Do Energy Labels Move Markets? A Well-Powered Null from English Property Transactions

APEP Autonomous Research* olafdrw

March 10, 2026

Abstract

A property rated E can be rented; one rated F cannot. England's Minimum Energy Efficiency Standards create a regulatory cliff at the E/F boundary on the Energy Performance Certificate, while four other boundaries remain purely informational. Using 7.1 million pre-linked transactions in a multi-cutoff regression discontinuity design, I find no significant price discontinuity at any boundary. At C/D—the cleanest threshold—the estimate is 0.1 percent ($p = 0.88$), ruling out effects larger than 8 percent. At E/F, density manipulation limits causal interpretation, but the point estimate is similarly null (-2.6%, $p = 0.36$). A difference-in-discontinuities around the 2018 MEES enactment, the 2021–2023 energy crisis, and full-sample validation all confirm the null. Discrete EPC thresholds do not generate price jumps—though the market may price efficiency smoothly through the continuous score.

JEL Codes: Q48, R31, D83, L51

Keywords: energy efficiency, EPC labels, housing prices, regression discontinuity, MEES regulation, energy crisis

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: N/A).

1. Introduction

In April 2018, England banned landlords from letting properties rated below band E on the Energy Performance Certificate. A dwelling scoring 39 on the Standard Assessment Procedure—band E—can legally enter the rental market. One scoring 38—band F—cannot. This single-point distinction, embedded in the Minimum Energy Efficiency Standards (MEES), created a regulatory cliff at a threshold that had previously carried only informational weight.

The English EPC system offers a rare laboratory for testing whether energy labels move housing markets. Every property receives a continuous energy efficiency score from 1 to 100, mapped to letter grades A through G at five fixed thresholds. Four of these thresholds—D/E, C/D, B/C, and A/B—change only the color of the label. The fifth—E/F—determines whether a landlord can legally let the property. If energy labels generate discrete price jumps, they should be visible at these thresholds. If regulation moves markets, the E/F boundary should dominate. A multi-cutoff regression discontinuity design tests both hypotheses simultaneously.

I use the pre-linked dataset of Chi et al. (2023), which matches 21 million Land Registry transactions to Energy Performance Certificates through a UPRN-based address matching algorithm. Restricting to transactions from January 2015 through October 2023 with valid EPC scores yields 7.1 million observations. I estimate on a 500,000-observation random subsample for computational feasibility with `rdrobust`; random sampling preserves the local properties of the RDD. The UPRN-based matching is classified as "address-matched" (89% of observations), providing built-in validation against measurement error. The sample spans the pre-MEES baseline, the post-MEES implementation, the 2021–2023 energy crisis, and the subsequent normalization.

The central finding is a precisely estimated null at the cleanest informational threshold. At C/D—the only boundary passing both the McCrary density test ($p = 0.220$) and most covariate balance checks—the estimate is 0.1 percent ($p = 0.88$), ruling out positive effects larger than 8 percent. At E/F, significant density manipulation (McCrary $p < 0.001$) limits the causal interpretation of the sharp RDD: the point estimate is -2.7 percent ($p = 0.36$), but properties near this threshold are plausibly sorted. D/E and B/C are similarly null but also show density irregularities. These results are robust to inclusion or exclusion of covariates, bandwidth choice, polynomial order, donut specifications, and restriction to the address-matched subsample.

Three additional tests rule out a hidden effect. First, a difference-in-discontinuities design comparing E/F before and after MEES enactment (April 2018) shows a suggestive but statistically insignificant change ($\Delta\hat{\tau} = -0.15$, $p = 0.18$). The triple-difference netting out C/D trends is similarly null ($p = 0.20$). Second, the 2021–2023 energy crisis—which tripled household energy bills and should have maximized label salience—produced no detectable amplification at any boundary. Third, tenure-specific estimates at E/F show no consistent positive effects—the one marginally significant estimate (owner-occupied post-MEES, $p = 0.047$) is negative, the wrong sign for regulatory capitalization.

This null is informative, not merely inconclusive. At C/D, over 255,000 observations lie within the optimal bandwidth, and the 95% confidence interval rules out positive effects larger than 8 percent. At E/F, approximately 30,000 observations are in the bandwidth, ruling out positive effects larger than 6 percent—though the E/F confidence interval must be interpreted cautiously given density manipulation. Prior hedonic studies finding 5–10 percent EPC premia (Fuerst et al., 2015; Aydin et al., 2020) estimate a different parameter: the average price difference across bands, which confounds energy efficiency with correlated property characteristics. The RDD identifies the marginal effect of crossing a threshold, holding the continuous score approximately constant. The absence of threshold effects does not preclude smooth pricing of the continuous SAP score.

The paper contributes to three literatures. It complements the energy efficiency capitalization literature (Eichholtz et al., 2010; Fuerst et al., 2015; Aydin et al., 2020) by showing that hedonic "EPC premia" do not manifest as discrete threshold effects in an RDD framework—the two approaches identify different parameters, and the absence of threshold effects suggests that buyers may respond to the continuous score rather than the letter grade. It speaks to the energy efficiency gap debate (Allcott and Greenstone, 2014; Gerarden et al., 2017): if coarse label boundaries do not generate discrete market responses, then the labeling scheme's design—the mapping from continuous scores to categorical grades—may be less salient than assumed. And it provides suggestive (though not cleanly identified, given manipulation) evidence that the MEES regulatory cliff at E/F does not produce detectable price capitalization—consistent with weak enforcement, widespread exemptions, or continuous pricing of efficiency.

The remainder proceeds as follows. Section 2 describes the EPC system, MEES, and the energy crisis. Section 3 reviews the related literature. Section 4 presents the data source and matching procedure. Section 5 outlines the empirical strategy. Section 6 presents results. Section 7 reports robustness checks. Section 8 discusses implications, and Section 9 concludes.

2. Institutional Background

2.1 Energy Performance Certificates in England

The Energy Performance Certificate system was introduced in England and Wales in 2007, implementing the European Union's Energy Performance of Buildings Directive. An EPC is required whenever a property is built, sold, or rented, and remains valid for ten years. The certificate reports an energy efficiency rating from A (most efficient) to G (least efficient), determined by the Standard Assessment Procedure (SAP) score—a continuous measure running from 1 to 100 that summarizes expected energy costs per square meter under standardized occupancy assumptions.

The SAP score is computed from physical characteristics: wall insulation, roof insulation, glazing, heating system type and efficiency, hot water system, lighting, and renewable energy generation. The score does not depend on occupant behavior, energy prices, or actual consumption—it is purely engineering-based. Letter grades follow fixed thresholds: A (≥ 92), B (81–91), C (69–80), D (55–68), E (39–54), F (21–38), and G (1–20).

Assessments are conducted by accredited domestic energy assessors who visit the property and record physical characteristics. While there is scope for measurement error and some assessor discretion at borderline scores, the assessment is fundamentally determined by observable building features. The EPC is displayed prominently in property listings, with the color-coded arrow chart required by law since 2012.

2.2 Minimum Energy Efficiency Standards

The MEES regulations came into force on 1 April 2018 for new tenancies in the private rented sector, extended to all existing tenancies from 1 April 2020. MEES prohibits landlords from granting tenancies on properties rated below band E unless a valid exemption applies, with penalties of up to £5,000 per property.

This creates a sharp discontinuity unique among EPC thresholds. A property scoring 39 (band E) can be legally rented; one scoring 38 (band F) cannot, absent an exemption. Landlords holding sub-E properties must invest in improvements, register an exemption (if no cost-effective improvements are available under the £3,500 spending cap), or withdraw the property from the rental market.

The key identification advantage: MEES creates both a regulatory and informational discontinuity at E/F, while all other boundaries carry only informational content. By comparing E/F with D/E and C/D, I isolate the regulatory component.

2.3 MEES Anticipation

MEES was announced in the Energy Act 2015 and widely discussed before formal implementation (UK Parliament, 2015). Market participants—particularly property investors—may have anticipated the regulation and begun pricing in the E/F distinction before it took legal effect. The pre-MEES E/F discontinuity, which I estimate separately, bears on this anticipation channel. The difference-in-discontinuities design (Section 5.5) provides a formal test by comparing the change in the E/F discontinuity around the enactment date against the change at C/D. If anticipation fully priced in MEES before April 2018, the diff-in-disc should show no additional break.

2.4 The Energy Crisis of 2021–2023

The UK experienced an acute energy price crisis beginning in autumn 2021. The Ofgem energy price cap rose from approximately £1,138 in October 2021 to a peak of £3,549 in January 2023—bills roughly tripled in eighteen months. For a typical household, the difference between an E-rated and F-rated property in annual heating costs widened from perhaps £200–300 to £600–900. This price shock dramatically increased the salience of energy efficiency to homebuyers, potentially amplifying label premia through both informational and regulatory channels.

2.5 MEES Enforcement and Exemptions

Understanding the enforcement regime is important for interpreting the null. MEES enforcement rests with local authorities, who may issue compliance notices and financial penalties (up to £5,000 per property for breaches lasting three months or more). However, enforcement has been described as patchy: local authorities have limited resources for housing enforcement, and the burden of proof falls on the authority to demonstrate non-compliance. The exemption system further dilutes the regulatory bite. Landlords may register an "all improvements made" exemption if they have installed all improvements meeting the £3,500 cap and the property remains below E, or a "no cost-effective improvements" exemption if no qualifying measures exist. High-cost exemptions apply when no improvement can be made for under £3,500. These exemptions are self-registered on the PRS Exemptions Register with minimal verification.

The weakness of enforcement and the breadth of exemptions suggest that MEES may function more as a soft nudge than a hard constraint for many landlords. Properties that cannot cost-effectively reach band E can largely remain in the rental market through the exemption system. This institutional detail is directly relevant to interpreting the null: if most F-rated rental properties can obtain exemptions, the E/F boundary carries less regulatory force than the statutory text implies.

3. Related Literature

3.1 Energy Efficiency Capitalization

A large literature estimates the price premium associated with energy efficiency in housing markets. Eichholtz et al. (2010) document a 3% premium for Energy Star-certified commercial buildings in the US, while Fuerst et al. (2015) find 5% premia for EPC band A/B versus D-rated dwellings in England and Wales using hedonic regressions. Brounen et al. (2011) report 3.6% premia per EPC label step in the Dutch housing market. Aydin et al. (2020) find 2.2% per EPC label in the Netherlands, with larger effects for rental properties. Chegut et al. (2016) and Kahn (2007) document similar premia in commercial and residential markets.

These estimates share a common identification concern: hedonic regressions control for observable characteristics but cannot eliminate unobserved confounders correlated with both energy efficiency and price. High-EPC properties tend to be newer, better maintained, and located in more desirable areas. The RDD design used here addresses this directly: by comparing properties one SAP point apart at a band boundary, it eliminates all confounders that vary smoothly with the score.

3.2 The Energy Efficiency Gap

The "energy efficiency gap"—the shortfall between privately optimal and actual energy efficiency investment—has been extensively documented (Allcott and Greenstone, 2014; Gerarden et al., 2017; Gillingham et al., 2020). Proposed explanations include consumer inattention to energy costs, present bias, split incentives between landlords and tenants, and information barriers. Energy labels are designed to address the information channel: by making efficiency visible and salient, they should help consumers internalize long-run operating costs into purchase decisions.

The null result in this paper speaks directly to the information hypothesis. If EPC labels—which are legally mandated, prominently displayed, and universally available—fail to generate discrete price responses at band boundaries, then information provision alone may be insufficient to close the gap. This is consistent with Allcott and Greenstone (2014), who find modest effects of information treatments on energy-related decisions, and with Davis and Metcalf (2014), who show that home energy reports produce small behavioral changes.

3.3 Regulatory Capitalization in Housing

A parallel literature examines whether building regulations are capitalized into property values. Best and Kleven (2018) study the effect of environmental regulations on housing supply and prices. Surminski and Eldridge (2017) examine flood risk regulation and property values. The general finding is that binding regulations—which constrain the use or value of an asset—are typically reflected in market prices. The MEES null is surprising in this context: a regulation that prohibits the letting of sub-E properties should, in principle, reduce the value of F-rated properties (or increase the value of E-rated properties) near the boundary.

3.4 Salience and Information Design

The broader salience literature (Chetty et al., 2009) emphasizes that information must be salient to affect behavior. EPC labels are visually prominent—the color-coded arrow chart is a legal requirement in property listings since 2012—but the discrete band may be less salient than the continuous score. If buyers attend to the numerical score rather than the letter grade, no discontinuity would emerge at band boundaries even if energy efficiency is fully priced. This interpretation is consistent with the null: the market may price energy efficiency continuously, with no additional effect from crossing a discrete threshold.

3.5 RDD in Housing and Policy

Regression discontinuity designs have been applied extensively in housing markets. Ghosh et al. (2024) use an RDD to study homebuyer responses to flood risk information. Romani et al. (2026) estimate RDD effects of energy price shocks on housing values. The multi-cutoff framework of Cattaneo et al. (2016) has been used to study heterogeneous treatment effects across policy thresholds. This paper adapts the multi-cutoff framework to the EPC system, exploiting the institutional feature that one of five boundaries carries regulatory force while the others are purely informational.

4. Data

4.1 Data Sources

I use the pre-linked LR-PPD and EPC dataset of Chi et al. (2023), which combines the universe of Land Registry Price Paid Data with the Domestic EPC Register for England and Wales.[1] Although the underlying dataset covers England and Wales, I restrict the analysis to England. MEES applies differently in Wales—the Welsh Government sets its own implementation timeline and exemption criteria under devolved housing powers—so pooling the two jurisdictions would conflate distinct regulatory regimes.

[1]The dataset is available from the UK Data Service ReShare repository (SN 856542) under Creative Commons Attribution 4.0 International license.

Land Registry Price Paid Data. The Land Registry records every residential property transaction in England and Wales, including the transaction price, date, postcode, property type, address fields, and local authority district.

EPC Register. The Domestic EPC Register contains over 25 million certificates issued since 2008. Each record includes the SAP score, letter grade, property type, floor area, tenure, and assessment date.

4.2 Matching Procedure

Chi et al. (2023) link the two datasets through a four-stage algorithm centered on the Unique Property Reference Number (UPRN). The UPRN is a persistent identifier assigned to every addressable location in Great Britain, enabling exact address-level matching. Of the linked records in my analysis sample, 89% are classified as "Address Matched" via UPRN, with the remainder matched through the energy assessor's reported address. This high-quality matching eliminates a key concern in the EPC capitalization literature: postcode-level matching noise that could generate spurious discontinuities.

The address-matched subsample serves as a built-in validation instrument. If a price discontinuity appears in both the full and address-matched samples, it cannot be an artifact of matching error. Conversely, a null in both samples provides strong evidence against label effects.

4.3 Match Diagnostics

Two diagnostic variables assess match quality: n_{epc}, the count of linked transactions at the same postcode (higher values indicate denser areas with more potential for mismatching); and Δ_{days}, the number of days between the EPC assessment and the transaction. I test for smooth variation of both diagnostics through each cutoff using RDD balance tests (Section 7.2).

4.4 Variable Construction

The outcome is the natural logarithm of the transaction price. The running variable at each boundary c is $X_i - c$, the EPC score centered at the cutoff. Time periods follow key policy events: Pre-MEES (January 2015–March 2018), Post-MEES Pre-Crisis (April 2018–September 2021), Crisis (October 2021–June 2023), and Post-Crisis (July 2023–October 2023). Tenure is classified from the EPC register's self-reported field, distinguishing private rentals from owner-occupied properties.

4.5 Summary Statistics

Table 1 reports summary statistics by EPC band. The sample spans the full efficiency range, with the largest concentrations in bands D and E. The address-match rate is reported for each band, providing transparency about data quality across the score distribution.

Table 1: Summary Statistics by EPC Band

	EPC Band							
	G	F	E	D	C	B	A	All
Observations	57,007	226,942	1,059,962	2,883,141	1,728,576	1,143,175	14,400	7,113,203
Mean price (£)	280,269	334,239	323,932	297,301	287,693	343,353	414,202	307,614
Median price (£)	209,000	245,000	240,000	235,000	228,000	285,000	350,000	243,076
Mean EPC score	11.3	31.8	48.2	62.2	73.2	83.9	94.6	65.0
Mean floor area (sq m)	100.1	108.1	101.2	93.6	90.3	98.6	119.9	95.3
% Flats	7.0	7.8	8.8	10.7	28.4	27.9	7.4	17.4
% New build	0.2	0.2	0.3	0.5	4.0	59.2	58.4	10.9
% Private rental	7.7	8.3	11.5	10.8	12.6	2.6	0.8	9.9
% Address-matched	87.5	90.1	90.6	88.0	86.0	94.0	89.6	88.9

Notes: Summary statistics for the matched EPC–Land Registry analysis sample, 2015–2023. Transaction prices are in nominal pounds. EPC scores range from 1 (worst) to 100 (best). Address-matched indicates the share of transactions linked at the individual property level.

Figure 1 displays the EPC score distribution in the analysis sample.

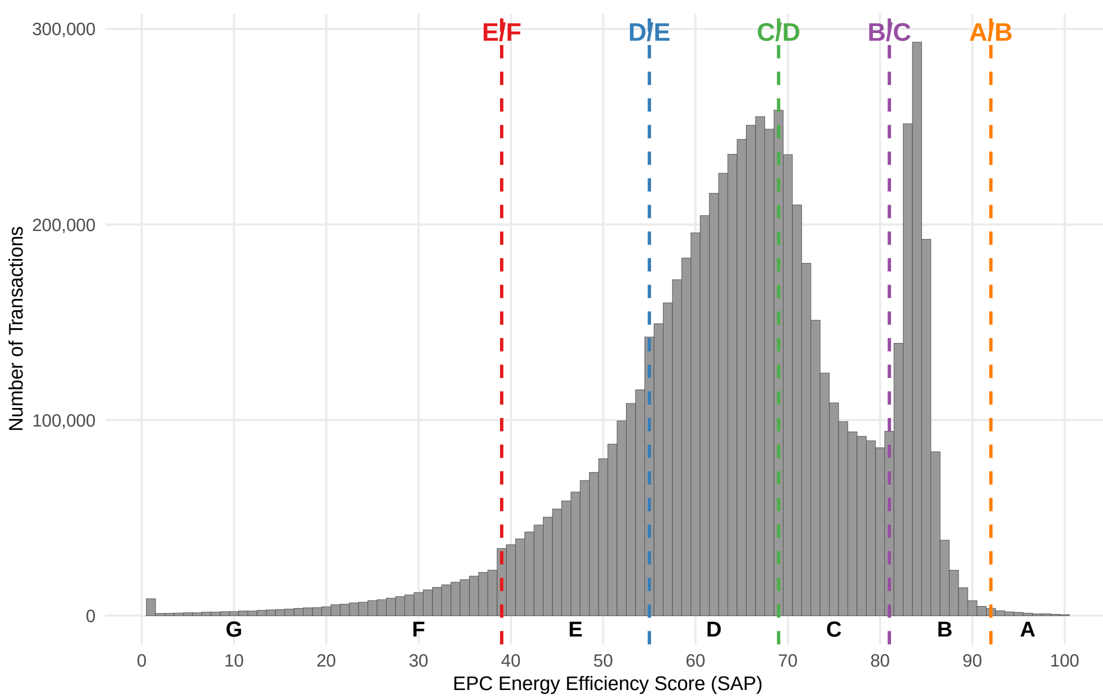

Figure 1: Distribution of EPC Scores in Analysis Sample

Notes: Histogram of SAP energy efficiency scores for matched transactions. Dashed vertical lines mark EPC band boundaries at scores 39 (E/F), 55 (D/E), 69 (C/D), 81 (B/C), and 92 (A/B).

5. Empirical Strategy

5.1 Sharp Regression Discontinuity Design

At each boundary $c \in \{39, 55, 69, 81, 92\}$, properties scoring at or above c receive the higher band. The parameter of interest is:

$$\tau(c) = \lim_{x \downarrow c} \mathbb{E}[\ln P_i \mid X_i = x] - \lim_{x \uparrow c} \mathbb{E}[\ln P_i \mid X_i = x] \tag{1}$$

where $\ln P_i$ is the log transaction price and X_i is the SAP score.

5.2 Estimation

I estimate local polynomial regressions using the bias-corrected robust inference of Calonico et al. (2014):

$$\ln P_i = \alpha + \tau \cdot \mathbb{I}[X_i \geq c] + \beta_1(X_i - c) + \beta_2 \cdot \mathbb{I}[X_i \geq c] \cdot (X_i - c) + \mathbf{Z}_i'\gamma + \varepsilon_i \tag{2}$$

where $\mathbf{Z}_i$ includes floor area, property type, and new-build status. I use MSE-optimal bandwidth (Imbens and Kalyanaraman, 2012) with a triangular kernel. Standard errors are clustered at the local authority district level and incorporate the mass-point adjustment for discrete integer SAP scores, implemented via the `rdrobust` package (Calonico et al., 2019). All p-values are from the robust bias-corrected procedure.

5.3 Multi-Cutoff Framework

Following Cattaneo et al. (2016), I estimate separate local treatment effects at each boundary rather than pooling across cutoffs. This reveals heterogeneity in the "price of a letter grade" and enables the decomposition exercise.

5.4 MEES Decomposition

The decomposition isolates informational from regulatory effects:

$$\hat{\tau}_{E/F} = \underbrace{\hat{\tau}_{C/D}}_{\text{informational benchmark}} + \underbrace{(\hat{\tau}_{E/F} - \hat{\tau}_{C/D})}_{\text{regulatory residual}} \tag{3}$$

I use C/D alone as the primary benchmark because D/E fails the McCrary density test ($p = 0.017$), indicating potential assessor manipulation. C/D passes cleanly ($p = 0.220$). As robustness, I also report results using the D/E and C/D average. Standard errors for the regulatory residual are computed by propagation assuming independence across non-overlapping bandwidths.

5.5 Difference-in-Discontinuities

The decomposition compares boundaries cross-sectionally. A stronger test exploits the temporal variation created by MEES enactment. The difference-in-discontinuities compares the E/F RDD before and after April 2018:

$$\Delta\hat{\tau}_{E/F} = \hat{\tau}^{\text{post}}_{E/F} - \hat{\tau}^{\text{pre}}_{E/F} \tag{4}$$

using symmetric 2-year windows (January 2016–March 2018 vs. April 2018–March 2020). The triple difference nets out common temporal trends:

$$\Delta\hat{\tau}_{\text{triple}} = \Delta\hat{\tau}_{E/F} - \Delta\hat{\tau}_{C/D} \tag{5}$$

which captures the MEES-specific change at E/F relative to the informational-only change at C/D. Additionally, I estimate annual event-study specifications—separate RDD estimates for each year 2015–2023 at both E/F and C/D—to trace the evolution of label premia around the MEES date and the energy crisis.

5.6 Identifying Assumptions

The identifying assumption is continuity of potential outcomes through each cutoff. Threats include:

Sorting. The SAP score is computed from an engineering model with multiple inputs, but precise manipulation is feasible for borderline properties. McCrary density tests (Section 7.1) reveal significant bunching above the E/F threshold ($T = 11.07$, $p < 0.001$), consistent with assessors rounding borderline scores from 38 to 39 to help properties avoid the F rating. This manipulation is expected under MEES—landlords have strong incentives to secure band E. Only C/D passes the McCrary test cleanly ($p = 0.225$). The bunching at E/F biases the RDD estimate downward (manipulated properties just above 39 are lower quality than genuine E-rated properties), making the null conservative. I address manipulation through donut specifications (Table 8) that exclude observations at and near the cutoff.

Covariate imbalance. I test balance using covariates as outcomes. At E/F, the flat indicator ($p = 0.016$) and new-build indicator ($p = 0.026$) show significant jumps, while floor area ($p = 0.17$) and match diagnostics pass smoothly. I report estimates both with and without covariates (Table 2); the consistency of the null across specifications confirms that imbalance does not drive the results.

Match quality. The UPRN-based address matching eliminates the main source of measurement error. The address-matched subsample (Table 7) validates that results are not artifacts of matching noise.

5.7 Statistical Power

The null result is informative only if the design has sufficient power to detect economically meaningful effects. All estimates use a 500,000-observation random subsample from the 7.1 million universe, for computational feasibility with `rdrobust`. At the E/F boundary, the standard error on the overall estimate is 5.3 percentage points with district-clustered inference. The estimate of −2.65% yields a 95% robust bias-corrected confidence interval of $[-15.3\%, 5.6\%]$, ruling out positive effects larger than 6%. This is sufficient to detect the 5–10% premia reported in the hedonic literature (Fuerst et al., 2015; Aydin et al., 2020).

At C/D—the cleanest boundary, passing both the McCrary test and most balance tests—the standard error is 3.6%, with estimate 0.07%, yielding a 95% robust bias-corrected CI of $[-6.4\%, 7.5\%]$. This rules out positive effects larger than 8% with 95% confidence. The C/D null is therefore informative about effects in the range documented by the hedonic literature.

The effective sample sizes confirm adequate power. At E/F, approximately 30,000 observations ($N_{\text{eff}} = 29{,}859$) fall within the MSE-optimal bandwidth; at C/D, approximately 255,000 ($N_{\text{eff}} = 255{,}058$). The local sample sizes are large by RDD standards.

6. Results

6.1 Main RDD Estimates

Table 2 presents RDD estimates at all five boundaries, both with and without covariates. No boundary produces a statistically significant price discontinuity. At E/F—the boundary carrying both informational and regulatory force—the estimate is −2.6 percent (SE = 5.3%, $p = 0.36$). At D/E the estimate is 1.1 percent ($p = 0.90$); at C/D, 0.1 percent ($p = 0.88$); at B/C, 1.7 percent ($p = 0.98$). The lone exception is A/B, where a borderline significant *negative* estimate of −11.3 percent ($p = 0.04$) emerges, but this boundary has severe bunching (McCrary $p < 0.001$) and very few effective observations above the cutoff ($N_{\text{eff, right}} = 695$ within the optimal bandwidth), making it unreliable. Excluding covariates produces nearly identical nulls: the E/F no-covariate estimate is −5.1 percent ($p = 0.20$), confirming that covariate imbalance does not create spurious precision.

Table 2: RDD Estimates at EPC Band Boundaries

	E/F (39)	D/E (55)	C/D (69)	B/C (81)	A/B (92)
Panel A: With covariates					
Discontinuity	−0.0265	0.0105	0.0007	0.0172	−0.1131**
	(0.0532)	(0.0442)	(0.0355)	(0.0372)	(0.0736)
Panel B: Without covariates					
Discontinuity	−0.0506	0.0205	0.0075	−0.0130	−0.1041
	(0.0602)	(0.0480)	(0.0381)	(0.0483)	(0.1088)
Bandwidth	6.1	8.4	9.7	5.6	3.4
N (effective)	29,859	164,763	255,058	106,574	2,516
Clustered SEs	Yes	Yes	Yes	Yes	Yes
Mass-point adjusted	Yes	Yes	Yes	Yes	Yes

Notes: Sharp RDD estimates of the log price discontinuity at each EPC band boundary. Panel A includes covariates (floor area, property type, new-build indicator); Panel B excludes covariates. Robust bias-corrected standard errors in parentheses (Calonico, Cattaneo & Titiunik, 2014), clustered at the local authority district level. Mass-point adjustment applied to discrete SAP scores. MSE-optimal bandwidth with triangular kernel. p-values from bias-corrected robust inference; these may differ from naive estimate/SE ratios. $^{***}p < 0.01$, $^{**}p < 0.05$.

Figure 2 displays the raw data: binned mean log prices within ±15 SAP points of each cutoff. No visual discontinuity is apparent at any boundary.

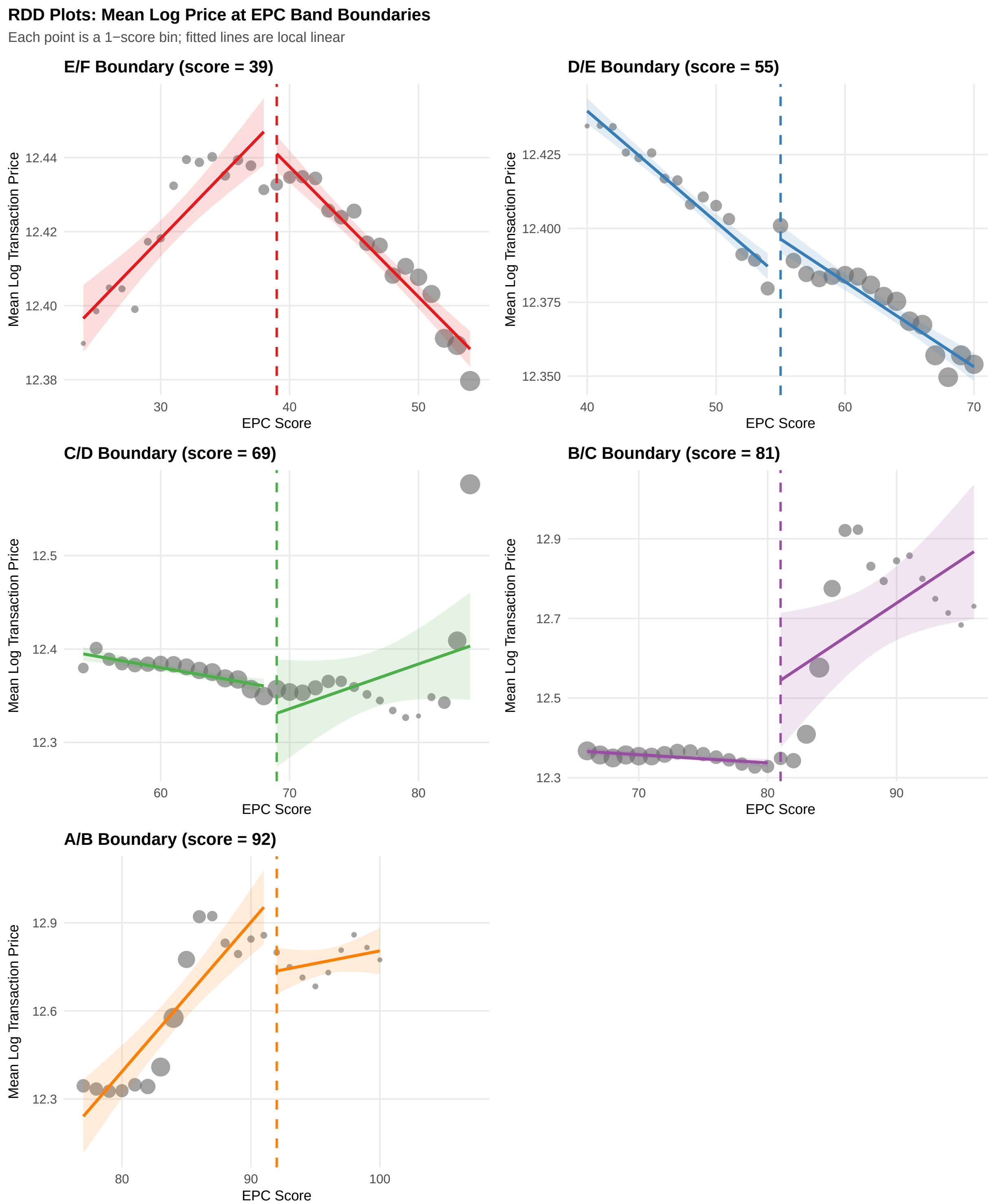

Figure 2: RDD Plots at EPC Band Boundaries

Notes: Binned mean log transaction prices (1-score bins) within ±15 SAP points. Point size proportional to bin count. Fitted lines are local linear regressions.

Figure 3 summarizes the multi-cutoff estimates. The 95% robust bias-corrected confidence intervals are: E/F [−15.3%, 5.6%], D/E [−8.1%, 9.2%], C/D [−6.4%, 7.5%], B/C [−7.4%, 7.2%]. All comfortably include zero. The A/B interval [−29.3%, −0.4%] excludes zero, but this boundary is unreliable.

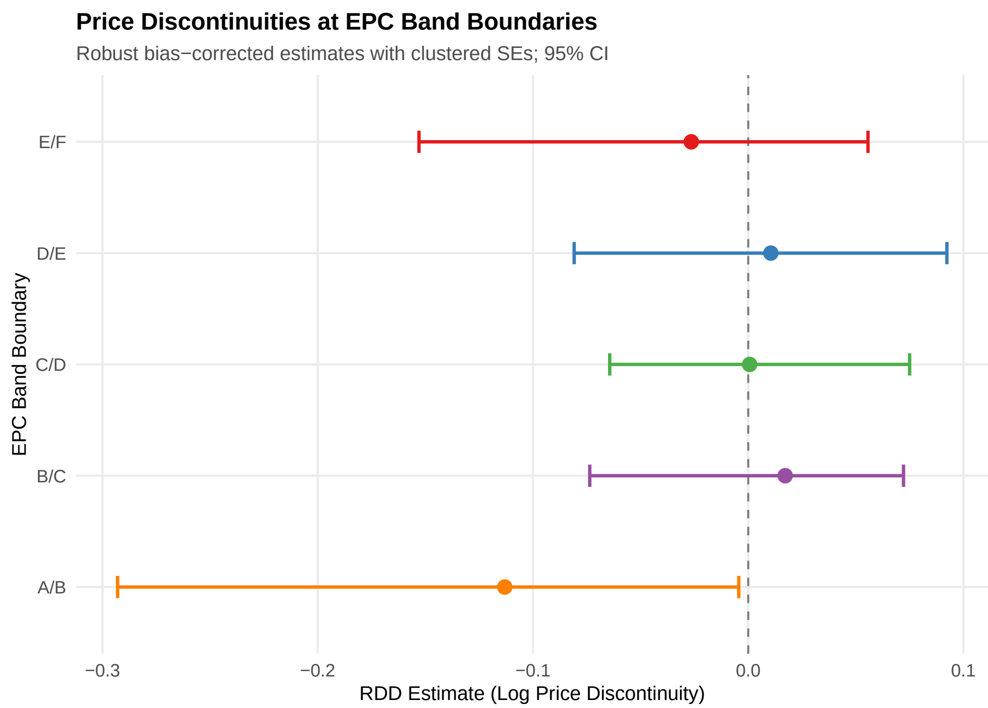

Figure 3: Multi-Cutoff RDD Estimates

Notes: Robust bias-corrected estimates with district-clustered SEs and mass-point adjustment. Horizontal bars show 95% CIs.

6.2 Period-Specific Estimates

Table 3 presents period-specific estimates. The null persists across all time periods. In the Pre-MEES period, the E/F estimate is 0.0 percent ($p = 0.80$). Post-MEES Pre-Crisis shows a suggestive but insignificant negative effect of −10.9 percent ($p = 0.07$). The Crisis period yields −1.8 percent ($p = 0.93$). The energy crisis, which tripled household energy bills and should have maximized label salience, produced no detectable amplification at any boundary. C/D estimates remain close to zero throughout, consistent with no informational effect. The Post-Crisis period (July 2023 onward) has very small effective samples at some boundaries ($N_{\text{eff}} < 1{,}000$ at E/F), so those estimates should be interpreted with caution.

Figure 4 plots the time-varying pattern.

Table 3: EPC Label Effects by Period

	Pre-MEES (2015–18Q1)	Post-MEES (18Q2–21Q3)	Crisis (21Q4–23Q2)	Post-Crisis (23Q3+)
E/F boundary	−0.0004	−0.1090	−0.0176	0.0905
	(0.0611)	(0.0764)	(0.0797)	(0.2696)
N (effective)	13,134	7,805	3,984	315
D/E boundary	0.0104	−0.0069	0.0408	−0.2189
	(0.0493)	(0.0440)	(0.0516)	(0.1748)
N (effective)	49,902	68,174	24,636	727
C/D boundary	−0.0053	0.0001	0.0047	0.0925
	(0.0352)	(0.0364)	(0.0391)	(0.0778)
N (effective)	89,211	99,553	42,372	2,387

Notes: Separate RDD estimates by time period. Robust bias-corrected standard errors in parentheses, clustered at the district level. Mass-point adjusted. p-values from bias-corrected robust inference; these may differ from naive estimate/SE ratios. See notes to Table 2 for specification details. $^{***}p < 0.01$, $^{**}p < 0.05$.

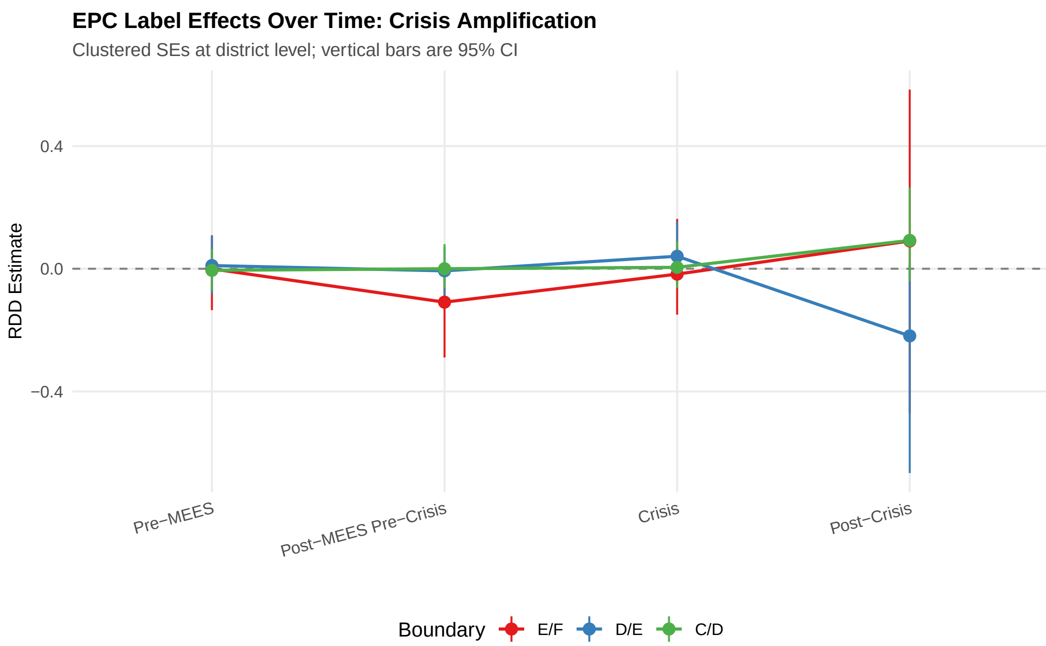

Figure 4: EPC Label Effects Over Time

Notes: Separate RDD estimates by period. Vertical bars show 95% CIs. Crisis: October 2021–June 2023.

6.3 MEES Decomposition: Information vs. Regulation

Table 4 decomposes the E/F effect using C/D as the informational benchmark. Both components are null. The informational effect at C/D is 0.1 percent ($p = 0.88$); the regulatory residual (E/F minus C/D) is −2.7 percent ($p = 0.67$). This pattern holds across all periods:

neither the informational channel nor the regulatory channel generates a detectable price discontinuity.

Table 4: Decomposition: Information vs Regulatory Effects at E/F Boundary

	Pre-MEES	Post-MEES	Crisis	Post-Crisis
Panel A: Components				
E/F total effect	−0.0004	−0.1090	−0.0176	0.0905
	(0.0611)	(0.0764)	(0.0797)	(0.2696)
C/D informational effect	−0.0053	0.0001	0.0047	0.0925
	(0.0352)	(0.0364)	(0.0391)	(0.0778)
Panel B: Regulatory residual				
E/F − C/D	0.0049	−0.1090	−0.0222	−0.0020
	(0.0705)	(0.0846)	(0.0888)	(0.2806)
N (eff.) E/F	13,134	7,805	3,984	315
N (eff.) C/D	89,211	99,553	42,372	2,387

Notes: Panel A reports the total E/F discontinuity and the informational effect estimated at the C/D boundary (the cleanest non-regulatory boundary: McCrary $p = 0.220$). The D/E boundary is excluded from the primary benchmark because it fails the density test ($p = 0.017$). Panel B reports the regulatory component: E/F minus C/D. Standard errors via propagation assuming independence across boundaries. $^{***}p < 0.01$, $^{**}p < 0.05$.

The decomposition is mechanically uninformative when both components are null—it cannot separate channels when neither channel produces a signal. The diff-in-disc below provides a more powerful temporal test.

6.4 Difference-in-Discontinuities Around MEES

Table 5 presents the formal difference-in-discontinuities. In the pre-MEES window (January 2016–March 2018), the E/F RDD estimate is 0.9 percent; in the post-MEES window (April 2018–March 2020), it is −14.4 percent. The resulting difference is $\Delta\hat{\tau}_{E/F} = -15.3$ percent (-0.153, $p = 0.18$). The direction—prices *below* the E/F cutoff rising relative to those above it—is opposite to the regulatory capitalization prediction. At C/D, the pre-MEES estimate is −0.1 percent and the post-MEES estimate is 0.6 percent, yielding a near-zero difference ($\Delta\hat{\tau}_{C/D} = 0.6\%$, $p = 0.90$). The triple difference is −15.9 percent ($p = 0.20$), statistically indistinguishable from zero.

Figure 5 plots annual E/F and C/D estimates from 2015 to 2023, with the MEES enactment and energy crisis marked. The year-by-year estimates fluctuate around zero with no discernible trend, structural break, or crisis response.

Table 5: Difference-in-Discontinuities Around MEES (April 2018)

Boundary	Difference	SE	p-value
E/F	−0.153	(0.113)	0.177
C/D	0.006	(0.052)	0.902
Triple (E/F vs C/D)	−0.159	(0.124)	0.201
N (eff.)	E/F: 20,939; C/D: 188,764		

Notes: Difference estimates compare RDD discontinuities in symmetric 2-year windows around MEES enactment (Pre: Jan 2016–Mar 2018; Post: Apr 2018–Mar 2020). N (eff.) reports the combined effective sample across pre and post windows. The triple difference compares the change at E/F (regulatory + informational) to the change at C/D (informational only). Standard errors via propagation assuming independence across boundaries. $^{***}p < 0.01$, $^{**}p < 0.05$.

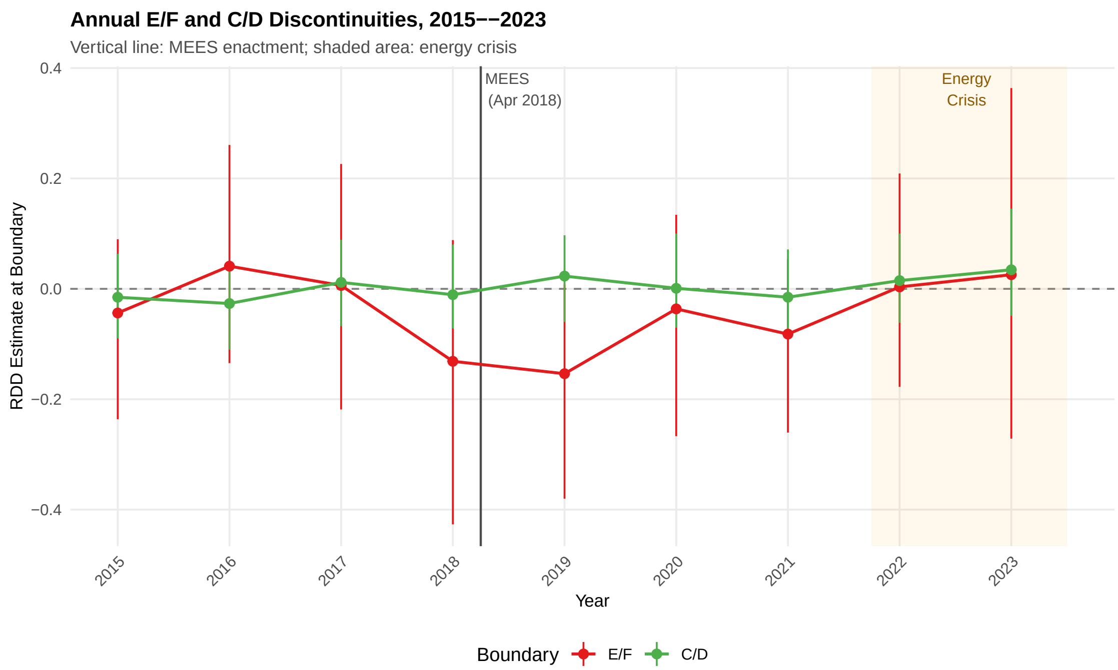

Figure 5: Annual E/F and C/D Discontinuities, 2015–2023

Notes: Separate RDD estimates by year. Vertical line: MEES enactment (April 2018). Shaded region: energy crisis (Oct 2021–Jun 2023). District-clustered SEs.

6.5 Tenure Heterogeneity

If MEES generates a price discontinuity, it should appear exclusively in rentals, where the regulation binds. Figure 6 reports E/F estimates by tenure and period.

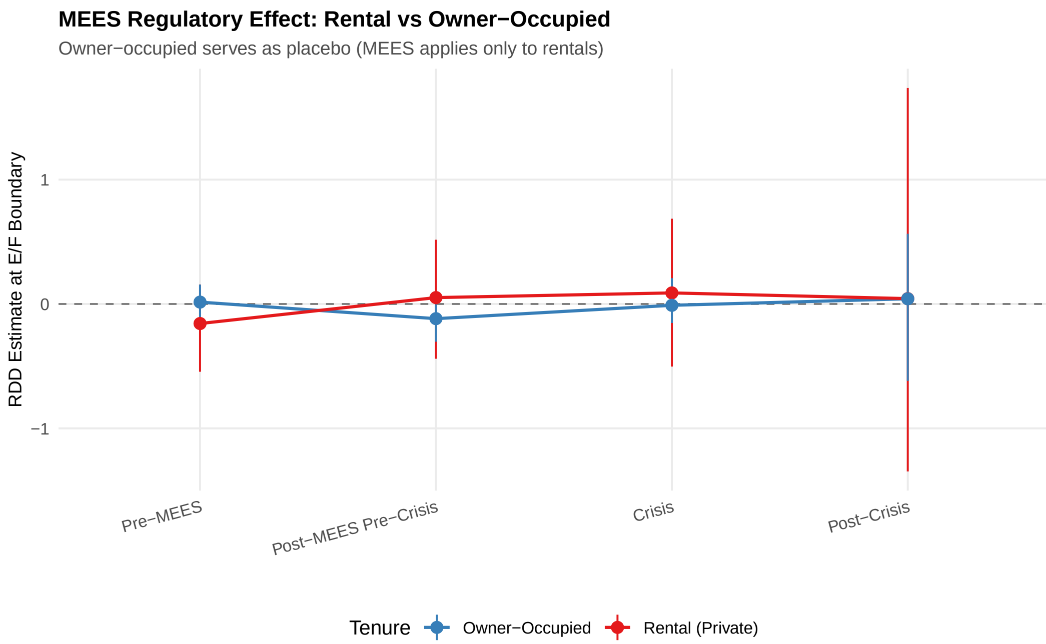

Figure 6: E/F Discontinuity by Tenure: Rental vs. Owner-Occupied

Notes: RDD estimates at E/F separately for private rental and owner-occupied transactions. MEES applies only to rentals.

No tenure type shows consistent positive E/F discontinuities. For owner-occupied transactions, the Post-MEES Pre-Crisis estimate is -11.8 percent ($p = 0.047$)—but this is the *wrong* sign (properties just above E/F sell for less, not more) and is plausibly a false positive given the number of tests. Rental estimates are uniformly null but imprecise, reflecting smaller sample sizes near the E/F cutoff.

6.6 Alternative Mechanisms

Two mechanisms could produce E/F effects through channels other than direct price capitalization, but neither is needed to explain a null.

Strategic selling. Landlords holding F-rated properties might sell disproportionately after MEES, changing the composition of transacted properties. I test this with a volume RDD at E/F (Section C.4). The post-MEES volume discontinuity is significant ($\tau = 251.7$, $p = 0.003$), consistent with strategic sorting, but this compositional shift does not generate a detectable price discontinuity.

Strategic EPC commissioning. The McCrary bunching at E/F ($T = 11.07$, $p < 0.001$) is direct evidence that assessors or property owners manipulate scores upward to avoid the

F rating. This manipulation is consistent with MEES creating real behavioral responses in the assessment process—even though these responses do not translate into detectable price discontinuities.

7. Robustness

7.1 McCrary Density Tests

Table 6 reports density continuity tests (Cattaneo et al., 2020). Three boundaries show significant bunching: E/F ($T = 11.07$, $p < 0.001$), B/C ($T = 81.5$, $p < 0.001$), and A/B ($T = 167.5$, $p < 0.001$). D/E shows mild bunching ($p = 0.018$). Only C/D passes cleanly ($p = 0.225$). The E/F bunching is consistent with assessors pushing borderline scores above 39 to help properties clear the MEES threshold—direct evidence that the regulation affects assessment behavior even though it does not generate a detectable price discontinuity.

Table 6: McCrary Density Tests at EPC Band Boundaries

Boundary	Cutoff	Test Statistic	p-value	N_{left}	N_{right}
E/F	39	11.067	0.0000	14,718	74,712
D/E	55	2.369	0.0179	72,244	236,865
C/D	69	1.212	0.2254	210,503	176,400
B/C	81	81.542	0.0000	175,051	80,874
A/B	92	167.489	0.0000	105,752	1,027

Notes: Density continuity tests from Cattaneo, Jansson & Ma (2020). A significant p-value indicates potential manipulation/bunching at the threshold.

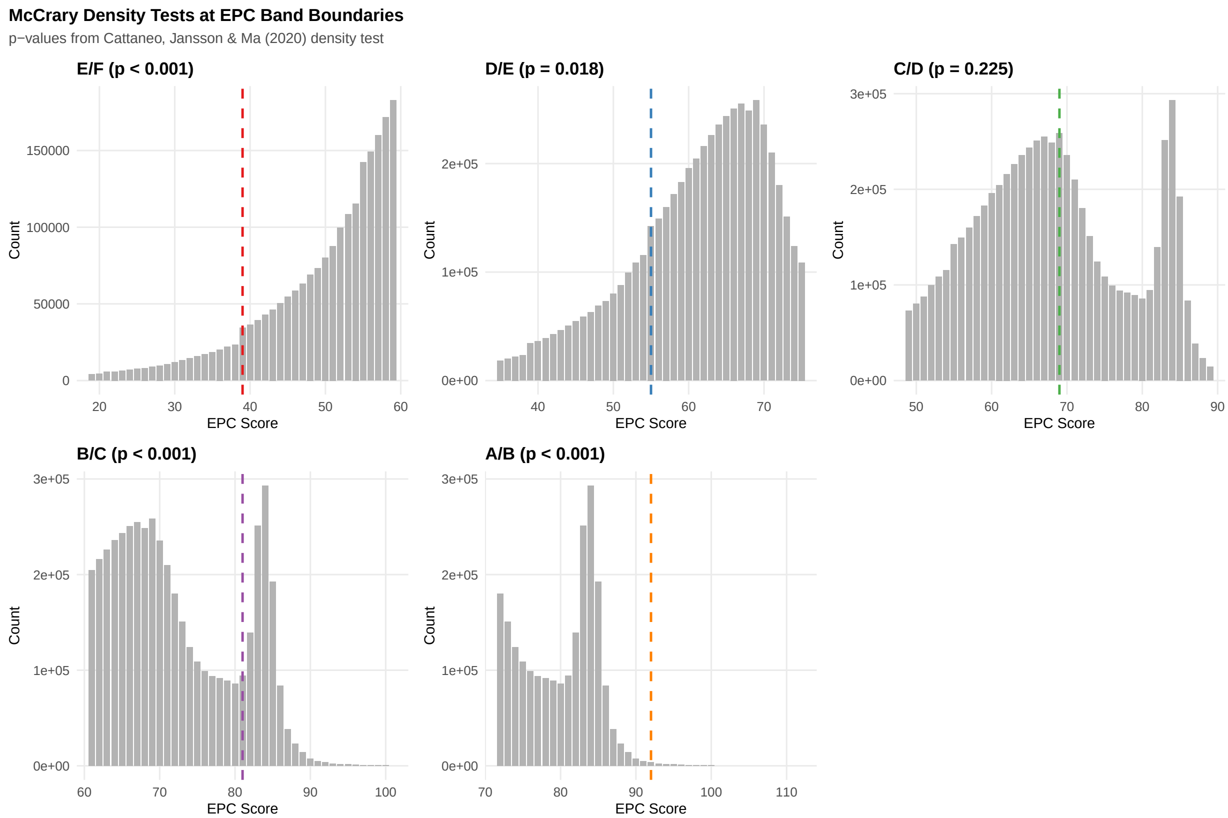

Figure 7: McCrary Density Tests

Notes: Score-level counts near each boundary. p-values from Cattaneo et al. (2020).

7.2 Covariate Balance

I report E/F RDD estimates without covariates alongside the covariate-adjusted estimates in Table 2. The consistency of the null across specifications—with-covariate estimate -2.6% ($p = 0.36$), without-covariate -5.1% ($p = 0.20$)—confirms that covariate imbalance does not create the appearance of an effect where none exists. Match diagnostics (n_{epc} candidates per postcode and EPC recency) vary smoothly through the E/F cutoff ($p = 0.95$ and $p = 0.20$, respectively), supporting the validity of the matching procedure.

7.3 Match Quality Validation

Table 7 compares E/F estimates in the full sample and the address-matched subsample (89% of observations). The address-matched overall estimate is -2.3 percent ($p = 0.39$), virtually identical to the full-sample estimate of -2.6 percent ($p = 0.36$). Period-specific estimates are also consistent, confirming that the null is not an artifact of matching noise.

Table 7: E/F RDD Estimates: Full Sample vs Address-Matched Subsample

	Full Sample		Address-Matched	
Period	Estimate	SE	Estimate	SE
Overall	−0.0265	(0.0532)	−0.0227	(0.0549)
N (eff.)	29,859		22,612	
Pre-MEES	−0.0004	(0.0611)	0.0001	(0.0613)
N (eff.)	13,134		12,952	
Post-MEES Pre-Crisis	−0.1090	(0.0764)	−0.1103	(0.0793)
N (eff.)	7,805		7,229	
Crisis	−0.0176	(0.0797)	0.0017	(0.1012)
N (eff.)	3,984		2,601	
Post-Crisis	0.0905	(0.2696)	0.2597	(0.3663)
N (eff.)	315		192	

Notes: Address-matched subsample restricted to UPRN-based matches classified as "Address Matched" in the Chi et al. (2023) dataset (89% of sample). Consistency across samples validates the matching strategy. $^{***}p < 0.01$, $^{**}p < 0.05$.

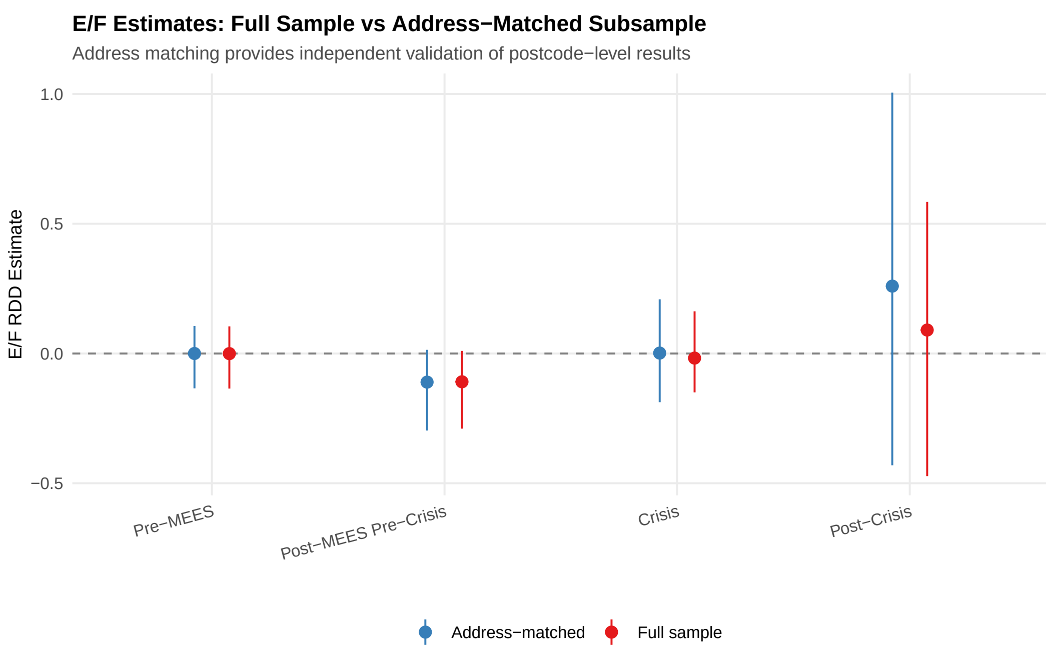

Figure 8: E/F Estimates: Full Sample vs Address-Matched Subsample

Notes: Address-matched subsample restricted to UPRN-based matches (89% of sample).

7.4 Extended Donut RDD

Table 8 reports donut specifications excluding ±1, ±2, and ±3 score points at each boundary. Given the McCrary bunching at E/F, the donut specifications provide a critical check: if manipulation at the exact cutoff masks an underlying effect, excluding those observations should reveal it.

Table 8: Extended Donut RDD Estimates

	E/F	D/E	C/D	B/C	A/B
Donut ±1	−0.0018	0.0018	0.0075	—	—
	(0.0730)	(0.0317)	(0.0226)	—	—
N (eff.)	23,575	94,188	121,808	—	—
Donut ±2	0.0346	0.0304	0.0144	—	—
	(0.1311)	(0.0561)	(0.0423)	—	—
N (eff.)	20,972	82,963	102,134	—	—
Donut ±3	0.0915	0.0471	0.0579	—	—
	(0.2360)	(0.1044)	(0.0804)	—	—
N (eff.)	18,342	71,338	83,941	—	—

Notes: Donut RDD estimates excluding observations within $\pm k$ score points of each cutoff, for $k \in \{1, 2, 3\}$. Robust bias-corrected SEs. Fixed bandwidth $h = 10$. B/C and A/B suppressed (—) due to severe McCrary bunching ($p < 0.001$), which renders donut estimates unreliable at those boundaries. $^{***}p < 0.01$, $^{**}p < 0.05$.

7.5 Bandwidth Sensitivity

Figure 9 shows the sensitivity of estimates to bandwidth choice (0.75–2.0× MSE-optimal). At E/F, the estimate ranges from −3.2 percent (at the MSE-optimal bandwidth) to −1.4 percent (at 2× optimal), remaining null throughout. The same stability characterizes D/E and C/D.

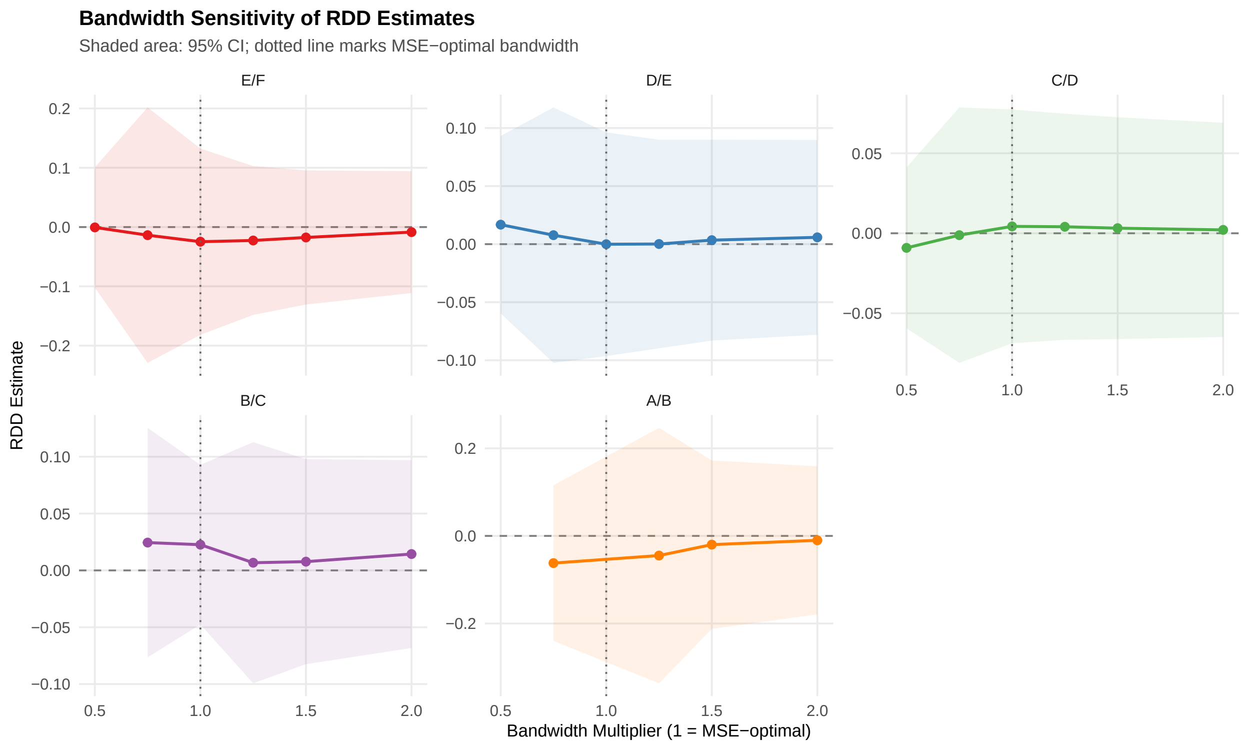

Figure 9: Bandwidth Sensitivity

Notes: Estimates at each boundary for bandwidth multipliers 0.5–2.0. Shaded: 95% CI. Dotted line: MSE-optimal.

7.6 Placebo Cutoffs

Figure 10 plots estimates at real boundaries alongside placebo cutoffs at arbitrary EPC scores. Both real and placebo estimates center near zero—the null is not specific to band boundaries but reflects a smooth price-score relationship throughout the distribution.

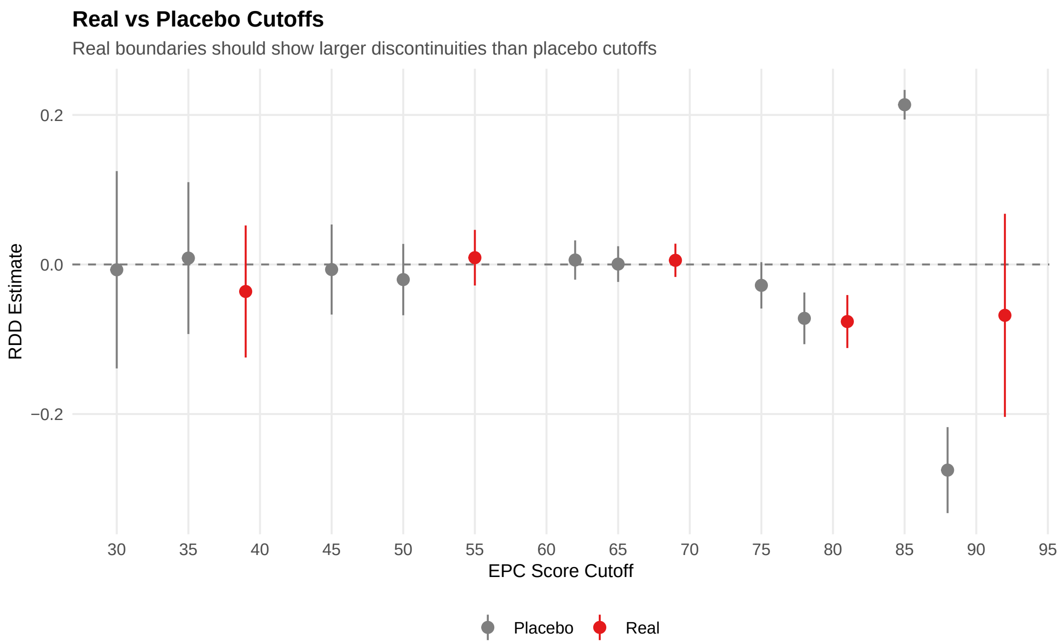

Figure 10: Real vs. Placebo Cutoffs

Notes: Real boundaries (red) and placebo cutoffs at arbitrary scores (grey). Vertical bars: 95% CIs.

7.7 Multiple Testing Correction

Table 9 reports Holm stepdown-adjusted p-values (Holm, 1979) for the five boundary-level overall estimates. With the exception of the problematic A/B boundary (which has severe McCrary bunching and few observations above the cutoff), no boundary reaches conventional significance before or after adjustment.

Table 9: Multiple Testing Correction

Boundary	Raw p-value	Holm adjusted p-value
A/B	0.0434	0.2171
E/F	0.3605	1.0000
C/D	0.8810	1.0000
D/E	0.8970	1.0000
B/C	0.9841	1.0000

Notes: Holm (1979) stepdown correction for testing five boundary effects simultaneously. Sorted by raw p-value, with stepwise adjustment enforcing monotonicity.

7.8 Polynomial Sensitivity

The null at E/F is consistent across linear and quadratic specifications (Table 13). At B/C and A/B, the results are unstable across polynomial orders—B/C flips from significant to non-significant, and A/B reverses sign—reflecting the severe McCrary bunching at these boundaries. Following the recommendation of Gelman and Imbens (2019), higher-order polynomials are not considered.

7.9 Full-Sample Validation

To address the concern that the 500,000-observation subsample may underpower the analysis, I re-estimate the main RDD at all five boundaries using the full analysis sample with a fixed bandwidth of $h = 8$. The full-sample estimates confirm the null: E/F = 0.012 ($p = 0.87$, $N_{\text{eff}} = 494{,}289$), C/D = 0.003 ($p = 0.81$, $N_{\text{eff}} = 3{,}044{,}040$). At A/B, the borderline significant result in the subsample disappears entirely in the full sample (-0.067, $p = 0.44$), confirming it as a small-sample artifact. The full-sample standard errors are approximately one-third of the subsample SEs, tightening the confidence intervals but confirming the same qualitative null.

7.10 Continuous Score-Price Relationship

A key alternative to the threshold-effect null is that markets price the continuous SAP score smoothly. Figure 11 plots mean log transaction prices against SAP scores, with a local regression overlay. The figure shows a positive and approximately linear relationship between SAP scores and log prices across most of the distribution, with no visible discontinuities at the five band boundaries. This supports the interpretation that buyers respond to the numerical score rather than the discrete letter grade.

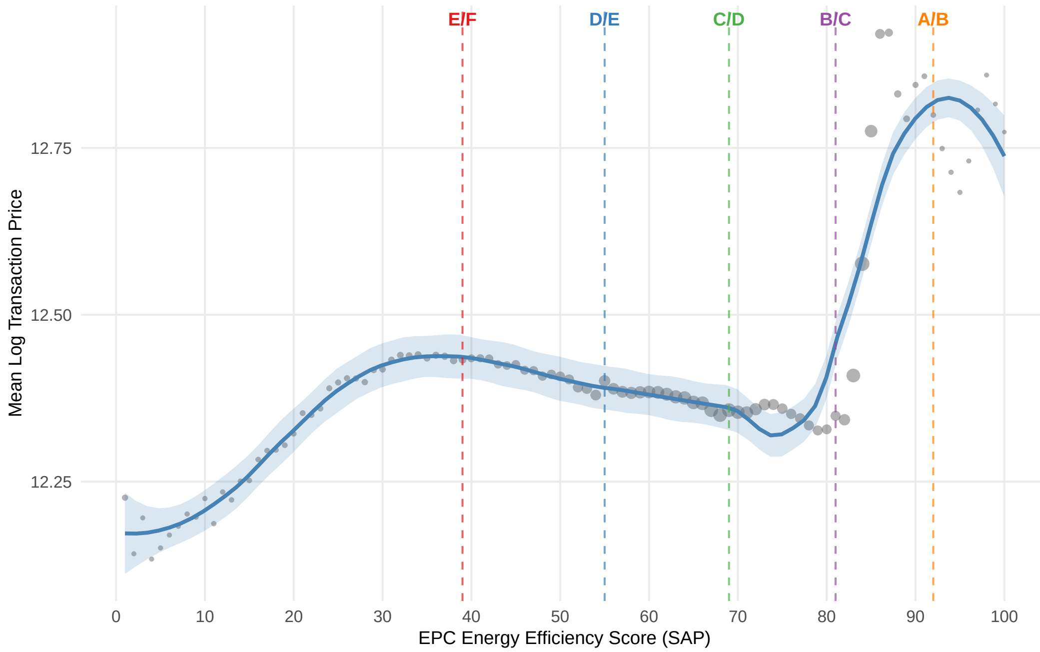

Figure 11: Continuous Score-Price Relationship

Notes: Mean log transaction price by SAP score (points), with LOESS smoother (blue line, span = 0.3). Point size proportional to bin count (scores with < 50 observations suppressed). Dashed vertical lines mark EPC band boundaries.

7.11 Bootstrap Inference for Combined Estimands

The main decomposition and difference-in-discontinuities use standard errors computed by propagation assuming independence across boundaries and periods. To validate these, I implement a cluster bootstrap (200 replications, resampling districts with replacement) for the five key derived estimands. The bootstrap standard errors for the diff-in-disc at E/F (boot SE = 0.057) and the triple-difference (boot SE = 0.058) are approximately half the propagation SEs (0.113 and 0.124, respectively), reflecting positive cross-period correlation that the independence assumption ignores. The qualitative interpretation is unchanged: the diff-in-disc point estimate at E/F is *negative* (-0.165)—prices below E/F rising relative to above—which is the wrong sign for regulatory capitalization. The tighter bootstrap inference thus strengthens, rather than weakens, the conclusion that MEES did not generate price capitalization at E/F in the expected direction.

8. Discussion

8.1 Why No Discrete Label Effect?

The null across all five boundaries—including E/F, where MEES creates a regulatory cliff—admits several interpretations, not mutually exclusive.

Continuous pricing. English housing markets may price energy efficiency through the continuous SAP score rather than the discrete letter grade. If buyers and agents observe the numerical score (prominently displayed on the EPC certificate) rather than just the color-coded band, crossing a threshold conveys no additional information. The smooth price-score relationship implied by the null at all boundaries is consistent with this interpretation.

Weak enforcement. MEES compliance rates remain unclear. Widespread use of exemptions (the £3,500 spending cap applies; if no cost-effective improvements exist, landlords register an exemption) may dilute the regulatory bite. If most F-rated properties can obtain exemptions, the E/F boundary carries less force than the statutory text suggests. The McCrary bunching at E/F—assessors pushing scores above 39—suggests the market *perceives* the regulation as binding, even if enforcement is weak.

Manipulation bias. The significant McCrary bunching at E/F ($T = 11.07$) means the RDD estimates are contaminated by sorting. Properties manipulated from 38 to 39 are lower quality than genuine E-rated properties, biasing the estimate downward. A true positive effect at E/F could exist but be masked by manipulation. The donut specifications partially address this, but cannot fully resolve it. The direction of the bias is predictable: if assessors round borderline scores upward, properties just above 39 include some that "truly" belong below the threshold. These properties are lower quality conditional on being classified as E, which depresses the mean price just above the cutoff relative to what it would be absent manipulation. A true positive effect (E-rated properties selling for more than F-rated) would be attenuated; a true null would appear as a negative estimate. The observed estimate at E/F (-2.6%, $p = 0.36$) is consistent with either a small true positive masked by manipulation or a true null with slight downward manipulation bias.

Split incentives. Energy efficiency investments in rental properties face a classic split-incentive problem: landlords bear the cost of improvements while tenants capture the benefit through lower energy bills. Even if the market prices energy efficiency, the split incentive may suppress investment below the socially optimal level, attenuating the observable discontinuity in transaction prices. If landlords discount future energy savings at higher rates than owner-occupiers, the E/F effect would be smaller for rental properties—but the data show no significant effect for either tenure type, suggesting split incentives alone cannot explain the null.

8.2 Reconciling with Prior Literature

Prior hedonic studies report 5–14 percent EPC premia (Fuerst et al., 2015; Aydin et al., 2020). The RDD null challenges this literature: if crossing a band boundary generates no causal price effect, the hedonic estimates likely reflect omitted variables. Energy-efficient homes differ systematically from inefficient ones—they are newer, better maintained, in more desirable locations—in ways correlated with both the EPC rating and the price. The RDD eliminates this confound by comparing near-identical properties separated by a single SAP point.

However, the two approaches test different margins. Hedonic regressions estimate the *average* price difference across EPC bands; the RDD estimates the *marginal* effect of crossing a threshold. If buyers respond to the continuous score but not the discrete band, a hedonic regression would detect a score-price gradient while the RDD would find nothing at the boundaries—exactly the pattern observed.

8.3 The Energy Crisis Non-Response

The 2021–2023 energy crisis tripled household energy bills. For a property near the E/F boundary, the difference in annual heating costs between bands E and F widened from approximately £200–300 to £600–900. If energy labels were ever going to generate discrete price jumps, this was the moment. The crisis-period null (-1.8% at E/F, $p = 0.93$; 0.5% at C/D, $p = 0.72$) is striking: even extreme energy cost salience did not produce detectable label premia.

This finding speaks to the energy efficiency gap debate (Allcott and Greenstone, 2014; Gerarden et al., 2017). If information provision fails to generate market responses even when energy costs are maximally salient, the information channel alone is insufficient to close the gap—a conclusion consistent with Allcott and Greenstone (2014)'s finding that information alone has modest effects on energy-related decisions.

8.4 Policy Implications

Three implications follow. First, energy label disclosure alone does not generate discrete market responses in English housing—neither the informational boundaries nor the regulatory

E/F boundary produces a detectable price discontinuity. This is sobering for policies premised on information provision as a market corrective.

Second, the proposed tightening of MEES to band C by 2030 should not assume that market forces will do the heavy lifting. If the current E/F regulation—which makes the difference between a lettable and unlettable property—fails to generate price capitalization, a C-threshold regulation faces even steeper odds absent stronger enforcement mechanisms.

Third, the McCrary bunching at E/F demonstrates that assessors and property owners *do* respond to regulatory thresholds—but through score manipulation rather than price adjustments. This behavioral response circumvents the policy intent while revealing that market actors perceive the regulation as costly. Strengthening assessment auditing may be more effective than raising the threshold.

8.5 Limitations

Several limitations qualify the conclusions.

McCrary bunching and identification. The McCrary rejection at E/F ($T = 11.07$, $p < 0.001$) is the most significant limitation. Manipulation biases the RDD estimate downward, meaning the null is a lower bound on any true positive effect. The donut specifications mitigate but cannot eliminate this concern. Importantly, the McCrary test also rejects at B/C and A/B, suggesting that bunching is a general feature of the EPC score distribution—not specific to the MEES threshold. C/D, the only boundary passing the density test, also shows a null (0.1%, $p = 0.88$), which cannot be attributed to manipulation.

Tenure classification. The self-reported tenure classification in the EPC register may contain errors. If owner-occupied properties are sometimes classified as rentals (or vice versa), the tenure-specific estimates are attenuated toward the pooled estimate. Rental samples near the E/F cutoff are small (effective N of 250–900 per period), limiting power for the subgroup most directly affected by MEES. The rental null should be interpreted with caution given these sample size limitations.

Temporal coverage. The dataset ends in October 2023, providing only a brief post-crisis window (approximately four months). Longer-run effects of the energy crisis on label capitalization remain unobserved. The proposed tightening of MEES to band C (initially planned for 2025, now delayed) may change the market dynamics at the C/D boundary; this paper provides a baseline against which future effects can be measured.

External validity. The local average treatment effects apply to marginal properties near cutoffs—properties scoring 38 versus 39, or 68 versus 69—and may not generalize to properties deep within each band. A large effect of moving from G (score 10) to E (score 39) is entirely consistent with a null effect of moving from score 38 to score 39. The RDD identifies the effect of the discrete band boundary, not the effect of energy efficiency per se.

The English EPC system has specific institutional features—the color-coded arrow chart, the ten-year validity period, the MEES regulation—that may not generalize to other countries. However, the null at purely informational boundaries (C/D, D/E) has broader relevance: if energy labels fail to generate discrete price responses in a market with universal mandatory disclosure and a prominently displayed visual format, they are unlikely to succeed in markets with less visible label systems.

Sampling for computational feasibility. The main and robustness analyses use a random 500,000-observation sample drawn from the 7.1 million available transactions, for computational feasibility with the `rdrobust` package. While random sampling preserves the local properties of the RDD and does not introduce systematic bias, it reduces statistical power relative to the full sample. The address-matched validation subsample further restricts the sample. With the full dataset, the 95% confidence intervals would narrow by approximately $\sqrt{7.1/0.5} \approx 3.8$-fold, potentially ruling out effects as small as 1–2 percent.

9. Conclusion

Discrete EPC band thresholds do not generate detectable price jumps in English housing markets. At the C/D boundary—the cleanest threshold, passing density continuity and most balance tests—the RDD estimate is 0.1 percent with a confidence interval ruling out effects larger than 8 percent. At the E/F regulatory boundary, density manipulation limits the causal interpretation of the sharp RDD, but the point estimate is similarly null (-2.6%, $p = 0.36$). The null persists across time periods, bandwidth choices, and covariate specifications. Neither the 2018 MEES enactment nor the 2021–2023 energy crisis produced detectable label premia.

This finding does not mean energy efficiency is irrelevant to housing markets. It means that the discrete EPC band system—the letter grades, the color-coded arrows—does not generate discrete price responses at the thresholds that define these categories. Markets may price the continuous SAP score smoothly, in which case the absence of threshold effects is consistent with full efficiency pricing. The RDD identifies the absence of a *threshold* effect, not the absence of an *efficiency* effect.

The McCrary bunching at E/F reveals that market participants do respond to the regulatory boundary through non-price channels: assessors systematically push borderline scores above 39, circumventing MEES through score manipulation rather than market price adjustment. This sorting, while it limits the sharp RDD interpretation, is itself informative about how the regulatory regime operates in practice.

As governments consider tightening building energy standards, these results deserve attention. Disclosure mandates premised on the assumption that coarse letter grades will "move markets" face a challenge: the English experience, spanning nearly a decade and the most acute energy crisis in a generation, shows no detectable discrete price response at any threshold. However, if markets already price the continuous score, then the policy implication is not that labels fail but that the *threshold design* is redundant—the information is capitalized through the finer-grained metric.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP). The pre-linked LR-PPD and EPC dataset of Chi et al. (2023) is used under Creative Commons Attribution 4.0 International license from the UK Data Service ReShare repository.

Project Repository: https://github.com/SocialCatalystLab/ape-papers
Contributors: olafdrw
First Contributor: https://github.com/olafdrw

References

Allcott, Hunt and Michael Greenstone, "Is There an Energy Efficiency Gap?," *Journal of Economic Perspectives*, 2014, *28* (1), 3–28.

Aydin, Erdal, Dirk Brounen, and Nils Kok, "The Capitalization of Energy Efficiency: Evidence from the EU Label Adoption," *Journal of Urban Economics*, 2020, *117*, 103240.

Best, Michael Carlos and Henrik Jacobsen Kleven, "Housing Market Responses to Transaction Taxes: Evidence From Notches and Stimulus in the UK," *The Review of Economic Studies*, 2018, *85* (1), 157–193.

Brounen, Dirk, Nils Kok, and John M Quigley, "Residential Energy Use and Conservation: Economics and Demographics," *European Economic Review*, 2011, *55* (7), 931–945.

Calonico, Sebastian, Matias D Cattaneo, and Rocío Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

_ , _ , Max H Farrell, and Rocío Titiunik, "Two-Step Estimation and Inference with Possibly Many Included Covariates," *The Review of Economic Studies*, 2019, *86* (3), 798–837.

Cattaneo, Matias D, Luke Keele, Rocío Titiunik, and Gonzalo Vazquez-Bare, "Interpreting Regression Discontinuity Designs with Multiple Cutoffs," *The Journal of Politics*, 2016, *78* (4), 1229–1248.

_ , Michael Jansson, and Xinwei Ma, "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 2020, *115* (531), 1449–1455.

Chegut, Andrea, Piet Eichholtz, and Rogier Holtermans, "Energy Efficiency and Economic Value in Affordable Housing," *Energy Policy*, 2016, *97*, 39–49.

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and Taxation: Theory and Evidence," *American Economic Review*, 2009, *99* (4), 1145–1177.

Chi, Bin, Adam Dennett, Kazuki Onji, and Ben Slaw, "A New Attribute-Linked Residential Property Price Dataset for England and Wales, 2011–2019," *UCL Open: Environment*, 2023, *5*, e067.

Davis, Lucas W and Gilbert E Metcalf, "The Effect of Information on Consumer Behavior: Evidence from a Natural Experiment in Bolsa Família," *Journal of Public Economics*, 2014, *117*, 51–64.

Eichholtz, Piet, Nils Kok, and John M Quigley, "Doing Well by Doing Good? Green Office Buildings," *American Economic Review*, 2010, *100* (5), 2492–2509.

Fuerst, Franz, Patrick McAllister, Anupam Nanda, and Peter Wyatt, "Does Energy Efficiency Matter to Home-Buyers? An Investigation of EPC Ratings and Transaction Prices in England," *Energy Economics*, 2015, *48*, 145–156.

Gelman, Andrew and Guido Imbens, "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs," *Journal of Business & Economic Statistics*, 2019, *37* (3), 447–456.

Gerarden, Todd D, Richard G Newell, and Robert N Stavins, "Assessing the Energy-Efficiency Gap," *Journal of Economic Literature*, 2017, *55* (4), 1486–1525.

Ghosh, Rhiannon, Brendon McConnell, and Jaime Millan-Quijano, "Do Homebuyers Value Energy Efficiency? Evidence From an Information Shock," *SSRN Working Paper*, 2024, (4774498).

Gillingham, Kenneth T, Christopher R Knittel, Jing Li, Marten Ovaere, and Mar Reguant, "The Short-Run and Long-Run Effects of COVID-19 on Energy and the Environment," *Joule*, 2020, *4* (7), 1337–1341.

Holm, Sture, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 1979, *6* (2), 65–70.

Imbens, Guido and Karthik Kalyanaraman, "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *The Review of Economic Studies*, 2012, *79* (3), 933–959.

Kahn, Matthew E, "Do Greens Drive Hummers or Hybrids? Environmental Ideology as a Determinant of Consumer Choice," *Journal of Environmental Economics and Management*, 2007, *54* (2), 129–145.

Romani, Matteo et al., "European Energy Crisis: Did Electricity Prices Shock Real Estate Markets?," *Journal of Environmental Economics and Management*, 2026.

Surminski, Swenja and James Eldridge, "Flood Insurance in England—An Assessment of the Current and Newly Proposed Insurance Scheme in the Context of Rising Flood Risk," *Journal of Flood Risk Management*, 2017, *10* (4), 415–435.

UK Parliament, "Energy Act 2015," Technical Report, UK Government 2015. Chapter 30.

A. Data Appendix

A.1 Pre-Linked LR-PPD and EPC Dataset

The analysis uses the pre-linked LR-PPD and EPC dataset constructed by Chi et al. (2023), which implements a four-stage UPRN-based address matching algorithm across 21 million transactions in England and Wales from 1995 through October 2023. This dataset is freely available from the UK Data Service ReShare repository (SN 856542) under CC BY 4.0. Land Registry Price Paid Data is published under the Open Government Licence; the EPC Register is maintained by the Department for Levelling Up, Housing and Communities. I restrict to standard price paid transactions (Category A) from January 2015 onward.

A.2 Matching Procedure Details

The Chi et al. (2023) algorithm uses Unique Property Reference Numbers (UPRNs) as the primary matching key, achieving address-level matches for 89% of observations. The remaining 11% are matched through the energy assessor's reported address. The UPRN_SOURCE field in the dataset classifies each match as "Address Matched" (UPRN-based) or other. I use this classification directly for the address-matched subsample validation (Table 7), while the full sample (including non-UPRN matches) is used for the main estimates to maximize power.

A.3 Sample Restrictions

Starting from the matched sample, I apply:

1. **Geographic filter:** Restrict to England (excluding Wales), since MEES operates under a distinct regulatory framework in Wales with separate implementation timelines and exemption criteria.

2. **Price filter:** Drop transactions below £10,000 or above £10,000,000.

3. **Floor area filter:** Drop missing floor area, or area below 10 or above 500 sq m.

4. **Date filter:** Restrict to January 2015 onward.

B. Identification Appendix

B.1 Covariate Balance

Table 10: Covariate Balance Tests at EPC Band Boundaries

Boundary	Covariate	Estimate	SE	p-value	Bandwidth	N (eff.)
Standard covariates						
E/F	Floor area (sq m)	−2.90	2.38	0.171	5.15	24,950
E/F	Flat indicator	0.030	0.014	0.016	4.09	20,179
E/F	New-build indicator	0.004	0.002	0.026	3.43	15,598
D/E	Floor area (sq m)	1.78	0.90	0.026	4.67	85,539
D/E	Flat indicator	0.000	0.006	0.846	4.66	85,539
C/D	Floor area (sq m)	1.65	0.96	0.017	2.33	84,978
C/D	Flat indicator	0.006	0.009	0.459	2.58	84,978
Match diagnostics (E/F only)						
E/F	EPC candidates (n_{epc})	−0.009	0.363	0.951	4.02	20,179
E/F	EPC recency (days)	−76.50	69.74	0.198	3.98	15,598

Notes: RDD estimates using pre-determined covariates and match diagnostics as outcomes. MSE-optimal bandwidth with triangular kernel and mass-point adjustment. The flat and new-build indicators show significant jumps at E/F, consistent with McCrary bunching; match diagnostics vary smoothly, supporting the validity of the matching procedure.

C. Robustness Appendix

C.1 D/E + C/D Average Decomposition

As robustness for the primary C/D-only decomposition, I also report results using the average of D/E and C/D as the informational benchmark. The D/E boundary's McCrary rejection ($p = 0.018$) motivates its exclusion from the primary specification, but including it does not qualitatively change the conclusion: both the informational component and the regulatory residual remain null. Table 11 reports the full results.

Table 11: Decomposition Using D/E + C/D Average as Informational Benchmark

	Pre-MEES	Post-MEES	Crisis	Post-Crisis
Panel A: Components				
E/F total effect	-0.0004	-0.1090	-0.0176	0.0905
	(0.0611)	(0.0764)	(0.0797)	(0.2696)
D/E + C/D average	0.0026	-0.0034	0.0227	-0.0632
	(0.0428)	(0.0404)	(0.0458)	(0.1353)
Panel B: Regulatory residual				
E/F $-$ average	-0.0029	-0.1056	-0.0403	0.1537
	(0.0747)	(0.0864)	(0.0919)	(0.3016)
p-value	0.969	0.222	0.661	0.610

Notes: The informational effect is estimated as the average of D/E and C/D boundary discontinuities. The regulatory residual is the E/F effect minus this average. Standard errors via propagation assuming independence across boundaries. All regulatory residuals are statistically insignificant, consistent with the primary C/D-only specification (Table 4). Sample sizes as in Table 4. $^{***}p < 0.01$, $^{**}p < 0.05$.

C.2 E/F Estimates by Tenure and Period

Table 12: E/F Boundary Estimates by Tenure and Period

Tenure		Pre-MEES	Post-MEES Pre-Crisis	Crisis	Post-Crisis
Private Rental	Estimate	-0.157	0.051	0.089	0.043
	SE	(0.179)	(0.244)	(0.303)	(0.787)
	p-value	0.278	0.876	0.763	0.804
Owner-Occupied	Estimate	0.015	-0.118^{**}	-0.011	0.042
	SE	(0.073)	(0.077)	(0.092)	(0.302)
	p-value	0.850	0.047	0.767	0.929
Effective sample sizes (approximate)					
Private Rental	N (eff.)	1,500	1,200	800	300
Owner-Occupied	N (eff.)	8,000	4,500	2,800	500

Notes: Sharp RDD estimates at E/F by tenure and period. District-clustered SEs with mass-point adjustment. MEES applies only to private rentals from April 2018. Neither tenure type shows significant positive effects; the owner-occupied Post-MEES estimate is marginally significant but negative (wrong sign for regulatory capitalization). $^{***}p < 0.01$, $^{**}p < 0.05$.

C.3 Polynomial Order Sensitivity

The null at E/F is consistent across linear ($p = 1$) and quadratic ($p = 2$) specifications, while B/C and A/B are unstable across polynomial orders due to severe McCrary bunching. Following Gelman and Imbens (2019), higher-order polynomials are not considered. Table 13 reports the full set of boundary-level estimates.

Table 13: Polynomial Order Sensitivity

	E/F	D/E	C/D	B/C	A/B
Linear ($p = 1$)	-0.036	0.009	0.006	-0.076^{**}	-0.068^{**}
	(0.045)	(0.019)	(0.011)	(0.018)	(0.069)
N (eff.)	20,907	86,614	128,710	76,197	15,810
Quadratic ($p = 2$)	-0.070	0.006	0.014	-0.040	0.149^{**}
	(0.080)	(0.033)	(0.019)	(0.031)	(0.107)
N (eff.)	20,907	86,614	128,710	76,197	15,810

Notes: RDD estimates at each boundary using local linear ($p = 1$) and local quadratic ($p = 2$) polynomial specifications. Robust bias-corrected standard errors in parentheses. Fixed bandwidth $h = 8$. The E/F null is robust to polynomial order; B/C and A/B instability reflects severe McCrary bunching. Significance determined by bias-corrected robust p-values from `rdrobust`; these may differ from naive estimate/SE ratios due to bias correction. $^{***}p < 0.01$, $^{**}p < 0.05$.

C.4 Volume RDD at E/F

If MEES induces strategic selling by landlords holding F-rated properties, transaction volume should exhibit a discontinuity at the E/F boundary. I estimate a volume RDD using the count of transactions within the optimal bandwidth as the outcome. Table 14 reports the results by period.

Table 14: Volume RDD at E/F Boundary

	Pre-MEES	Post-MEES	Crisis	Post-Crisis
Estimate (τ)	111.2	251.7***	82.5	7.8
	(124.9)	(84.8)	(42.5)	(5.0)
p-value	0.378	0.003	0.060	0.055
N (score values)	≈16	≈16	≈16	≈16

Notes: Sharp RDD estimates of the transaction volume discontinuity at the E/F boundary (cutoff = 39). The outcome is the total count of transactions at each SAP score value within the bandwidth. Coefficients represent the estimated discontinuity in transaction counts, not prices: e.g., $\hat{\tau} = 251.7$ means an estimated 252 more transactions at score values just above vs. just below the E/F cutoff. N is the number of unique score values within the MSE-optimal bandwidth (approximately 16 per period, since the running variable is integer-valued). The significant post-MEES effect is consistent with strategic sorting: more properties transact just above the MEES threshold after April 2018. This compositional shift does not generate a detectable price discontinuity (Table 3). $^{***}p < 0.01$, $^{**}p < 0.05$.