather than to provide a stable longitudinal identifier for research purposes. Record-linkage methodology recognizes that identifier fields intended to track a single unit over time can be assigned inconsistently, particularly where the underlying administrative system was not designed with research reproducibility as a priority [41]; the EPC register's own guidance separately notes anomalies in the published bulk data [42]. Linkage errors of this kind are distinct from measurement errors in the outcome variable itself: they affect which observations are treated as repeated measures of the same unit, and can contaminate panel-based estimates without being detectable from the resulting estimates alone.

Within the EPC register specifically, the Unique Property Reference Number (UPRN) is the standard field used to link successive inspections of a dwelling into a panel. Although EPC-based studies increasingly integrate certificate records with property-level and spatial information, the literature reviewed for this study did not identify a documented procedure specifically validating whether consecutive same-UPRN records refer to the same physical dwelling. Section Constructing a Conservatively Filtered Same-Unit Panel documents that this assumption does not hold uniformly across the register and proposes a conservative same-unit confirmation filter, reproduced across two independent computational pipelines, which may be of independent use to other researchers constructing panels from this or similarly structured administrative sources; this filter has not been checked against an external ground truth such as an address gazetteer, Land Registry records, or a manually verified sample, and is not described as validated in that sense.

2.6. Endogeneity and Post-Treatment Bias in Evaluation

The broader econometric literature recognizes that conditioning on a variable measured after treatment can introduce bias of an unpredictable sign, a concern discussed as "bad control" or post-treatment bias [37,38].

The retrofit literature more commonly examines substantive barriers to energy-efficient design and construction and homeowners' perceived costs and benefits [43,44], rather than the timing of outcome-adjacent classification variables. RDDs offer methodological precedents for addressing related selection concerns around policy thresholds in other empirical contexts, including energy-consumption responses to price shocks and running-variable choices in regional business-incentive evaluation, rather than green-finance retrofit evidence specifically [45,46].

Our analysis connects measurement timing and threshold-based selection, showing both occur in one construction of the EPC-based threshold status and must be addressed jointly.

2.7. Monetary Policy Transmission and Capital-Intensive Investment

Macroeconomic conditions shape corporate financing and investment decisions more broadly [47]. Monetary policy influences financial conditions through interest-rate adjustments and credit market dynamics [48], changes in borrowing costs affect investment in capital-intensive sectors such as real estate [10,49], and monetary policy can affect corporate investment more generally through firms' balance sheets and financing structures [50], but comparatively little attention has been paid to how it interacts with sustainability-related investment specifically. As shown here, isolating that interaction turns out to face substantial identification challenges of its own, as detailed in Section 4.4.2 and Supplementary File S1; accordingly, this study treats the interest-rate channel as a secondary, exploratory objective rather than a co-equal focus, subordinate to the measurement-timing contribution that motivates the paper.

2.8. Research Gap

The retrofit-financing and housing-decarbonization literature [6,23] discusses green-finance eligibility conditions descriptively, at the level of policy instruments rather than as a property-level empirical treatment, while regional green-finance studies [14] measure green finance as an aggregate financial-development construct rather than a threshold-based eligibility indicator; neither strand examines whether the timing convention used to construct an eligibility measure, before or after the retrofit event it helps explain, materially affects the estimated relationship with recorded EPC-score-change outcomes, because neither strand constructs such a property-period-specific measure in the first place. A further gap concerns the administrative data underlying this literature. The literature search undertaken for this study did not identify a documented procedure specifically assessing whether successive same-UPRN records refer to the same physical unit. Section Constructing a Conservatively Filtered Same-Unit Panel therefore applies a conservative same-unit confirmation filter, reproduced across two computational pipelines, while recognizing that the procedure has not been externally validated against an independent ground-truth source. Taken together, the retrofit-financing and housing-policy literature (Sections 2.1–2.3) does not construct a property-level green-finance eligibility variable in the first place, so it neither treats such a variable as exogenous nor examines its timing; the measurement and administrative-data literature (Sections 2.4 and 2.5), largely situated in other empirical domains, has not previously been connected to green-finance eligibility construction specifically; and the post-treatment-bias literature (Section 2.6) provides the conceptual apparatus for why this connection matters, without having been applied to this setting. The literature search undertaken for this study did not identify a previous study that jointly examines measurement timing, administrative record linkage, and causal identification within EPC-based threshold-status evaluations of the kind used to proxy green-finance eligibility. Table 1 summarizes these literature streams and the specific gap each leaves open.

This study addresses this gap: the measurement-timing choice is associated with a reversal in the sign of the estimated eligibility–retrofit relationship, corroborated by an RDD applying a different estimation strategy to the same underlying score variable, while a companion attempt on monetary policy conditions does not yet yield a credible estimate.

Table 1. Synthesis of literature streams reviewed in Section 2, indicating how each treats eligibility or a related construct, whether timing is addressed, and the gap this study addresses. Streams are characterized only to the extent that their own stated scope permits; several are policy reviews, cross-domain measurement studies, or general econometric treatments not designed to estimate a property-level green-finance eligibility variable, and are described as such, not assessed against a standard they were not built to meet.

Literature Stream	Treatment of Eligibility or Related Construct	Timing Addressed?	Gap Addressed in This Study
Retrofit-finance instruments and housing-decarbonization policy [6,23]	Financing conditions and qualification requirements are discussed descriptively or institutionally; no property-level eligibility treatment is estimated.	Not applicable to an empirical property-level treatment.	Section Timing of the Band C Threshold Status Measure operationalizes eligibility as a property-period-specific measure and examines its relation to observed recorded EPC-score-change outcomes.
Regional green-finance and energy-transition studies [14]	Green finance is measured as a regional financial-development construct rather than as threshold-based property eligibility.	Province-year variation is observed, but eligibility timing is not defined.	Section Timing of the Band C Threshold Status Measure directly distinguishes eligibility measured before and after the retrofit event.
Corporate green-bond studies [25]	Green-bond issuance and third-party certification are observed financing events; no EPC-based property eligibility threshold is constructed.	Yes; announcement and post-issuance timing are central.	This study separates the timing of eligibility measurement from the timing of the subsequent recorded EPC change.
Green-building certification and mortgage performance [9]	LEED and ENERGY STAR status is modelled as time-varying: the analysis distinguishes buildings that were certified green at loan origination from those becoming certified later.	Yes, certification timing relative to loan origination and subsequent loan performance is used directly in the identification strategy, although not as a pre- versus follow-up EPC eligibility distinction.	Section 2.4 extends certification-based measurement to EPC threshold status.
ESG and sustainability measurement divergence [39]	Ratings differ systematically because of scope, measurement, and weighting choices.	Not applicable to treatment timing.	Section 2.4 identifies measurement timing as an additional source of classification discrepancy for EPC-based eligibility.
EPC-based energy-performance prediction [16]	EPC, text, and spatial data are combined to predict energy-performance scores and retrofit scenarios.	Longitudinal same-unit eligibility timing is not the principal research question.	Section Constructing a Conservatively Filtered Same-Unit Panel applies a conservative same-unit confirmation filter, reproduced across two computational pipelines, before constructing threshold-status histories.
Program-evaluation and post-treatment-control econometrics [37,38]	Provides general treatment-effect and post-treatment-variable principles rather than a green-finance eligibility application.	Conceptually addressed, but not in the EPC green-finance setting.	Section 3.3.1 applies these principles to distinguish baseline eligibility from eligibility generated by the retrofit itself.

3. Conceptual Framework and Hypotheses

3.1. EPC Classification, Financial Context, and Measurement Timing

Energy retrofit investment in residential buildings is influenced by financial conditions, particularly borrowing costs and access to credit. From a macro-financial perspective, monetary policy affects investment decisions through several transmission channels, including interest rates and credit availability [11,12]. In capital-intensive sectors such as real estate, changes in borrowing costs can significantly influence real-estate market behaviour and prices [49]; higher interest rates increase financing costs for property owners, which this paper's H2 examines as a possible channel discouraging investment in energy retrofit activities specifically.

Green-finance instruments have emerged as mechanisms designed to alleviate these financing constraints for sustainable investments. Financial products such as green mortgages and preferential lending schemes aim to improve access to credit and reduce the effective cost of capital for energy-efficient properties [23]. Because these instruments operate within broader financial markets, their effectiveness may in principle depend on prevailing macro-financial conditions: this paper hypothesizes that when borrowing costs increase due to monetary tightening, the financial advantages associated with green finance may become less effective in encouraging retrofit investment, drawing on evidence regarding green central-bank policy transmission [51] and on the use of green bonds in financing building retrofits, including their limitations as a financing tool [52].

This framework motivates a straightforward prediction. Under the institutional rationale motivating the initial specification, properties classified at or above an EPC threshold used as a proxy for a favourable policy or financial status would be expected to exhibit greater subsequent recorded EPC improvements, an association that may narrow as interest rates rise. It also contains an assumption interrogated directly here: that this threshold classification can be observed independently of the outcome it helps explain. In practice, eligibility is typically built from the same performance data used to measure retrofit activity. This creates two concerns: eligibility may be endogenous to unobserved characteristics correlated with retrofit propensity, and its value can depend structurally on measurement timing. Section 4 shows the second concern is not merely hypothetical here.

3.2. Conceptual Framework

Figure 1 illustrates the conceptual framework, distinguishing links this study empirically evaluates from links that remain theoretical. Baseline EPC Band C threshold status (Section 4.2) feeds into the subsequent recorded EPC-score change through a link empirically evaluated via regression discontinuity (Section 4.4.1); the monetary policy rate feeds into the same outcome through a hypothesized interest-rate transmission channel that could not be credibly identified (Section 4.4.2). The recorded EPC-score change is used throughout as an observable proxy for retrofit investment and energy efficiency improvement, which remain unobserved constructs in this study's data; this proxy relationship in turn connects to broader urban sustainability transitions.

The framework integrates insights from the retrofit financing literature [23], evidence that the timing and funding design of retrofit programmes can influence wider economic outcomes [53], the financing constraint literature [51,52], and the energy efficiency investment literature [54,55]. Unlike the original framework, this version does not treat both links as equally established a priori: the tested, solid-line link (Band C threshold status) and untested, dashed-line link (monetary policy rate) are distinguished in Figure 1, in direct response to the identification issues addressed in Section 4.4.

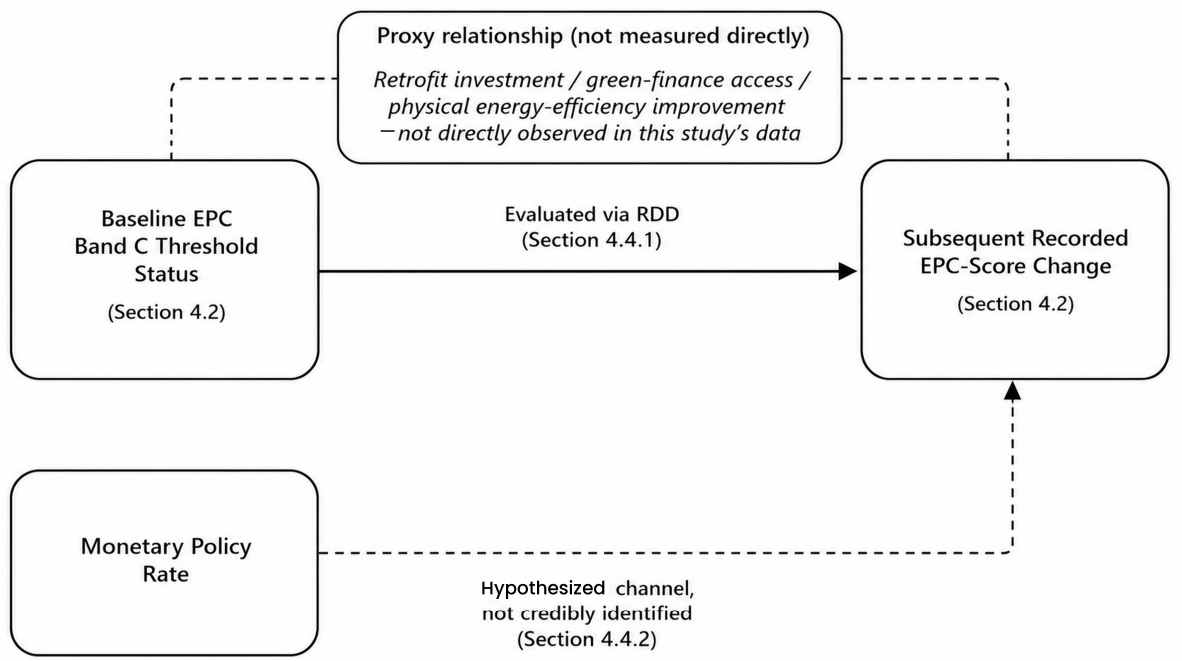

Figure 1. Conceptual framework linking baseline EPC Band C threshold status and monetary policy conditions to subsequent recorded EPC-score change. Retrofit investment, green-finance access, and physical energy-efficiency improvements are not directly observed in this study's data and are not shown as measured constructs. The solid line indicates the link empirically evaluated in this study via regression discontinuity around the Band C threshold; the dashed line indicates the hypothesized interest-rate transmission channel, which could not be credibly identified with the available data (Section 4.4.2).

3.3. Post Hoc Analysis Framework

The hypotheses below are developed in light of the identification strategies in Section 4.4 and the results in Section 5; they are accordingly presented as a post hoc analysis framework rather than in the conventional ex ante hypothesis-testing format, so that the reader understands the order in which the analysis proceeded. The underlying research question concerns how the EPC Band C threshold status is associated with the subsequent recorded EPC-score change, and whether macro-financial conditions may moderate that association; this question is both relevant and only partially answered by the literature. As shown below, the direction of one relationship differs from what a conventional specification suggests, and two others cannot be tested with the available data.

3.3.1. H1

Under the institutional rationale motivating the initial specification, properties classified at or above an EPC threshold used as a proxy for a favourable policy or financial status were expected to exhibit a greater subsequent recorded EPC improvement, and an initial specification corresponding to Equation (2) (Section 4.3), using recorded EPC improvement as the dependent variable, follow-up-score Band B threshold status, the filtered panel, postcode and year fixed effects, and the policy-rate interaction, indeed yields a correlational estimate consistent with this (threshold-status coefficient: +0.311, $p < 0.01$); this differs from the simplified follow-up/filtered/Band B specification reported in Table 2 (+0.281, Section Timing of the Band C Threshold Status Measure) only by the inclusion of the policy-rate interaction term. Yet there is a specific mechanical reason to doubt this estimate: buildings that already meet the eligibility threshold are, by construction, more energy-efficient at baseline, and may therefore be both more likely to be classified as eligible and more likely to undertake further improvement for reasons unrelated to green finance, or, conversely, to

have less remaining scope for improvement left. Investigating this concern revealed a more fundamental problem than endogeneity alone, motivating the hypothesis below.

Table 2. Sensitivity of the EPC threshold-status coefficient to measurement timing, panel construction, and threshold choice. Dependent variable: recorded EPC improvement, a binary indicator equal to 1 when the later matched score exceeds the baseline score and 0 otherwise (Section 4.2); coefficients are reported in probability units; for example, −0.288 corresponds to a 28.8-percentage-point difference. The naive baseline specification (Section 3.3) yields +0.311. All specifications: postcode and year fixed effects, clustered standard errors by postcode. The filtered-panel N reported here (15,281) differs by 4 observations from the 15,277 figures used in Section Constructing a Conservatively Filtered Same-Unit Panel and Table 3, reflecting a minor difference in missing-value handling between the original production run and this verification pass; this does not affect any reported coefficient beyond rounding.

Measurement Timing	Panel Construction	EPC Classification Threshold	Coefficient (β)	95% CI	N	*p*-Value
Baseline	Unfiltered	Band C (SAP $\geq$ 69)	−0.319	[−0.359, −0.278]	20,972	<0.001
Baseline	Unfiltered	Band B (SAP $\geq$ 81)	−0.472	[−0.514, −0.430]	20,972	<0.001
Baseline	Filtered	Band C (SAP $\geq$ 69)	−0.288	[−0.328, −0.248]	15,281	<0.001
Baseline	Filtered	Band B (SAP $\geq$ 81)	−0.450	[−0.494, −0.407]	15,281	<0.001
Follow-up	Filtered	Band C (SAP $\geq$ 69)	0.244	[0.209, 0.280]	15,281	<0.001
Follow-up	Filtered	Band B (SAP $\geq$ 81)	0.281	[0.249, 0.313]	15,281	<0.001
Follow-up	Unfiltered	Band C (SAP $\geq$ 69)	0.172	[0.142, 0.203]	16,540	<0.001
Follow-up	Unfiltered	Band B (SAP $\geq$ 81)	0.250	[0.220, 0.279]	16,540	<0.001

Table 3. Descriptive statistics, reconstructed high-floor panel.

Statistic	Value
Filtered panel observations	15,277
Unique properties	12,161
Mean recorded EPC improvement	58.30%
Mean recorded EPC-score change (SAP points)	3.16
Properties at or above the Band C threshold (SAP $\geq$ 69)	68.60%
Properties at or above the Band B threshold (SAP $\geq$ 81)	9.80%
Baseline score: mean	65.32
Baseline score: SD	14.62
Baseline score: median	70
Baseline score: range	1 to 104
Baseline score: deciles (P10–P90)	44, 57, 62, 67, 70, 73, 74, 76, 79
Recorded EPC improvement: SD	0.493
Recorded EPC-score change: SD	10.12
Recorded EPC improvement: range	0 to 1
Recorded EPC-score change: range	−59 to 67

H1: *If threshold status is constructed from the baseline rather than the follow-up EPC score, thereby avoiding direct incorporation of the subsequent score-change outcome, the positive association in the initial specification will reverse in sign, becoming negative rather than merely attenuated.*

This prediction is tested directly in Section Timing of the Band C Threshold Status Measure, which examines the sensitivity of the Band C threshold-status coefficient to measurement timing, and evaluated against the baseline estimates reported in Section 5.2.

Section 5.3 provides a second test of H1, using the regression discontinuity design described in Section 4.4.1; because this design uses the baseline score as its running variable, it is immune to the timing problem addressed by the test above. The two tests differ in

estimation strategy, full-sample regression versus local randomization around a threshold, but draw on the same underlying panel and the same baseline score variable, so they should be read as two specifications of related evidence rather than as fully independent confirmations. Both tests are reported and interpreted together in Sections 5.2 and 5.3.

3.3.2. H2 and H3

As a secondary, exploratory objective, subordinate to H1, a further hypothesis is that higher interest rates would reduce retrofit investment (H2) and weaken the positive influence of Band C threshold status (H3). H2 and H3 face distinct identification problems: the policy rate's main effect varies only by year and is collinear with the year fixed effects, so it cannot be separately identified (H2), while its interaction with threshold status retains cross-sectional variation and is statistically estimable but causally uninterpretable, given the endogenous construction of threshold status, common shocks correlated with both variables, and limited rate variation over most of the sample period (H3; Section 4.4.2). Two strategies were attempted to resolve this (Section 4.4.2): a difference-in-differences design around the 2018 MEESs introduction, confounded by a pre-trend predating the policy by several years, and a predetermined exposure-by-rate interaction, driven almost entirely by 2023, when rate increases coincided with the European energy price crisis; neither produced a stable estimate.

H2 and H3 are therefore reported as untested rather than supported by an untrustworthy estimate. This is discussed further as a limitation in Section 6.3, alongside the data (area-level mortgage-type or rate-exposure data spanning a period with more interest-rate variation than 2013–2021) needed to test them credibly in future work.

3.3.3. Conceptual Framework: Acknowledged Scope Limitation

A further consideration is that the conceptual framework (Section 3.2) does not model the decision-making of individual stakeholders (property owners, lenders, energy service companies) whose incentives jointly determine whether an eligible building is actually retrofitted. This is a genuine limitation not resolved here, focusing on establishing credible identification for the aggregate relationships already specified, and is flagged as a direction for future work (Section 6.3): a stakeholder-level extension incorporating household risk aversion, information frictions, and lender underwriting behaviour.

4. Data and Methodology

4.1. Data Sources and Sample Construction

The study uses building-level EPC register data for England and Wales, drawn from the full public register (29,069,352 records), not a pre-aggregated extract, so every filtering decision is transparent. Scotland and Northern Ireland maintain separate registers, regulatory arrangements, data structures, and potentially different implementation versions or conventions, and are not included. Existing Scottish dwellings are, in fact, assessed using RdSAP [56], the same underlying methodology used in England and Wales, so the exclusion is not based on a methodological incomparability claim; rather, the separate registers, administrative conventions, and regulatory context are excluded to preserve a consistent administrative and regulatory sample. Including them would introduce a further measurement-consistency problem of the same kind this paper investigates, rather than help resolve the one already under study; this restricts generalization of the findings to the England and Wales regulatory context specifically. Attention is restricted to flats and maisonettes: the FLOOR_LEVEL field, used to identify high-floor dwellings, is missing for 99.5–100% of houses and bungalows but only 0.2–0.5% of flats and maisonettes. Evaluating

coverage only after this restriction avoids treating register-wide missingness (65–70%) as a genuine data-quality problem.

Within this population, certificates lodged between 2013 and 2023 with a non-missing Unique Property Reference Number (UPRN) and a current energy-efficiency score of 150 or below are retained (224 records with implausible scores, evidently data-entry errors, are dropped). The sample is then restricted to dwellings on the sixth floor or above, yielding 132,738 certificate records.

The study operationally defines the sample as flats and maisonettes recorded on floor six or above. This high-floor proxy cannot confirm statutory building height, total storey count, or coverage under the England-specific higher-risk-building regime established by the Building Safety Act 2022 [57], and is not presented as equivalent to that legal definition. The focus on this population is nonetheless substantively motivated, not an arbitrary restriction: high-floor retrofit work plausibly involves higher per-unit cost and greater technical complexity than low-floor retrofit, given shared building fabric, restricted access, and common-parts works requiring building-wide coordination; this study does not test this cost-and-complexity claim directly, and the further suggestion that this is therefore the segment of the housing stock where financing constraints and Band C threshold status are most likely to bind is an untested economic hypothesis rather than an empirical finding of this paper. As the EPC register does not report height or storey count directly, floor level is used as the only field permitting a high-floor criterion at scale; total building height cannot be independently verified, a limitation discussed further in Section 6.3.

Constructing a Conservatively Filtered Same-Unit Panel

A central methodological consideration in this study concerns how repeat inspections of the same dwelling are linked into a panel. Linking successive EPC records by UPRN alone is not reliable throughout the register: UPRNs are occasionally recorded at the building rather than the unit level. In one illustrative case, a single UPRN links 300 separate inspection records lodged on the same calendar day, corresponding to 300 physically distinct flats of differing floor area within one building rather than repeated inspections of a single dwelling.

This is addressed by requiring, for every pair of consecutive same-UPRN inspections, that (i) total floor area not differ by more than 5 m^2 and (ii) the normalized property address match exactly (address fields concatenated, lower-cased, and stripped of punctuation and repeated whitespace). Both thresholds are deliberately conservative rather than optimized for confirmation rate: the 5 m^2 floor-area tolerance was chosen to accommodate ordinary measurement variation between assessors rather than to maximize the number of pairs matched; alternative tolerances and an approximate-matching alternative to the exact-address requirement were subsequently tested and are reported in Section 6.3. The filter is a conservative operational rule designed to reduce obvious cross-unit linkage contamination, not a validated record-linkage classifier. The exact-match address requirement is similarly conservative. It will fail to retain some genuinely matching pairs affected by minor transcription variation (e.g., abbreviation or spacing differences) that a fuzzy-matching procedure would accept, so bias from this criterion is expected to run predominantly toward false negatives—discarding valid same-unit pairs—rather than toward false positives that would contaminate the panel with cross-unit comparisons. This is a deliberate design choice, since false positives are the error this procedure is designed to rule out. It is not a guarantee, however. A building with duplicated or missing unit-level address strings and units of similar floor area could in principle still pass both criteria despite being physically distinct. The construction-age-band field was considered as a third confirmation criterion but rejected: the rate at which this field disagrees between consecutive inspections of

the same UPRN rises smoothly from 16% (2008) to 58% (2020 onward), consistent with successive revisions to the RdSAP methodology's age-band categories rather than genuine differences in dwelling identity.

Within the high-floor, 2013–2023 population, this procedure identifies 21,680 candidate consecutive-inspection pairs, of which 15,277 (70.5%) satisfy the conservative same-unit confirmation criteria and are retained in the filtered panel, covering 12,161 distinct properties. An earlier diagnostic applying the same two criteria to all consecutive-inspection pairs in the unrestricted flat/maisonette register (not limited to high-floor dwellings) found a similar confirmation rate of 60.5–60.8%, cross-checked by independently reproducing the procedure across two computational pipelines (R/data.table and SQL/DuckDB), which gives confidence that the filtering rule itself, and not a toolchain artefact, drives the result; this computational cross-check confirms implementation consistency only and is not a claim of validation against ground truth. The confirmation rate is not stable across the sample period (Table 4): it exceeds 79% in every year from 2013 to 2017, falls sharply from 2018 onward, and reaches a low of 52.5% in 2022 before a modest recovery to 56.9% in 2023. The single largest year-on-year drop (13.4 percentage points) occurs between 2017 and 2018, broadly consistent with the RdSAP methodology revisions occurring within this period rather than a change in the underlying quality of address or floor-area recording, though the exact revision date cannot be pinned to a single year from the confirmation-rate series alone. This means the confirmed panel's later years are drawn from a substantially more selected subset of candidate pairs than its earlier years; this compositional shift over time is noted as a limitation in Section 6.3.

Table 4. Same-unit confirmation rate by year, 2013–2023. Accepted-pair totals sum to 15,281, consistent with the verification-session N documented in Table 2's caption (4 more than the 15,277 production-run figure reported in Section Constructing a Conservatively Filtered Same-Unit Panel and Table 3); the year-by-year pattern and overall rate (70.5%) are unaffected by this known, minor discrepancy.

Year	Candidate Pairs	Accepted Pairs	Confirmation Rate
2013	1738	1635	94.1%
2014	1955	1705	87.2%
2015	1725	1372	79.5%
2016	1288	1114	86.5%
2017	1832	1615	88.2%
2018	2029	1518	74.8%
2019	1344	896	66.7%
2020	1270	730	57.5%
2021	2249	1242	55.2%
2022	2340	1228	52.5%
2023	3910	2226	56.9%

This construction also implies a form of selection into the panel: only the roughly 15% of high-floor flats that receive a second EPC inspection within the sample window are usable, since the timing correction in Section Timing of the Band C Threshold Status Measure requires two inspections of the same unit. A sensitivity check compares high-floor properties that were reinspected against those inspected only once (using an independently reconstructed approximation of the same high-floor filter, since the exact production pipeline could not be re-run for this check). Reinspected properties have a substantially lower initial energy-efficiency score on average than single-inspection properties (65.9 versus 71.6 on the underlying scale; $t = -50.6$, $p < 0.001$), and a markedly lower owner-occupied share (1.6% versus 8.1%; $\chi^2 = 908.1$, $p < 0.001$). This mean (65.9) is computed on the reconstructed approximation described above and describes only the reinspected-

versus-single-inspection comparison; it is not directly comparable to the 68.6% Band C share reported for the full filtered panel in Table 3, since the two statistics are drawn from different samples for different purposes rather than describing the same population at the same point. This is consistent with reinspection being concentrated in rented-sector stock undergoing sale, MEESs-driven compliance work, or refinancing, and in worse-performing units generally. The filtered panel used throughout this paper should therefore be read as representative of reinspected, disproportionately rented, high-floor flats specifically, not of the high-floor stock as a whole. This narrower scope is a limitation and is revisited in Section 6.3.

The resulting panel differs from a UPRN-only panel in composition, not in scale: mean recorded EPC improvement is 58.3% and mean recorded EPC-score change is 3.16 EPC points in the filtered panel, versus 64.9% and 3.08 without same-unit confirmation. This similarity is not evidence that the unfiltered panel was adequate; roughly three in ten candidate pairs fail matching, and the unmatched dependent variable mixes genuine retrofit events with cross-unit comparisons that would bias any estimate unpredictably.

4.2. Variable Construction

Recorded EPC Improvement is a binary indicator equal to one where CURRENT_ENERGY_EFFICIENCY increases between matched consecutive inspections of the same unit; recorded EPC-Score change is the corresponding change in EPC points.

EPC Band C threshold status is defined as an indicator equal to one where the dwelling's earlier matched (baseline) EPC score is at least 69. Band C is treated throughout as a policy-relevant benchmark, not as confirmed eligibility for an individual green mortgage; this distinction requires clarification given a discrepancy in the underlying evidence. Individual-mortgage green product eligibility at two major UK high-street lenders, NatWest and Barclays, is documented in their own current published product terms to require Band A or B (SAP $\geq$ 81), not Band C; the academic sources cited above [6,7] describe green mortgages in general terms and do not themselves specify a Band C individual-loan threshold. Band C is nonetheless retained as the primary threshold for the discontinuity design for reasons distinct from individual-loan eligibility. First, it carries independent policy relevance. NatWest has adopted a portfolio-level ambition for 50% of its mortgage book to be at or above EPC Band C by 2030. The UK Government has subsequently confirmed a single compliance date of 1 October 2030 for the higher minimum energy-efficiency standard in the private rented sector, subject to the applicable EPC metrics and exemptions [58]. Band C therefore remains a meaningful policy benchmark, even though it is not, on the evidence reviewed here, an individual-loan qualifying criterion. Second, and decisive for the RDD design specifically, Band C offers denser support on both sides of the cutoff than Band B: density above score 84 falls sharply, limiting usable bandwidths for a local-randomization design centred on Band B. The stricter Band B threshold (score $\geq$ 81), which current lender documentation indicates is the individual-mortgage-level criterion actually applied by NatWest and Barclays, is examined as a parallel timing-sensitivity specification in Table 2, and is also reported descriptively in Table 3, not treated only as a secondary robustness check; the correlational estimates in Table 2 are, if anything, larger in magnitude at Band B than at Band C, so the paper's substantive conclusions do not depend on which threshold is treated as primary. A Band D threshold was not examined, since it was not examined as a mortgage-product or lender-portfolio threshold, though it is used as a policy-eligibility criterion by programmes such as the Energy Company Obligation and the Warm Homes: Social Housing Fund (Section 2.3); extending the sensitivity analysis to Band D is noted as a direction for future work (Section 6.3).

A further scope limitation follows directly from the discrepancy documented above. This study observes an EPC score and its position relative to the Band C threshold; it does not observe whether a dwelling's owner applied for, was offered, or received any green mortgage product, nor the interest rate actually charged. The regression discontinuity design therefore identifies a discontinuity in subsequent EPC-recorded score-change outcomes around a policy-relevant classification threshold, not a verified first-stage jump in the probability of receiving green financing at that threshold; no lender-level loan-origination or application data are available to test whether such a first-stage discontinuity exists in this sample and period. Throughout the remainder of the paper, "Band C threshold status" denotes this EPC-based classification, consistent with its Section 4.2 definition, rather than a confirmed financing outcome, and readers should interpret the RDD estimates accordingly.

Timing of the Band C Threshold Status Measure

Whether Band C threshold status is measured at a dwelling's baseline or follow-up score is not a minor detail: it determines whether the variable can be interpreted as a predictor of subsequent recorded EPC improvement at all. If measured from the follow-up score, a dwelling can only newly satisfy the threshold by having already improved enough to cross it, making recorded improvement a near-necessary condition rather than a pre-existing condition encouraging adoption. The literature reviewed for this study did not identify sufficient documentation to determine how comparable indicators elsewhere were timed.

This was tested directly. Table 2 reports the EPC threshold-status coefficient, with recorded EPC improvement as the dependent variable, from the simplest possible specification (postcode and year fixed effects only, no interest-rate term) under all eight combinations of panel construction (same-unit filtered versus unfiltered), threshold (Band C versus Band B), and timing (baseline versus follow-up score). Standard errors are clustered by postcode throughout; the filtered panel spans 3206 distinct postcodes, well above the range at which few-cluster inference concerns typically arise, so no wild-cluster-bootstrap correction is applied.

Every combination using the temporally prior baseline construction is negative and highly significant, regardless of construction or threshold. Every follow-up combination is positive. Of the eight cells reported, follow-up/filtered/Band B (+0.281) is numerically closest to the naive baseline coefficient (+0.311, Section 3.3), a difference of 0.030; the corresponding unfiltered follow-up cells (+0.250 for Band B, +0.172 for Band C) are the next closest, followed by the filtered follow-up/Band C cell (+0.244). The mechanism was verified directly: among dwellings with no recorded improvement, exactly zero newly cross the Band B threshold, a logical necessity of the follow-up construction. Among dwellings that do improve, 9.8% cross it. This discrete jump is sufficient to generate a spurious positive association without any economic mechanism linking green finance to retrofit behaviour. This 9.8% figure and Table 3's 9.8% share of the panel at or above Band B at baseline are two conceptually distinct statistics (crossing rate among improving dwellings versus baseline threshold share) that happen to coincide in their rounded value; this is confirmed to be a numerical coincidence rather than a duplicated or mislabelled statistic.

Whether similarly constructed eligibility indicators elsewhere share this undocumented timing convention is not established here (Section 6.3). This finding is presented not as an accusation of any particular study but as the most plausible general explanation, consistent with all evidence in this paper, for why a naive specification yields a positive H1 estimate: a specification choice that makes the eligibility indicator partially definitional of the outcome it explains. All results reported from Section 4.3 onward use the baseline score exclusively.

Potential headroom is defined as POTENTIAL_ENERGY_EFFICIENCY minus the current (baseline) score, and % gap closed as recorded EPC-score change divided by headroom, for the 99.2% of the filtered panel with positive headroom. The remaining 0.8% record a potential score below the current score (again attributable to methodology revisions across inspection dates, not a genuine ceiling) and are excluded from this outcome measure only. This outcome's denominator (headroom) shrinks systematically as the running variable increases toward the ceiling. Because headroom declines as the baseline score rises, otherwise comparable positive EPC-score changes mechanically translate into larger, not smaller, gap-closed ratios above the cutoff. This scaling feature therefore works against, rather than mechanically generating, the observed negative discontinuity: finding a negative effect despite this mechanical pull toward larger positive ratios above the threshold is, if anything, a conservative result. Mean and SD of % gap closed among dwellings with positive headroom are 0.52 and 3.20, respectively; the wide SD reflects the small-denominator dwellings discussed next. Within the ±5 window specifically (N = 4977, matching the sample used in Table 5), mean headroom is 5.7 SAP points; it differs sharply either side of the cutoff, averaging 7.3 points below the threshold versus 3.9 points at or above it, directly illustrating the mechanical scaling difference discussed above rather than inferring it from the ratio of point estimates. Dwellings with very small headroom generate disproportionately extreme ratios: 26% of the positive-headroom subsample has 2 or fewer points of headroom, and this subgroup's % gap closed ranges from −59 to +53, versus −14 to +15 for dwellings with more than 2 points of headroom. Re-estimating the ±5 window with the outcome winsorized at the 1st/99th percentiles yields an attenuated, non-significant estimate (−0.085, $p = 0.166$), while restricting to dwellings with more than 5 points of headroom yields a similar-magnitude but more significant estimate (−0.181, $p < 0.001$) than the full-sample estimate reported above. This divergence suggests the small-headroom dwellings contribute noise that winsorizing does not fully address, since capping extreme values leaves the underlying small-denominator observations in the sample; restricting the sample directly removes this noise source and, if anything, strengthens the result. Given this, the headroom-restricted estimate is treated as the more informative robustness check, and the full-sample estimate in Tables 5 and 6 should be read as a conservative baseline rather than an upper bound on the effect.

Table 5. RDD estimates, Band C threshold, window ±5.

Outcome	Estimate	95% CI	N	*p*-Value
Recorded EPC improvement (raw)	−0.058	[−0.082, −0.034]	5797	<0.001
Recorded EPC-score change (SAP points; raw)	−1.923	[−2.283, −1.563]	5797	<0.001
% of headroom closed (ceiling-adjusted)	−0.189	[−0.339, −0.039]	4977	0.014

N for the raw outcomes reflects observations with non-missing score within the ±5 window; N for the ceiling-adjusted outcome is lower because rdrandinf's internal sample construction additionally excludes observations with missing headroom, which is undefined for a small number of dwellings (Section Timing of the Band C Threshold Status Measure).

Table 6. Bandwidth sensitivity, ceiling-adjusted outcome, with 95% confidence intervals and effective sample sizes, for the four windows (±2 to ±5) satisfying the two-covariate balance criterion established in Section 4.4.1. Balance fails from ±6 onward under that criterion ($p < 0.05$), and estimates at those wider windows are not reported here. Under an extended balance test additionally including owner–occupier status, private-rented status, and construction-age band, ±2 ($p = 0.024$) and ±1 ($p = 0.015$) fail balance, while ±3 to ±5 continue to pass; the ±2 estimate is retained here for transparency but should be read with this additional caveat. Unique properties and postcode clusters for each window are reported in the table below.

Bandwidth (±SAP)	Estimate	*p*-Value	95% CI	N (Left, Right)	Unique Properties	Postcode Clusters
2	−0.296	0.001	[−0.45, −0.15]	753, 1381	2038	897
3	−0.388	<0.001	[−0.55, −0.20]	1050, 1855	2732	1115
4	−0.176	0.049	[−0.35, −0.05]	1331, 2769	3790	1326
5	−0.189	0.014	[−0.30, −0.05]	1639, 3338	4563	1526

4.3. Baseline (Correlational) Specification

As a baseline for comparison, a specification following common practice in the literature is estimated first, as an explicit point of reference rather than a credible estimate of any causal quantity:

$$RecordedEPCImprovement_{it} = \beta_0 + \beta_1 \ ThresholdStatus_{it} + \delta_p + \gamma_t + \varepsilon_{it} \quad (1)$$

Here δ_p denotes postcode fixed effects, and γ_t denotes year fixed effects, with standard errors clustered by postcode; this is the specification reported in Table 2. It cannot support a causal interpretation of β_1, since Band C threshold status is, by construction, related to a building's baseline energy performance. As shown in Section Timing of the Band C Threshold Status Measure, β_1's sign is also highly sensitive to whether eligibility is measured from the baseline or follow-up score; Equation (1) is estimated throughout using the baseline measure and reported in Table 2 purely as a baseline, not a credible estimate in its own right.

A second, exploratory specification incorporates the policy rate and its interaction with threshold status, to investigate H2 and H3 (Section 5.4):

$$RecordedEPCImprovement_{it} = \beta_0 + \beta_1 \ PolicyRate_t + \beta_2 \ ThresholdStatus_{it} + \beta_3 \ (PolicyRate_t \times ThresholdStatus_{it}) + \delta_p + \gamma_t + \varepsilon_{it} \quad (2)$$

Equation (2) cannot identify β_1 separately, since $PolicyRate_t$ varies only by year and is collinear with the year fixed effect γ_t; β_3 retains cross-sectional variation from threshold status itself and can be estimated statistically, though its causal interpretation remains problematic for the reasons discussed in Section 5.4. No table in this paper reports Equation (2) as its primary specification; the correlational estimates it produces are discussed narratively in Section 5.4.

4.4. Identification Strategies

4.4.1. Regression Discontinuity Design for Band C Threshold Status (H1)

To address the endogeneity of Band C threshold status, an RDD is implemented around the Band C threshold (score = 69), using the baseline score as the running variable. The treatment indicator here, whether score ≥ 69, is definitional: it is constructed directly from the running variable, and no data on actual loan applications, offers, or receipt are available (Section 4.2). In the absence of loan-level evidence that crossing Band C changes access to green finance, the RDD estimates a discontinuity in subsequent recorded EPC improvement at the Band C threshold. It should not be interpreted as the causal effect of

green-finance access. Standard manipulation testing uses local polynomial density estimation with the package's mass-point adjustment for repeated running-variable values [59], rejecting the null of no manipulation (T = 18.58, $p < 0.001$) even after this adjustment. The persistence of the density discontinuity after this adjustment indicates that exact duplication alone does not account for the result; it does not eliminate broader concerns arising from discreteness, rounding, or measurement error, and does not by itself establish that assignment near the cutoff is random. Placebo tests at four non-policy cutoffs are also strongly significant, in inconsistent directions, and donut-hole exclusion increases the test statistic, both consistent with global heaping in RdSAP scores rather than manipulation specific to this threshold. Continuity-based density testing beyond the mass-point adjustment is judged uninformative here.

The local randomization approach to regression discontinuity is therefore adopted [60,61], which selects an estimation window based on covariate balance rather than continuity of the running variable's density; the latter reference covers the rdwinselect window-selection procedure and its local-randomization implementation directly. Using pre-determined covariates (tenure and prior floor area), windows of ± 2 to ± 5 score points around the cutoff pass covariate balance tests (p = 0.505, 0.551, 0.178 and 0.097, respectively), with balance failing from ± 6 onward (p = 0.016). An extended balance test additionally including owner–occupier status, private-rented status, and construction-age band shows a less clean pattern: ± 3 to ± 5 continue to pass (p = 0.560, 0.183, 0.089), but ± 1 and ± 2 fail (p = 0.015 and 0.024); ± 6 and ± 7 also fail under this extended test (p = 0.014 and 0.005), consistent with their failure of the narrower two-covariate test reported above, indicating that balance in this design is sensitive to which covariates are tested, a limitation discussed further in Section 6.3. Across the four reported window widths, three (± 3 to ± 5) satisfy both the primary and extended balance diagnostics, whereas ± 2 is retained only as a sensitivity specification because it fails the extended balance test. The ± 5 window is adopted as the working specification; sensitivity to this choice, across the full range of windows satisfying the primary two-covariate balance criterion (± 2 to ± 5), is reported in Section 5.3 and Table 6.

The choice between continuity-based and local-randomization RDD is not merely a robustness preference here but a response to a specific feature of the running variable. Continuity-based estimators (e.g., rdrobust) rely primarily on the identifying assumption that the conditional expectation of potential outcomes varies smoothly through the cutoff; this assumption is not directly testable, since potential outcomes are only ever observed on one side of the threshold for any given unit. The density of the running variable is a separate, testable quantity: a discontinuity in this density is a diagnostic that may be consistent with sorting or manipulation around the cutoff, and its absence is often treated as indirectly supportive of the potential-outcome-continuity assumption, but a smooth density does not prove that assumption, and a discontinuous density does not by itself disprove it, since density discontinuities can arise from causes other than outcome-relevant sorting. The standard density-based manipulation test assesses only this density diagnostic: whether the distribution of the running variable exhibits a discontinuity at the cutoff. Section 4.4.1 above shows that this density diagnostic is uninformative here for a reason unrelated to any manipulation of eligibility: RdSAP-generated EPC scores are heavily rounded by assessors, producing discrete mass points (heaping) that generate spurious density discontinuities at every threshold tested, policy-relevant or not. A continuity-based estimator applied to a heaped running variable can therefore reject smoothness, or produce an unstable point estimate, purely as an artefact of assessor rounding, independent of whether the eligibility threshold itself induces any behavioural response. Local randomization sidesteps this specific problem: rather than modelling the density of the running

variable, it treats observations within a data-driven window around the cutoff as if from a small, randomized experiment. The intuition is that immediately around the threshold, small differences in EPC score are plausibly driven by assessor rounding and other sources of measurement noise rather than by any systematic difference between dwellings, so which side of the threshold a given dwelling happens to fall on resembles a coin flip within a narrow enough window; if this holds, permutation-based inference within that window can substitute for the smooth-density assumption that continuity-based methods require. This local-randomization assumption is not simply assumed: its plausibility is assessed empirically using covariate-balance diagnostics on variables plausibly unrelated to the eligibility threshold (tenure, prior floor area) rather than on the running variable's density itself. This approach avoids directly modelling a smooth running-variable density, making it more suitable for the discrete, heaped score used here, at the cost of external validity being confined to the estimation window rather than the full support of the running variable; its local-randomization assumption nonetheless still requires substantive and diagnostic justification, rather than following automatically from the choice of estimator, which is why Table 6 reports sensitivity across multiple window widths rather than relying on a single specification. A matching-based alternative (propensity-score matching, entropy balancing, or inverse-probability weighting) was not pursued as the primary strategy because these methods condition on observed covariates to approximate exogenous assignment away from the threshold, and so cannot address the same threshold-proximate endogeneity concern that motivates using an RDD in the first place; they would, in addition, still require deciding how to treat the heaping problem in the running variable used to define the matching sample. The continuity-based estimate is nonetheless reported in Section 5.3 for completeness, alongside an explicit caveat about its reduced credibility in this setting.

4.4.2. Attempted Identification of the Interest-Rate Effect (H2, H3)

Equation (2)'s collinearity between the policy rate and year fixed effects cannot be resolved by within-sample re-specification alone, because the policy rate has no cross-sectional variation. Two independent identification strategies were attempted for the interest-rate channel; both are reported as unsuccessful, together with the diagnostic evidence supporting that conclusion, in the interest of transparency, not as supporting results.

The first strategy exploits the phased introduction of the Minimum Energy Efficiency Standards (MEESs) from April 2018 as a source of policy-induced retrofit variation. A triple-difference specification shows no discrete break at the policy date: an event-study version of the interaction is already elevated in 2013, five years before MEESs existed, and the estimated coefficient changes its sign across reasonable specifications. Private-rented, F/G-rated dwellings were already on a differential retrofit trajectory before MEESs took effect, so this design cannot separate a MEESs-specific effect from a pre-existing trend. Full diagnostics are provided in Supplementary File S1, Part A.

The second strategy constructs a predetermined exposure-by-rate interaction, combining the national policy rate with a pre-determined, area-level measure of interest-rate exposure (local-authority private-rented tenure share, 2008–2012). This interaction varies in both time and cross-section, but within-year variation is small before 2022 and driven almost entirely by 2023, a year that also coincided with the European energy price crisis. A leave-one-year-out analysis confirms the estimate is not robust to excluding 2023 alone, so an interest-rate channel cannot be separated from an energy-price response. Full diagnostics are provided in Supplementary File S1, Part B.

The collinearity between the national policy rate and year fixed effects, and the consequent inability to identify H2 specifically, together with the residual causal-interpretation

problems affecting H3 despite its statistical estimability, is reported as an acknowledged limitation of this study (Section 6.3).

5. Empirical Results

5.1. Descriptive Statistics

Table 3 reports descriptive statistics for the reconstructed, same-unit-filtered high-floor panel (15,277 observations, 12,161 unique properties, 2013–2023). The mean recorded EPC improvement is 58.3% and the mean recorded EPC-score change is 3.16 EPC points. Under the Band C threshold (score $\geq$ 69), 68.6% of high-floor observations are at or above the threshold; under the stricter Band B threshold (score $\geq$ 81), the corresponding figure is 9.8% (Figure 2).

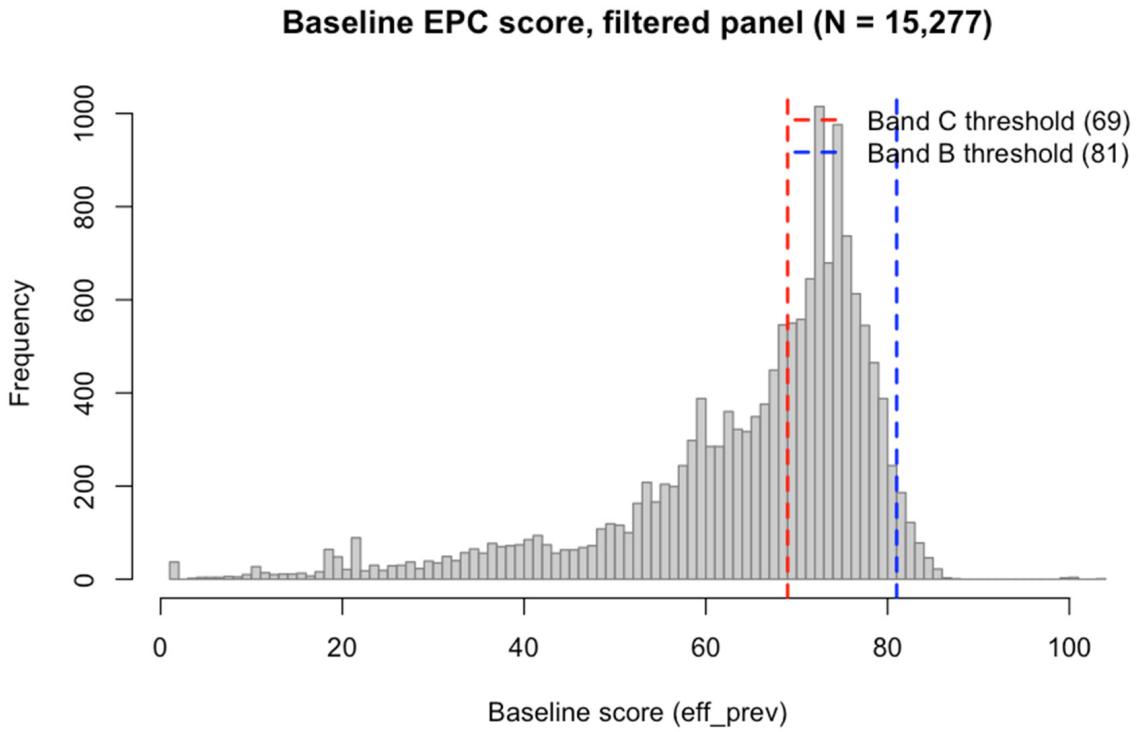

Figure 2. Distribution of the baseline EPC score, filtered panel (N = 15,277). Dashed lines mark the Band C (69) and Band B (81) classification thresholds. The distribution is left-skewed, with a pronounced mode just above the Band C threshold coinciding with the excess mass at score 73 discussed in Section 4.4.1.

5.2. Baseline Correlational Estimates

The timing-sensitivity specification described in Section Timing of the Band C Threshold Status Measure is exactly Equation (1) as defined in Section 4.3 (postcode and year fixed effects, no interest-rate term), estimated using the corrected, baseline Band C threshold-status measure across the timing/panel/threshold variations that are the paper's main focus, and is reported in Table 2. The separate, exploratory Equation (2), which additionally includes the policy rate and its interaction with Band C threshold status, is examined narratively in Section 5.4, not in a dedicated table, since Section 4.4.2 shows that its policy-rate term cannot be separately identified from the year fixed effects. Band C threshold status now carries a large, negative, highly significant coefficient, the central empirical finding of this paper: the naive baseline estimate (+0.311, Section 3.3) is closely reproduced (+0.244 to +0.281) only from the follow-up score, which makes adoption near-necessary for newly satisfying the threshold. Measured from the baseline instead, the coefficient is negative under every combination tested. By construction, a dwelling with no recorded improvement can

never newly cross an eligibility threshold under the post-inspection definition; the original association is thus consistent with a mechanical, near-circular relationship rather than a genuine financing effect. These results are consistent with H1 (Section 3.3.1): correcting the timing of Band C threshold status does not merely weaken the original positive association but reverses its sign under every specification tested. Section 5.3 provides a related robustness check using a different estimator, drawing on the same underlying panel and score variable; Section 5.4 reports unsuccessful attempts to identify the interest-rate channel and is not treated as additional evidence for H1.

5.3. Regression Discontinuity Estimates for H1

This section reports the regression discontinuity estimates for H1, using a different estimation strategy from the timing-correction test in Section 5.2; the two tests share the same underlying panel and baseline score variable (Section Timing of the Band C Threshold Status Measure), so the comparison below should be read as a robustness check across specifications, not as independent replication.

The raw outcomes are affected by a mechanical ceiling effect: buildings just above the cutoff have less scope left to improve (maximum attainable score 92–104 within Band A) independent of any financing channel. Adjusting for this using the headroom-based outcome (% of the remaining energy-efficiency gap closed) reduces the magnitude of the estimate, but the sign remains negative, and the estimate remains significant at the 5% level (-0.189, $p = 0.014$). A further caveat applies to all outcomes reported in this section: because both the running variable and the outcome are derived from the same EPC scoring system, and the potential (headroom-defining) score is itself generated by the same underlying model and certificate data, the estimates remain susceptible to mathematical coupling, regression to the mean, differential measurement error, and assessor or methodology changes between inspections, in addition to the ceiling effect corrected for above; dividing by potential headroom does not remove these shared-source risks. The RDD is accordingly best understood as testing for a discontinuity in the subsequent recorded EPC change at the threshold, which need not correspond one-to-one with the physical retrofit investment. All randomization-inference estimates in this section use 5000 permutation replications with a fixed random seed for reproducibility.

This estimate is not fully stable to bandwidth choice. Table 6 reports the ceiling-adjusted estimate across the windows (± 2 to ± 5) that satisfy the primary two-covariate balance criterion established in Section 4.4.1; wider windows (± 6 to ± 8) fail this balance test and are therefore excluded from the primary evidence for H1.

The sign is negative and statistically significant at the 5% level at every window reported in Table 6 (± 2 to ± 5 EPC score points), though the magnitude is not monotonic across this range: significance is weakest at the ± 4 window ($p = 0.049$) and strongest at ± 3 ($p < 0.001$), with ± 2 ($p = 0.001$) and ± 5 ($p = 0.014$) falling in between. With only three primary valid window widths (± 3 to ± 5), plus ± 2 as an additional sensitivity specification, no confidence-interval or Monte Carlo variability evidence is offered here to adjudicate between candidate explanations. One specific candidate, assessor rounding toward the round-number scores 65 and 70 (which enter the ± 4 and ± 5 windows but not ± 2 and ± 3), was tested directly: the running variable shows no excess mass at either value relative to neighbouring integer scores (ratios of 0.91 and 0.99), and a donut specification excluding both values leaves the same non-monotonic pattern virtually unchanged (± 2: -0.257; ± 3: -0.387; ± 4: -0.163; ± 5: -0.180, versus -0.296, -0.388, -0.176, -0.189 in the full sample). This rules out heaping at these specific values as the explanation. The full integer-by-integer distribution (Table 7) does, however, show pronounced excess mass at two other values within the wider windows: scores of 73 and 75 exceed their immediate neighbours by 42%

and 28%, respectively (counts of 1014 and 970 against neighbouring averages of 712 and 760, across scores 60–78). Score 73 falls at the outer edge of the ±4 window (65–73) and as an interior point of the ±5 window (64–74); score 75 falls outside all windows reported in Table 6. Because score 73 first enters the sample when the window expands from ±3 to ±4 and remains included at ±5, this mass point is a plausible candidate for the marked attenuation between the ±3 and ±4 estimates (−0.388 to −0.176), although it cannot by itself explain the comparatively small further change between ±4 and ±5; testing it with a further donut exclusion is left for future work given the limited number of remaining valid windows. No single account is therefore asserted as established. The variation across windows may reflect sampling variation, changing sample composition as the window widens, heterogeneous local effects, or measurement characteristics of the discrete running variable; it should not be read as evidence of a monotonic dose-response relationship. A genuinely monotonic pattern would not be expected under local randomization in any case, since each window is treated as a separate randomized experiment, not as points on a continuous bandwidth-selection curve. Windows of ±6 and beyond fail covariate balance (Section 4.4.1) and are not treated as valid local-randomization estimates; for transparency, the ceiling-adjusted point estimates obtained at those wider windows were also negative and nominally significant, but because the underlying randomization assumption does not hold there, they are not reported as evidence for H1 and should not be interpreted as confirmatory. A continuity-based cross-check (rdrobust, default MSE-optimal bandwidth h = 3.26, triangular kernel) on the same ceiling-adjusted outcome yields a positive, non-significant point estimate (+0.131, 95% CI [−0.302, 0.821], p = 0.365, effective N = 1050 left/1855 right), opposite in direction, though this check is itself compromised by the heaping problem documented in Section 4.4.1 and is reported for completeness, not as a decisive test; the full detail of this check is discussed in Section 6.1.

Table 7. Running variable (baseline score) frequency, integer scores 60–78, filtered panel. Referenced in Section 4.4.1's heaping diagnostic.

Score	N	Score	N
60	386	70	545
61	290	71	554
62	283	72	644
63	365	73	1014
64	319	74	680
65	320	75	970
66	346	76	734
67	370	77	610
68	450	78	540
69	555		

The combined evidence supports three conclusions. One is that the original, naive positive estimate motivating this investigation is explained largely by a measurement-timing artefact (Table 2) rather than a genuine effect. Another is that this is not the only route to a negative estimate: RDD around the eligibility threshold reaches the same sign under a different specification, addressing the remaining endogeneity concern, though it draws on the same underlying score variable as the timing-correction test, not a separate data source. Finally, no confirmed causal effect in the reversed direction is claimed, since the RDD magnitude is bandwidth-sensitive (Table 6) and a continuity-based check points the other way. Taken together, the naive hypothesis of a positive causal effect is not supported, while this paper's H1 (Section 3.3.1), predicting that the association would reverse in sign

under baseline timing, is supported; the magnitude of the reversal should nonetheless be read as suggestive rather than conclusive.

5.4. H2 and H3: Unresolved

As detailed in Section 4.4.2, neither strategy produces a reliable test of H2 or H3. A postcode-and-year fixed-effects model incorporating the interest-rate exposure measure alongside Band C threshold status shows a negative coefficient on Band C threshold status across every sample tested (coefficient range: -0.277 to -0.247, all $p < 0.001$; dependent variable: recorded EPC improvement), consistent in direction with the RDD finding in Section 5.3. This coefficient is not treated as independent corroborating evidence, however: the same specification's interest-rate term is confounded by the 2023 energy-price shock and is not robust to a leave-one-year-out check (Section 4.4.2), and there is no principled basis for assuming the Band C threshold-status coefficient estimated in that same regression is immune to whatever specification problem contaminates the interest-rate term. The pattern is reported for transparency and because it is at least consistent with, rather than contradicting, the sign found elsewhere in the paper, not as an additional confirmatory test of H1.

6. Discussion and Policy Implications

6.1. Interpretation of Findings

This investigation set out to estimate how Band C threshold status and monetary policy conditions jointly shape subsequent recorded EPC-score changes in high-floor residential buildings, following the framing common to the existing literature [6,14,23]. The empirical work reported here answers a different, arguably more consequential, question: whether the eligibility metric used to test that framing is itself reliably measured. It is not, at least not without close attention to timing.

Two lines of evidence converge on H1, though they are not fully independent. One is definitional: eligibility built from a follow-up score is close to tautologically related to the outcome it explains: the simplified follow-up-score, filtered-panel Band B specification in Table 2, reproduces the original coefficient to within 0.03 (+0.281 versus +0.311); the original specification (Equation (2)) produced +0.311; re-estimating the corresponding follow-up/filtered/Band B specification with the policy-rate interaction term on the current analysis data reproduces this almost exactly (+0.312). The other comes from an RDD immune to this timing problem. It uses covariate balance instead of continuity-based testing, because the running variable's heaping makes the density-test result hard to interpret as evidence of threshold-specific manipulation. Both tests draw on the same underlying filtered panel and the same baseline score variable, and the correlation between the score as used in each test exceeds 0.99. The convergence therefore reflects two specifications applied to related evidence, not two independent replications. It still rules out the possibility that the reversal is an artefact of one particular estimation strategy.

A related question is whether this shared score variable creates a common bias, not merely a common sample. If assessor rounding (heaping) systematically pushed scores toward a specific value near the Band C cutoff, it could inflate eligibility counts in Table 2 and distort the running variable near the cutoff in the RDD in the same direction at once—a more serious concern than sample overlap alone. The placebo-cutoff evidence in Section 4.4.1 speaks to this directly: manipulation-style tests at four non-policy cutoffs are also strongly significant, but in inconsistent directions across cutoffs. This indicates that heaping is a general feature of the RdSAP score distribution, not a pattern concentrated at the Band C threshold specifically. A threshold-specific, common-direction bias therefore looks less likely. A general classification-error channel is not ruled out, however: any binary threshold

measure derived from a heaped score remains vulnerable to rounding-induced misclassification, whether or not the heaping clusters at one particular value. A bootstrap co-variation check quantifies this directly. Resampling the underlying properties (with replacement) 5000 times, and re-estimating both the Table 2 coefficient and the RDD estimate from each resampled set of properties, yields a small but statistically distinguishable-from-zero positive correlation between the two statistics (r = 0.152, 95% CI [0.125, 0.179]). This indicates some shared sensitivity to sampling variation in the underlying score variable, but falls far short of the near-perfect co-movement that would be expected if the two tests were merely re-deriving the same result. This residual shared-noise channel is noted as a limitation in Section 6.3.

Existing retrofit-finance reviews discuss financing conditions primarily at the instrument or policy level [23], while Ma and Jiang [14] measure green finance as a regional financial-development construct rather than as a property-level EPC threshold indicator. The negative sign surviving correction is not fully bandwidth-stable (Table 6) and is contradicted in direction by a continuity-based check: rdrobust with its default MSE-optimal bandwidth (h = 3.26, triangular kernel) yields a positive, non-significant estimate (+0.131, p = 0.365, effective N = 1050 left/1855 right), the opposite sign from every local-randomization window reported in Table 6. This continuity-based estimate should be treated with caution, not as a genuine contradiction of the local-randomization result: the running variable's heaping triggers an explicit mass-points warning from the estimator, and the automatically selected bandwidth (3.26) does not correspond to any window tested under local randomization, so the two approaches are not estimating the same quantity on the same effective sample. A reversed causal claim is not proposed; the original positive claim was not supported in this sample, and a verifiable mechanism helps explain why.

H2 and H3 remain open questions. Both attempted identification strategies failed for reasons specific to the UK interest-rate environment during the sample period: near-zero rates for most of 2013–2021, a sharp increase concentrated in 2022–2023, and a policy shock (MEESs) whose treatment group was already on a differential trend before the policy existed, rather than for reasons specific to green finance or retrofit investment. This is best viewed as an unresolved data availability problem, not evidence against a moderating role for monetary policy conditions.

6.2. Policy Implications

For policymakers and lenders operating within the England and Wales EPC system specifically, the central implication is that an eligibility flag built from post-improvement performance data cannot demonstrate that a financial instrument caused the improvement, regardless of how favourable the correlation appears. Programme evaluations using before/after performance data, in this or comparable EPC-based regulatory contexts, should specify, and report sensitivity to, the timing convention used to determine eligibility.

For regulators and standard-setters overseeing green taxonomies and sustainable finance disclosure, this finding is related to broader concerns raised in the ESG measurement literature about the design, standardization, interpretation, and comparability of sustainability classifications and metrics [62,63].

For the specific case of UK residential retrofit policy, the inability to identify a credible interest-rate channel does not imply that borrowing costs are irrelevant to retrofit decisions; it implies that the 2013–2023 period, dominated by a decade of near-zero rates followed by a rapid tightening cycle confounded with an energy price shock, does not offer the variation needed to isolate this channel from the data sources available in the EPC register. Policymakers seeking to understand this channel would need financing-cost data with more

independent variation than a national policy rate observed over an eleven-year window can provide.

6.3. Limitations and Directions for Future Research

Several limitations are worth noting. The first concerns the timing convention used in similarly constructed eligibility indicators elsewhere. This often cannot be confirmed, since it goes undocumented; the Section Timing of the Band C Threshold Status Measure gives the most plausible account for this study's own specification. A second limitation is that window selection is sensitive to which covariates are tested. The two-covariate balance test used to select ±2 to ±5 is not the only reasonable specification. An extended test that adds owner–occupier status, private-rented status, and construction-age band fails at ±2 (Section 4.4.1). The RDD estimate is accordingly bandwidth-sensitive: Table 6 reports four bandwidth specifications, of which three (±3 to ±5) satisfy the extended covariate-balance diagnostics from Section 4.4.1. Furthermore, ±2 is retained only as a sensitivity specification, since it fails the extended balance test though it satisfies the narrower two-covariate criterion. This avoids searching across arbitrary widths, though it narrows the range of direct bandwidth evidence available. A related check examines the baseline floor area, a structural characteristic fixed at construction and not plausibly responsive to green-finance eligibility, for discontinuity at the Band C threshold: none is found within the ±5 window (difference in means = 0.13, $p = 0.77$). This check is not independent placebo evidence; however, floor area is one of the two covariates used to select the ±2–±5 estimation windows in Section 4.4.1, so the absence of a discontinuity here restates rather than independently corroborates that window-selection balance test. A placebo check using a variable not used in window selection is identified as a priority for further work in Section 6.3. Third, the policy-rate collinearity could not be resolved: neither MEESs 2018 nor the predetermined exposure-by-rate interaction produced a stable estimate. Fourth, a further issue concerns the same-unit filtering procedure (Section Constructing a Conservatively Filtered Same-Unit Panel). Roughly 30% of candidate pairs are excluded. This primarily reflects deliberate, conservative quality control: the exact-match criteria are designed to discard ambiguous pairs, not retain potentially cross-unit ones. The 5 m^2 floor-area tolerance underlying this procedure was swept across six alternative values (2, 3, 5, 8, 10, and 15 m^2) using an independently reconstructed version of the linkage pipeline: the confirmation rate rises from 55% at 2 m^2 to 79% at 15 m^2, while the sign of the headline H1 estimate remains negative and its magnitude and significance are, if anything, slightly stronger at looser tolerances (from −0.191, $p = 0.008$, at 2 m^2 to −0.219, $p < 0.001$, at 15 m^2 in this reconstruction), indicating the result is not an artefact of the specific tolerance chosen. Approximate (fuzzy) address matching was also tested as an alternative to exact normalized-address matching, using Jaro–Winkler string similarity at thresholds from 0.85 to 1.00: the confirmation rate is stable within a 1.1-percentage-point range (66.9% to 68.0%) across this whole range, indicating that floor-area tolerance, not address-matching strictness, is the more consequential filtering choice. The confirmation rate also varies substantially across the sample period (Section Constructing a Conservatively Filtered Same-Unit Panel), exceeding 79% in every year from 2013 to 2017, falling to 74.8% in 2018 and 66.7% in 2019, and reaching a low of 52.5% in 2022 before recovering modestly to 56.9% in 2023, broadly coinciding with RdSAP methodology revisions; this compositional shift means later-year observations in the filtered panel are drawn from a more selected subset of candidates than earlier-year observations, a limitation not addressed by the tolerance and matching-strictness checks reported here. This reconstruction reproduces the audit-trail counts closely but not exactly (approximately 130,000 base high-floor records and 21,100 candidate pairs, versus 132,738 and 21,680 reported in Section 4.1). It also

used a simplified specification (year fixed effects only, without postcode fixed effects) for computational tractability. The point estimates above are therefore directional evidence of stability, not a direct replication of Table 2. Manual verification of accepted and rejected pairs against an independent ground-truth source remains outstanding, and is reported as such, not as completed. A manual review based solely on the same EPC address and floor-area fields would not constitute independent validation of dwelling identity; external validation would require an authoritative property-level ground-truth source, which is beyond the scope and data access of the present study. Fifth, high-floor classification relies on floor position as a proxy for building height. This cannot be independently verified against the Building Safety Act 2022 threshold. Sixth, the eligibility threshold sensitivity analysis (Section 4.2) covers Band C and Band B but not Band D; extending it further is a direction for future robustness checking. Seventh, the analysis is restricted to England and Wales. Existing Scottish homes are also assessed using RdSAP [56], but Scotland maintains a separate register, implementation history, regulatory framework, and administrative conventions, so the findings should not be generalized across jurisdictions without separate verification. An eighth, broader limitation is that this study's results are bounded by the data-quality characteristics of the EPC register itself, discussed further below.

This limitation warrants direct discussion, since it is not specific to the identification strategies used here: it affects the underlying data source itself. Independent national-scale audits of the EPC register have documented non-trivial measurement errors in the assessor-recorded score. A study of 1.6 million repeated assessments in England and Wales estimated a one-standard-deviation error that declines with EPC performance, from approximately ± 8 points near a rating of 35 to approximately ± 2.4 points near a rating of 85—large enough in the lower range to shift a dwelling across an official band, with an estimated 24% of Band D dwellings potentially misclassified as Band C [33]. A separate audit of the open EPC database found a proportion of records flagged by specific data-quality rules, from which an estimated underlying error rate of between 36% and 62% was derived. This was concentrated disproportionately in flats and maisonettes [34], the property type examined in this study. A further, related concern is the energy performance gap between EPC-predicted and metered energy use. A national analysis of the existing housing stock in England and Wales found this gap to vary substantially by region, built form, and construction age [64]. Whatever its value as an eligibility threshold, the EPC score is an imperfect proxy for a dwelling's true operational energy performance. These data-quality characteristics are external to the identification strategies developed in Section 4, and would affect any study relying on the EPC register as its primary data source. They are noted here as a boundary condition on generalizability, not a limitation specific to this paper's methods.

A ninth limitation concerns the two evidence lines for H1 (Sections 5.2 and 5.3). Both rely on the same baseline score variable and the same filtered panel. A systematic measurement-error channel affecting that score could therefore bias both tests in the same direction, beyond mere sample overlap. A bootstrap co-variation check (Section 6.1) finds a small, statistically distinguishable-from-zero positive correlation between the two test statistics across 5000 property-level resamples ($r = 0.152$, 95% CI [0.125, 0.179]). This suggests a limited but non-zero shared-noise channel—far below the near-perfect co-movement that would mean the two tests are re-deriving the same result, but enough that the two lines of evidence are correlated, not fully independent, when assessing their combined weight.

A tenth limitation concerns selection into the filtered panel itself (Section 4.1): only high-floor flats that happened to receive a second EPC inspection within 2013–2023 are usable for the timing correction, roughly 15% of the eligible stock. A sensitivity check

finds that reinspected properties have a markedly lower initial energy-efficiency score on average than properties inspected only once (65.9 versus 71.6; $p < 0.001$), and a substantially lower owner-occupied share (1.6% versus 8.1%; $p < 0.001$). This is consistent with reinspection being concentrated among rented, worse-performing stock undergoing sale, MEESs-driven compliance, or refinancing. The filtered panel represents reinspected, disproportionately rented, high-floor flats, not the high-floor stock as a whole, and the findings do not necessarily generalize beyond that reinspected subpopulation.

An eleventh limitation concerns multiple comparisons: Table 2 reports eight timing–panel-threshold combinations, Table 6 reports four bandwidth specifications, of which three (± 3 to ± 5) satisfy the extended covariate-balance diagnostics, while ± 2 is retained as an additional sensitivity specification, and the paper additionally examines two identification strategies for H2/H3, without a formal correction (such as Bonferroni or Holm) applied across this set of tests. No such correction is applied, because these are not independent hypothesis draws in the multiple-testing sense. They are robustness checks of the same underlying claim—the sign reversal under corrected timing—across specifications that share substantial data overlap. The relevant benchmark is the qualitative consistency of sign and magnitude across the set, not a family-wise error rate. The nominal p-values reported throughout describe each individual specification; they are not corrected for the search across panel constructions, thresholds, and bandwidths collectively.

Three directions for future research follow from these limitations. The measurement-timing problem is, in principle, general to any evaluation setting where an eligibility flag is built from the same source as the outcome. A credible test of H2 and H3 would need loan-level financing-cost data with more monetary-policy variation than 2013–2021 provided. A stakeholder-level extension of the framework (Section 3.2) would help explain why the RDD-identified discontinuity is negative but modest—more consistent with a weak or heterogeneous pattern in recorded EPC-score changes than a large, uniform response—without asserting that this pattern reflects a confirmed financing channel.

The transferability of these findings depends on which feature of the setting is in question. The general methodological point, that an eligibility indicator built from the same source data as the outcome it explains can be close to definitionally related to that outcome when timed after rather than before it, does not depend on England and Wales, RdSAP, or Band C specifically: it applies to any administrative or regulatory classification constructed from repeated measurements of the same underlying quantity used later as an outcome, including other national EPC schemes, other environmental certifications (e.g., LEED, BREEAM, ENERGY STAR), and non-property sustainability metrics more broadly. Two features are less likely to transfer directly, however. First, the specific magnitude and pattern of measurement error documented here (assessor rounding, heaping, and revisions to the underlying calculation methodology) reflect RdSAP's reduced-data, assessor-recorded design; EPC schemes elsewhere in the EU vary in whether they rely on modelled inputs of this kind or on directly measured energy consumption, and metered schemes would not be expected to exhibit the same heaping pattern, though they may introduce other measurement-timing risks of their own. Second, certification schemes such as LEED and BREEAM are typically awarded once at a point in time, rather than reassessed on a comparable repeated basis as EPCs are, so the specific baseline-versus-follow-up timing choice examined here may not map onto them in the same way; the more general lesson, that the timing of the measurement relative to the outcome deserves explicit justification rather than being treated as incidental, still applies.

7. Conclusions

Methodologically, this study uses a conservative same-unit filtering procedure to identify and address a previously unreported panel-construction problem in UPRN-linked EPC data. It also shows that an eligibility indicator built from the same source data as the outcome it explains can become close to definitionally related to that outcome when timed after the event it predicts. This is the paper's central contribution: a measurement-timing risk that applies to any evaluation constructing an eligibility flag from repeated records of the same underlying quantity.

Empirically, using EPC register data for England and Wales (2013–2023) among reinspected high-floor residential properties, we find that a conventional specification's positive threshold-status–EPC-change association comes largely from this timing artefact, not a confirmed financing effect. Measured from the baseline score instead, the association turns negative across every combination tested. A regression discontinuity design, using a different estimation strategy on the same underlying score variable, corroborates this reversal. The reversed association is not stable across every bandwidth choice, so we do not claim a causal effect in the opposite direction. Two further strategies aimed at the role of monetary policy conditions produced no credible estimates; this moderating role remains an open question.

For policy and practice, the findings matter for evaluations that use EPC-based classifications as proxies for financial or policy eligibility. They do not provide evidence about the effectiveness of any specific green-finance product. Eligibility indicators built from the same data used to measure recorded EPC-score-change outcomes need explicit attention to measurement timing: an unadjusted correlation's direction can hinge on this choice alone.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su18179023/s1, File S1: H2 and H3 identification diagnostics. Part A reports full specification tests, event-study results, and robustness checks for the difference-in-differences design described in Section 4.4.2. Part B reports the construction, within-year variance diagnostics, and leave-one-year-out results for the predetermined exposure-by-rate interaction described in the same section [65].

Author Contributions: Conceptualization, K.T. and C.P.; methodology, K.T.; software, C.P.; validation, K.T.; formal analysis, C.P.; investigation, C.P.; resources, K.T.; data curation, C.P.; writing—original draft preparation, C.P.; writing—review and editing, K.T. and C.P.; visualization, C.P.; supervision, K.T. All authors have read and agreed to the published version of the manuscript.

Funding: The research was partially funded by the Thammasat University Research Unit in Project Development and Innovation in Real Estate Business, Thammasat University, grant number 6407035.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Energy Performance Certificate register used in this study is publicly available from the UK Government's open data portal. The same-unit filtering procedure and filter specification are described in detail in Section Constructing a Conservatively Filtered Same-Unit Panel, with additional sensitivity analyses reported in Section 6.3. The same-unit confirmation procedure code is deposited in a permanent, publicly accessible repository: https://doi.org/10.5281/zenodo.22124537. The derived analysis panel itself remains available from the corresponding author upon reasonable request, since it is built from the publicly available register using this deposited code.

Acknowledgments: During the preparation of this manuscript, the authors used Claude Sonnet 5 (Anthropic) and GPT-5 (OpenAI) as assistive tools for language editing and manuscript preparation. All research design decisions, data preparation, statistical analyses, coding, interpretation of results,

and conclusions were undertaken by the authors, who reviewed all AI-assisted output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. United Nations Environment Programme; Global Alliance for Buildings and Construction. *Global Status Report for Buildings and Construction: Beyond Foundations—Mainstreaming Sustainable Solutions to Cut Emissions from the Buildings Sector*; UNEP: Nairobi, Kenya, 2024.
2. IPCC. *Climate Change 2022: Mitigation of Climate Change*; Cambridge University Press: Cambridge, UK, 2022. [CrossRef]
3. Nolan, T.; Saeedian, A.; Taherpour, P.; Aghamolaei, R. Integrated Life Cycle Assessment of Residential Retrofit Strategies: Balancing Operational and Embodied Carbon, Lessons from an Irish Housing Case Study. *Sustainability* **2025**, *17*, 8173. [CrossRef]
4. Amirkhani, M.; Martek, I.; Luther, M.B. Mapping Research Trends in Residential Construction Retrofitting: A Scientometric Literature Review. *Energies* **2021**, *14*, 6106. [CrossRef]
5. Zhang, H.; Hewage, K.; Karunathilake, H.; Feng, H.; Sadiq, R. Research on policy strategies for implementing energy retrofits in the residential buildings. *J. Build. Eng.* **2021**, *43*, 103161. [CrossRef]
6. Hoeller, P.; Ziemann, V.; Cournède, B.; Bétin, M. *Home, Green Home: Policies to Decarbonise Housing*; (OECD Economics Department Working Papers No. 1751); OECD Publishing: Paris, France, 2023. [CrossRef]
7. Maina, P.; Gyenge, B.; Fekete-Farkas, M.; Parádi-Dolgos, A. Analyzing Trends in Green Financial Instrument Issuance for Climate Finance in Capital Markets. *J. Risk Financ. Manag.* **2024**, *17*, 145. [CrossRef]
8. Cubi Montanya, E.; Keith, D.W. LEED, Energy Savings, and Carbon Abatement: Related but Not Synonymous. *Environ. Sci. Technol.* **2011**, *45*, 1757–1758. [CrossRef] [PubMed]
9. An, X.; Pivo, G. Green Buildings in Commercial Mortgage-Backed Securities: The Effects of LEED and Energy Star Certification on Default Risk and Loan Terms. *Real Estate Econ.* **2020**, *48*, 7–42. [CrossRef]
10. Ahmed, I.; Socci, C.; Medabesh, A.; Severini, F.; Zotti, J. Economic impact of monetary policy: Focus on real estate sector in Italy. *Int. J. Financ. Econ.* **2021**, *26*, 1256–1269. [CrossRef]
11. Bernanke, B.S.; Gertler, M. Inside the black box: The credit channel of monetary policy transmission. *J. Econ. Perspect.* **1995**, *9*, 27–48. [CrossRef]
12. Mishkin, F.S. *The Channels of Monetary Transmission: Lessons for Monetary Policy*; NBER Working Paper No. 5464; National Bureau of Economic Research: Cambridge, MA, USA, 1996. [CrossRef]
13. Bacchiocchi, A.; Ille, S.; Giombini, G. The effects of a green monetary policy on firms financing cost. *J. Econ. Interact. Coord.* **2024**, *19*, 727–757. [CrossRef]
14. Ma, Z.; Jiang, X. The Impact of Green Finance on Energy Transition Under Climate Change. *Sustainability* **2025**, *17*, 7112. [CrossRef]
15. Desalegn, G.; Tangl, A. Enhancing Green Finance for Inclusive Green Growth: A Systematic Approach. *Sustainability* **2022**, *14*, 7416. [CrossRef]
16. Bai, Y.; Tsion, A.T.; Rosales, R.; Shollock, B.; He, W. Gated multimodal learning for interpretable property energy performance prediction and retrofit scenario analysis. *arXiv* **2026**. [CrossRef]
17. del Mar Solà, M.; de Ayala, A.; Galarraga, I.; Escapa, M. Promoting energy efficiency at household level: A literature review. *Energy Effic.* **2021**, *14*, 6. [CrossRef]
18. Azimi, S.; Hon, C.K.H.; Tyvimaa, T.; Skitmore, M. Barriers to Energy Efficiency: Low-Income Households in Australia. *Buildings* **2023**, *13*, 954. [CrossRef]
19. Tozer, L.; MacRae, H.; Smit, E. Achieving deep-energy retrofits for households in energy poverty. *Build. Cities* **2023**, *4*, 258–273. [CrossRef]
20. Huang, C.; Ma, J.; Song, K. Homeowners' willingness to make investment in energy efficiency retrofit of residential buildings in China and its influencing factors. *Energies* **2021**, *14*, 1260. [CrossRef]
21. Kerr, N.; Winskel, M. Household investment in home energy retrofit: A review of the evidence on effective public policy design for privately owned homes. *Renew. Sustain. Energy Rev.* **2020**, *123*, 109778. [CrossRef]
22. Yardımcı, P.; Oskay, C. The Impact of Green Finance and Financial Globalization on Environmental Sustainability: Empirical Evidence from Türkiye. *Sustainability* **2025**, *17*, 5696. [CrossRef]
23. Bertoldi, P.; Economidou, M.; Palermo, V.; Boza-Kiss, B.; Todeschi, V. How to finance energy renovation of residential buildings: Review of current and emerging financing instruments in the EU. *Wiley Interdiscip. Rev. Energy Environ.* **2021**, *10*, e384. [CrossRef]
24. Prabatha, T.; Hewage, K.; Sadiq, R. An Incentives Planning Framework for Residential Energy Retrofits: A Life Cycle Thinking-Based Analysis under Uncertainty. *Sustainability* **2023**, *15*, 5479. [CrossRef]
25. Flammer, C. Corporate green bonds. *J. Financ. Econ.* **2021**, *142*, 499–516. [CrossRef]

26. NatWest. Green Mortgages: Energy-Efficient Homes. Available online: https://www.natwest.com/mortgages/mortgage-comparison/green-mortgages.html (accessed on 22 July 2026).
27. Barclays. Green Home Mortgage. Available online: https://www.barclays.co.uk/mortgages/green-home-mortgage/ (accessed on 22 July 2026).
28. HM Government. *The Energy Efficiency (Private Rented Property) (England and Wales) Regulations 2015*; SI 2015/962; Legislation: Richmond, UK, 2015. Available online: https://www.legislation.gov.uk/uksi/2015/962 (accessed on 26 January 2026).
29. Ofgem. *Energy Company Obligation (ECO4) Guidance: Delivery*; Version 4.0; Ofgem: London, UK, 2026. Available online: https://www.ofgem.gov.uk/sites/default/files/2026-03/ECO4%20delivery%20guidance%20version%204.0.pdf (accessed on 26 January 2026).
30. HM Government. *Electricity and Gas (Energy Company Obligation) Order 2022*; SI 2022/875; Legislation: Richmond, UK, 2022. Available online: https://www.legislation.gov.uk/uksi/2022/875 (accessed on 26 January 2026).
31. Department for Energy Security and Net Zero. Warm Homes: Social Housing Fund. GOV.UK. 2024. Available online: https://www.gov.uk/government/publications/social-housing-decarbonisation-fund (accessed on 26 January 2026).
32. Building Research Establishment. *RdSAP 10 Technical Specification*, Version 10.2; BRE: Watford, UK, 2024. Available online: https://files.bregroup.com/SAP/RdSAP10-dt13.02.2024.pdf (accessed on 26 January 2026).
33. Crawley, J.; Biddulph, P.; Northrop, P.J.; Wingfield, J.; Oreszczyn, T.; Elwell, C. Quantifying the Measurement Error on England and Wales EPC Ratings. *Energies* **2019**, *12*, 3523. [CrossRef]
34. Hardy, A.; Glew, D. An analysis of errors in the Energy Performance Certificate database. *Energy Policy* **2019**, *129*, 1168–1178. [CrossRef]
35. Heo, Y.; Choudhary, R.; Augenbroe, G.A. Calibration of building energy models for retrofit analysis under uncertainty. *Energy Build.* **2012**, *47*, 550–560. [CrossRef]
36. Kim, J.; Myoung, J.; Lim, H.; Song, D. Efficiency Gap Caused by the Input Data in Evaluating Energy Efficiency of Low-Income Households' Energy Retrofit Program. *Sustainability* **2020**, *12*, 2774. [CrossRef]
37. Chen, Q.; Qi, J.; Yan, G. Are outcome variables always bad controls? The case of mediating variables. *Appl. Econ. Lett.* **2025**, *32*, 2066–2070. [CrossRef]
38. Imbens, G.W.; Wooldridge, J.M. Recent developments in the econometrics of program evaluation. *J. Econ. Lit.* **2009**, *47*, 5–86. [CrossRef]
39. Berg, F.; Kölbel, J.F.; Rigobon, R. Aggregate Confusion: The Divergence of ESG Ratings. *Rev. Financ.* **2022**, *26*, 1315–1344. [CrossRef]
40. Buyuklieva, B.; Dennett, A.; Bailey, N.; Morley, J. Energy Efficient Homes: The Social and Spatial Patterns of Residential Energy Efficiency in England. In Proceedings of the GISRUK 2023: GIS Research UK Conference, Glasgow, UK, 18–21 April 2023. [CrossRef]
41. Herzog, T.N.; Scheuren, F.J.; Winkler, W.E. *Data Quality and Record Linkage Techniques*; Springer: Berlin/Heidelberg, Germany, 2007.
42. Ministry of Housing, Communities and Local Government. Energy Performance of Buildings Data: England and Wales—Guidance. 2023. Available online: https://epc.opendatacommunities.org/docs/guidance (accessed on 26 January 2026).
43. Rodriguez, N.J.; Katooziani, A.; Jeelani, I. Barriers to energy-efficient design and construction practices: A comprehensive analysis. *J. Build. Eng.* **2024**, *82*, 108349. [CrossRef]
44. Tovar Reaños, M.A.; Meier, D.; Curtis, J.; Pillai, A. The role of energy, financial attitudes and environmental concerns on perceived retrofitting benefits and barriers: Evidence from Irish home owners. *Energy Build.* **2023**, *297*, 113448. [CrossRef]
45. Bastos, P.; Castro, L.; Cristia, J.; Scartascini, C. Does Energy Consumption Respond to Price Shocks? Evidence from a Regression-Discontinuity Design. *J. Ind. Econ.* **2015**, *63*, 249–278. [CrossRef]
46. Bondonio, D. Does the Running Variable Matter? A Second Look at Discontinuity Designs for Evaluating Regional Economic Development and Business Incentive Policies. *Eval. Rev.* **2022**, *46*, 598–625. [CrossRef] [PubMed]
47. Dong, Y.; Hou, W.; Lin, B.; Zhang, T. Recent advances and future directions in macro-finance: Macroeconomic conditions and corporate decisions. *Eur. J. Financ.* **2022**, *28*, 307–313. [CrossRef]
48. Aikman, D.; Lehnert, A.; Liang, N.; Modugno, M. Credit, Financial Conditions, and Monetary Policy Transmission. *Int. J. Cent. Bank.* **2020**, *16*, 141–179.
49. Vonlanthen, J. Interest rates and real estate prices: A panel study. *Swiss J. Econ. Stat.* **2023**, *159*, 6. [CrossRef]
50. Cloyne, J.; Ferreira, C.; Froemel, M.; Surico, P. Monetary policy, corporate finance, and investment. *J. Eur. Econ. Assoc.* **2023**, *21*, 2586–2634. [CrossRef]
51. Bacchiocchi, A.; Favaretto, F.; Giombini, G.; Tramontana, F. Greenflation, climateflation and monetary policy: The dynamics of sustainable transition. *Macroecon. Dyn.* **2025**, *29*, e137. [CrossRef]
52. MacRae, H.; Tozer, L. The use of green bonds in financing energy retrofits in buildings. *Energy Res. Soc. Sci.* **2024**, *112*, 103500. [CrossRef]

53. Katris, A.; Turner, K. Achieving economy-wide gains from residential energy efficiency improvements: The importance of timing and funding approach in driving the transition. *Environ. Policy Gov.* **2025**, *35*, 13–25. [CrossRef]
54. Bjelland, D.; Brozovsky, J.; Hrynyszyn, B.D. Systematic review: Upscaling energy retrofitting to the multi-building level. *Renew. Sustain. Energy Rev.* **2024**, *198*, 114402. [CrossRef]
55. Bergman, N.; Foxon, T.J. Reframing policy for the energy efficiency challenge: Insights from housing retrofits in the United Kingdom. *Energy Res. Soc. Sci.* **2020**, *63*, 101386. [CrossRef]
56. Scottish Government. Energy Performance Certificate (EPC) Reform: Consultation. Gov.Scot. 2023. Available online: https://www.gov.scot/publications/energy-performance-certificate-epc-reform-consultation/ (accessed on 26 January 2026).
57. Building Safety Act 2022, c. 30 (UK). Available online: https://www.legislation.gov.uk/ukpga/2022/30 (accessed on 26 January 2026).
58. Department for Energy Security and Net Zero. Improving the Energy Performance of Privately Rented Homes: Government Response. GOV.UK. 2026. Available online: https://www.gov.uk/government/consultations/improving-the-energy-performance-of-privately-rented-homes-2025-update/outcome/improving-the-energy-performance-of-privately-rented-homes-government-response-html (accessed on 26 January 2026).
59. Cattaneo, M.D.; Jansson, M.; Ma, X. Simple Local Polynomial Density Estimators. *J. Am. Stat. Assoc.* **2020**, *115*, 1449–1455. [CrossRef]
60. Cattaneo, M.D.; Frandsen, B.R.; Titiunik, R. Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate. *J. Causal Inference* **2015**, *3*, 1–24. [CrossRef]
61. Cattaneo, M.D.; Titiunik, R.; Vázquez-Bare, G. Inference in Regression Discontinuity Designs under Local Randomization. *Stata J.* **2016**, *16*, 331–367. [CrossRef]
62. Nipper, M.; Ostermaier, A.; Theis, J. Mandatory disclosure of standardized sustainability metrics: The case of the EU Taxonomy Regulation. *Corp. Soc. Responsib. Environ. Manag.* **2025**, *32*, 2171–2190. [CrossRef]
63. Sica, F.; Tajani, F.; Sáez-Pérez, M.P.; Marín-Nicolás, J. Taxonomy and Indicators for ESG Investments. *Sustainability* **2023**, *15*, 15979. [CrossRef]
64. Firth, S.; Allinson, D.; Watson, S. Quantifying the spatial variation of the energy performance gap for the existing housing stock in England and Wales. *J. Build. Perform. Simul.* **2025**, *18*, 593–610. [CrossRef]
65. Wolfers, J. Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results. *Am. Econ. Rev.* **2006**, *96*, 1802–1820. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.