

Shining Light on Nothing? The Null Effect of Mandatory Energy Benchmarking on NYC Property Values

APEP Autonomous Research* @ai1scl

February 13, 2026

Abstract

Governments worldwide mandate energy disclosure for buildings, expecting transparency to reshape real estate markets. I exploit a sharp regulatory threshold in New York City’s Local Law 84—which requires buildings above 25,000 square feet to publicly report energy and water consumption—to estimate the causal effect of mandatory benchmarking on property values. A first-stage compliance jump of 42.3 percentage points confirms the threshold binds. Yet the main regression discontinuity estimate on log assessed value is -0.040 (robust SE = 0.059, $p = 0.591$): a precise, well-powered zero. This null persists across bandwidths, polynomial orders, kernels, donut specifications, borough subsamples, and building cohorts. The results imply that commercial real estate markets already internalize energy performance information, and that mandatory benchmarking—whatever its other merits—does not alter how the market prices buildings.

JEL Codes: R31, Q48, D83

Keywords: energy disclosure, building benchmarking, property values, regression discontinuity, information asymmetry

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch

1. Introduction

Mandatory energy disclosure is the quiet workhorse of climate policy. Over thirty U.S. cities and a growing number of countries now require large buildings to publicly report their energy consumption, on the theory that sunlight is the best disinfectant: once buyers, tenants, and investors can compare buildings on energy performance, the market will reward efficiency and punish waste ([Joskow, 2012](#)). New York City’s Local Law 84 (LL84) is the flagship of this movement. Since 2009 it has required the city’s largest buildings to annually benchmark energy and water usage through the EPA’s ENERGY STAR Portfolio Manager and to make those results publicly available. In 2016, the law expanded to cover buildings above 25,000 gross square feet, bringing roughly 12,000 additional structures—many of them mid-rise commercial and mixed-use buildings—under the mandate.

The rationale is simple and appealing. Energy costs are a major operating expense in commercial real estate, yet buyers and tenants often lack reliable information about a building’s energy profile before committing to a purchase or lease. This information asymmetry, of the kind first formalized by [Akerlof \(1970\)](#), can suppress investment in efficiency: owners who upgrade their buildings cannot credibly signal the improvement, so the market fails to reward them. Mandatory disclosure, the argument goes, resolves this “lemons” problem. Once energy performance is public, efficient buildings trade at a premium, inefficient buildings trade at a discount, and the invisible hand redirects capital toward the greener stock ([Allcott and Greenstone, 2012](#); [Gerarden et al., 2017](#)).

But does any of this actually happen? The empirical evidence is surprisingly thin. A well-known literature documents that LEED and ENERGY STAR certifications correlate with higher rents and sale prices in commercial real estate ([Eichholtz et al., 2010, 2013](#); [Kahn and Kok, 2014](#)). A parallel literature shows that energy performance certificates affect residential prices in Europe ([Brounen and Kok, 2013](#); [Fuerst et al., 2015](#)). Yet correlation is not causation. Buildings that voluntarily pursue green certifications differ from non-certified buildings in numerous dimensions—age, location, tenant quality, management sophistication—that independently affect value. The fundamental question of whether *mandatory* disclosure changes market outcomes remains open.

This paper answers that question using a clean identification strategy. I exploit the sharp regulatory threshold at 25,000 square feet created by LL84’s 2016 expansion. Buildings with gross floor area (GFA) just above 25,000 sq ft must publicly disclose their energy and water performance; buildings just below face no such requirement. Since 2016, the 25,000 sq ft threshold has been the primary regulatory cutoff at this building size: the related Local Law 87 (mandatory energy audits) applies only at the 50,000 sq ft threshold, and Local Law 97

(building emissions caps), enacted in 2019, does not impose compliance costs until 2024. This institutional isolation is critical—the dominant discontinuity in regulatory burden at 25,000 sq ft is LL84’s disclosure requirement.

The regression discontinuity design (RDD) relies on comparing buildings just above and just below the 25,000 sq ft threshold, under the assumption that potential outcomes vary smoothly across the cutoff (Imbens and Lemieux, 2008; Lee and Lemieux, 2010). Several features of the setting support this assumption. GFA is determined by physical construction and recorded in Department of Finance property records, typically decades before LL84’s expansion to this threshold. Altering a building’s GFA requires costly construction or demolition. For the vast majority of NYC’s building stock—constructed well before 2016—the 25,000 sq ft cutoff was not a consideration at the time of construction. A McCrary density test confirms the absence of manipulation: $t = -0.010$, $p = 0.992$.

The first stage is strong and unambiguous. Unconditional compliance rates are approximately 80.5% just above the threshold and 7% just below—a 73.5 percentage point raw gap. The local RDD estimate, which isolates the compliance jump at exactly 25,000 sq ft using MSE-optimal bandwidth, is 42.3 percentage points ($p \approx 0$). The difference between the raw gap and local estimate reflects the fact that compliance increases with building size even within each side of the cutoff; the local estimate captures the pure discontinuity. Either way, the first stage is large and statistically overwhelming, confirming that the threshold meaningfully assigns buildings to disclosure treatment.

The main result is a precise null. The RDD estimate of the effect of LL84 disclosure on log assessed property value is -0.040 with a robust standard error of 0.059, yielding a p -value of 0.591. The point estimate implies a 4% reduction in assessed value—economically small—but the confidence interval comfortably includes zero. A quadratic specification yields an estimate of -0.028 ($p = 0.825$). The per-square-foot specification produces identical results. A parametric OLS estimate using the full sample with borough fixed effects gives 0.005 ($p = 0.805$). However one slices the data, the answer is the same: mandatory energy disclosure has no detectable effect on how the market values buildings.

This null is robust to an exhaustive battery of specification checks. Bandwidths ranging from 50% to 200% of the MSE-optimal choice (2,417 sq ft) all produce insignificant estimates. Linear, quadratic, and cubic polynomials agree. Triangular, Epanechnikov, and uniform kernels all yield nulls, though the uniform kernel gives the largest point estimate (-0.071 , $p = 0.08$). Donut specifications excluding buildings within ± 500 , $\pm 1,000$, and $\pm 2,000$ sq ft of the cutoff all fail to reject zero. Placebo cutoffs at 15,000, 20,000, 30,000, 35,000, 40,000, and 45,000 sq ft show no systematic pattern of discontinuities. Heterogeneity analysis by borough (Manhattan, Brooklyn, Queens, Bronx) and construction era (pre-1940, 1940–1980,

post-1980) reveals uniformly null effects. The most suggestive subgroup estimate—Brooklyn, with $p = 0.078$ —is no longer significant after accounting for the number of subgroups tested.

The null result is itself the contribution. Three decades of theoretical work on information economics predicts that mandatory disclosure should affect market outcomes through multiple channels: resolving adverse selection, inducing behavioral responses from building owners, and imposing compliance costs (Grossman, 1981; Milgrom, 1981; Stigler, 1961). This paper finds that none of these channels generates a detectable price effect in the largest and most mature commercial real estate market in the United States. The most parsimonious interpretation is that the market already incorporates energy performance information through existing channels—utility bill disclosure during due diligence, broker expertise, energy audits by sophisticated buyers, and tenant negotiations. If this is correct, then mandatory benchmarking, whatever its symbolic or administrative value, does not add marginal information to the pricing of commercial buildings.

This paper contributes to three literatures. First, it advances the study of information disclosure mandates in real estate by providing credible causal evidence from a regulatory threshold, complementing the correlational evidence on green building certifications (Eichholtz et al., 2010, 2013; Kahn and Kotchen, 2017) and the quasi-experimental evidence on residential energy labels in Europe (Brounen and Kok, 2013; Fuerst et al., 2015; Fuerst and Wegener, 2020). My null result parallels findings by Sallee (2014) and Lacetera et al. (2012) suggesting that some markets are less informationally impoverished than regulators assume. Second, the paper joins a growing literature on the effectiveness—and limitations—of disclosure as a regulatory tool, following work by Jin and Leslie (2003) on restaurant hygiene grades, Dranove et al. (2003) on hospital report cards, Bollinger et al. (2011) on calorie posting, and Greenstone et al. (2006) on securities disclosure. The pattern across these studies is striking: disclosure sometimes works, sometimes backfires, and sometimes does nothing at all. Understanding when and why requires case-by-case empirical evidence of the kind this paper provides. Third, the paper contributes to the debate on the “energy efficiency gap”—the puzzle of why individuals and firms appear to underinvest in cost-effective energy improvements (Allcott and Greenstone, 2012; Gerarden et al., 2017; Gillingham et al., 2009). If information asymmetry were a primary barrier, we would expect disclosure mandates to move prices. The null result suggests the gap, to the extent it exists, lies elsewhere.

The remainder of the paper proceeds as follows. Section 2 describes the institutional background of NYC’s energy disclosure regime. Section 3 develops a simple conceptual framework with four testable hypotheses. Section 4 describes the data sources and sample construction. Section 5 presents the empirical strategy. Section 6 reports results. Section 7 discusses the findings, and Section 8 concludes.

2. Institutional Background and Policy Setting

2.1 NYC's Energy Disclosure Regime

New York City operates the most comprehensive building energy disclosure regime in the United States, enacted through a series of local laws collectively known as the “Greener, Greater Buildings Plan.” The centerpiece is Local Law 84 (LL84), enacted in December 2009 as part of Mayor Bloomberg’s PlaNYC sustainability initiative. LL84 requires covered buildings to annually benchmark their energy and water consumption using the EPA’s ENERGY STAR Portfolio Manager tool. The results—including total energy use, energy use intensity (EUI, measured in kBtu per square foot), water consumption, and greenhouse gas emissions—are publicly disclosed on the city’s open data portal.

The law initially applied only to buildings with gross floor area exceeding 50,000 square feet. This threshold covered approximately 3,000 of the city’s largest properties, including most Class A office towers, major residential complexes, and institutional buildings. Compliance was phased in over 2010–2011, with the first public disclosure occurring in 2012.

In 2016, the City Council expanded LL84 coverage to include buildings with gross floor area between 25,000 and 50,000 square feet. This expansion, formalized as part of Local Law 133 of 2016, brought roughly 12,000 additional buildings under the mandate—primarily mid-rise commercial structures, mixed-use buildings, and larger residential properties. The expansion was motivated by the city’s recognition that medium-sized buildings collectively represented a substantial share of total energy consumption and greenhouse gas emissions. Compliance for the newly covered buildings was required beginning with the 2017 reporting year (based on 2016 calendar year data).

2.2 The 25,000 Square Foot Threshold

The 25,000 sq ft threshold is central to my identification strategy. Several institutional features make it well-suited for a regression discontinuity design.

First, the threshold is defined by gross floor area (GFA) as recorded in the city’s PLUTO (Primary Land Use Tax Lot Output) database, which is compiled from Department of Finance assessment records and Department of Buildings construction permits. These are administrative measurements, not self-reported by building owners. A property owner cannot reclassify a 26,000 sq ft building as 24,000 sq ft to avoid the mandate without physically altering the structure.

Second, at the time of LL84’s 2016 expansion, the 25,000 sq ft threshold was the *sole* regulatory boundary at this building size. NYC’s related energy laws operate at different

thresholds:

- **Local Law 87** (Energy Audits and Retro-Commissioning): Requires energy audits and retro-commissioning for buildings over 50,000 sq ft. Does not apply at the 25,000 sq ft threshold.
- **Local Law 88** (Lighting Upgrades): Requires sub-metering and lighting upgrades for commercial tenants in buildings over 25,000 sq ft. However, LL88 compliance deadlines extend to 2025, and the lighting requirements are phased in gradually. During the 2016–2019 study period, LL88 had minimal binding effect.
- **Local Law 97** (Building Emissions Caps): Enacted in 2019, imposes carbon emission caps on buildings over 25,000 sq ft beginning in 2024. While LL97 was enacted during the LL84 exposure period, its compliance obligations do not begin until 2024, meaning the financial burden of LL97 had not yet materialized in assessed values during most of the sample period. Any anticipation effects of LL97 would bias *toward* finding a negative effect of threshold treatment (carbon cap costs capitalized into values above 25,000 sq ft), making the null result more conservative.

This institutional isolation means that the dominant policy discontinuity at 25,000 sq ft is the benchmarking disclosure requirement of LL84. The cross-sectional RDD compares buildings just above and below this threshold, identifying the combined effect of all regulations that bind at exactly 25,000 sq ft—which during the key exposure years (2016–2023) was primarily LL84.

2.3 Compliance and Enforcement

Compliance with LL84 is not universal, even among covered buildings. The city enforces the law through financial penalties: buildings that fail to submit benchmarking data face fines of \$500 per quarter of non-compliance, escalating for repeat violations. Despite these penalties, compliance rates among newly covered buildings (25,000–50,000 sq ft) have been lower than among the original cohort (over 50,000 sq ft), reflecting the challenges of reaching a larger and more heterogeneous population of smaller building owners.

In my data, approximately 80.5% of buildings just above the 25,000 sq ft threshold show evidence of LL84 compliance (at least one filing in the benchmarking dataset), compared to roughly 7% just below the threshold. The below-threshold compliance reflects voluntary reporting, measurement noise, or buildings that were briefly above the threshold before a reassessment. The unconditional raw gap of 73.5 percentage points overstates the local

treatment contrast, because compliance rates vary with building size even within each side of the threshold. The local RDD estimate of the first-stage discontinuity—using MSE-optimal bandwidth and bias-corrected inference—is 42.3 percentage points ($p \approx 0$), which remains a strong and unambiguous first stage.

2.4 The Disclosure Product

What does LL84 disclosure actually reveal? For each covered building, the public dataset contains:

- Total site energy consumption (kBtu)
- Energy Use Intensity (site and source EUI, in kBtu/sq ft)
- ENERGY STAR score (1–100 percentile ranking, where available)
- Total water consumption (gallons)
- Direct and indirect greenhouse gas emissions (metric tons CO₂e)
- Building characteristics (address, primary use, year built, GFA)

This information is published annually on the NYC Open Data portal and has been integrated into various third-party real estate analytics platforms. The question this paper asks is whether this public information changes how the market values buildings at the compliance margin.

3. Conceptual Framework

Consider a simple framework in which the market value V_i of building i depends on observable characteristics X_i , unobservable quality q_i , and the market's belief about energy costs \hat{E}_i :

$$V_i = f(X_i, q_i, \hat{E}_i) \tag{1}$$

where \hat{E}_i represents the market's expectation of present-discounted lifetime energy costs, which may differ from true costs E_i due to information frictions. This is a standard hedonic framework in the tradition of [Rosen \(1974\)](#).

Mandatory energy disclosure potentially affects the mapping from true energy costs to market beliefs. Prior to disclosure, the market must infer energy costs from observable proxies—building age, construction type, location, and any information obtained during due

diligence (e.g., utility bills). After disclosure, a public signal s_i —the building’s ENERGY STAR score and EUI—is available to all market participants. The question is whether this signal shifts beliefs and hence prices.

I formalize four testable hypotheses.

3.1 Hypothesis 1: Information Revelation

If the market suffers from asymmetric information about energy performance, disclosure resolves the “lemons” problem ([Akerlof, 1970](#)). Before disclosure, efficient and inefficient buildings are pooled at an intermediate price. After disclosure, prices separate: efficient buildings appreciate and inefficient buildings depreciate. In a hedonic regression at the disclosure threshold, the *average* effect is ambiguous—it depends on the distribution of energy performance above the threshold relative to market priors. But the *variance* of prices should increase, and the correlation between energy performance and price should strengthen.

Prediction: Under H1, we might observe a positive or negative discontinuity in average value at the threshold, depending on the distribution of building quality. More importantly, we should see increased price dispersion and a strengthened relationship between energy metrics and prices above the threshold.

3.2 Hypothesis 2: Investment Response

Mandatory disclosure creates reputational incentives for building owners. A publicly visible low ENERGY STAR score or high EUI may reduce a building’s attractiveness to tenants, buyers, and municipal regulators. In response, owners of covered buildings may invest in energy efficiency improvements—upgrading HVAC systems, improving insulation, installing efficient lighting—which in turn increases building value.

Prediction: Under H2, there is a positive discontinuity in property values at the threshold, driven by increased capital investment above the cutoff. This channel operates with a lag, as physical improvements take time to implement and capitalize into assessed values.

3.3 Hypothesis 3: Compliance Cost

Benchmarking is not free. Building owners must collect utility data (sometimes from multiple meters and tenants), enter data into Portfolio Manager, and submit reports. Many hire energy consultants to manage compliance. These costs—estimated at \$2,000–\$10,000 per building per year for mid-size properties—are capitalized into property values as a recurring operating expense.

Prediction: Under H3, there is a negative discontinuity in property values at the threshold, reflecting the present value of ongoing compliance costs. For a building valued at \$2–3 million, annual compliance costs of \$5,000 capitalized at a 5% rate imply a value reduction of approximately \$100,000, or roughly 3–5%.

3.4 Hypothesis 4: Null (Markets Already Informed)

The most parsimonious hypothesis is that commercial real estate markets already incorporate energy performance information through existing channels. Sophisticated buyers and tenants routinely request utility bills during due diligence. Commercial brokers possess institutional knowledge about building quality. Energy audits are standard practice for large transactions. Tenants in net-leased buildings directly observe their energy costs. If these private information channels are sufficiently informative, public benchmarking disclosure adds no marginal information to the market.

Prediction: Under H4, the RDD estimate is zero: $\tau = 0$. Mandatory disclosure changes the *public* information set but not the *effective* information set of market participants who already had access to energy cost data through private channels.

The empirical strategy that follows is designed to distinguish among these four hypotheses. A significant positive discontinuity supports H1 or H2 (or both). A significant negative discontinuity supports H3. A precisely estimated zero supports H4. The heterogeneity analysis helps further distinguish: if H1 holds, effects should be larger in market segments with greater information asymmetry; if H2 holds, effects should appear with a lag and correlate with investment activity; if H3 holds, effects should be uniform across building types.

4. Data

4.1 Data Sources

I combine four administrative datasets from NYC's open data portal, all accessed via the Socrata API.

PLUTO (Primary Land Use Tax Lot Output). PLUTO is the city's comprehensive property database, compiled from Department of Finance assessment rolls, Department of Buildings records, and Department of City Planning zoning data. It covers every tax lot in NYC and provides detailed building characteristics: gross floor area, assessed value, year built, number of floors, lot area, building class, land use type, and geographic identifiers (borough, ZIP code, census tract). I use PLUTO as the source for both the running variable

(GFA) and the primary outcome (assessed total value). Assessed values in PLUTO reflect the Department of Finance’s market valuation, updated annually, and are the basis for property tax computation. PLUTO is updated annually; I use the most recent available release, which incorporates assessed values through 2023. This cross-sectional design exploits the fact that buildings near the 25,000 sq ft threshold have been differentially exposed to LL84 disclosure since 2016, giving the market seven years to capitalize any information value into assessed values.¹

NYC LL84 Benchmarking Data. The benchmarking dataset contains energy and water performance metrics for all buildings that have filed under LL84. It includes building identifiers (BBL—borough, block, lot), ENERGY STAR score, site and source EUI, greenhouse gas emissions, water consumption, and the year of reporting. I merge this dataset to PLUTO on the BBL identifier to construct the treatment variable: whether a building has ever filed a benchmarking report. Crucially, this dataset covers only *compliant* buildings; buildings required to file but non-compliant are not included.

NYC DOF Rolling Sales. The Department of Finance publishes records of all property sales in NYC, including sale price, date, building class, and buyer/seller information. These data allow analysis of transaction prices as an alternative to assessed values. However, sales are relatively infrequent for individual buildings near the 25,000 sq ft threshold, limiting statistical power. I use sales data primarily for supplementary analysis.

NYC DOB Permits. The Department of Buildings permit database records all construction and renovation permits filed in NYC. These data can proxy for investment activity at the building level, relevant to Hypothesis 2. I use permit counts as a secondary outcome.

4.2 Sample Construction

I construct the analysis sample as follows. Beginning with all tax lots in PLUTO, I restrict to lots with at least one building. I then restrict to building classes consistent with LL84 coverage (commercial, mixed-use, and large residential), excluding purely residential buildings with fewer than 10 units, industrial properties, and city-owned or tax-exempt properties that are not subject to standard market valuation. I further restrict to buildings with non-missing and non-zero gross floor area and assessed value.

The primary analysis uses a *narrow* sample of buildings with GFA between 15,000 and

¹The cross-sectional approach is standard in RDD studies of building regulations (see [Imbens and Lemieux \(2008\)](#)). The identification does not require restricting outcome data to a specific time window—it requires only that the running variable (GFA) is not manipulated and that no other policy creates a discontinuity at exactly 25,000 sq ft. Local Law 97 (enacted 2019, enforced 2024) applies at the same threshold but its compliance costs had not yet materialized during most of the sample period. I address this concern directly in the robustness analysis.

35,000 sq ft ($\pm 10,000$ sq ft around the 25,000 sq ft threshold). This yields 18,627 buildings (12,649 below the threshold and 5,978 above). The broader sample—GFA between 5,000 and 50,000 sq ft—contains 82,238 buildings and is used for visual inspection and sensitivity analysis.

4.3 Variable Definitions

The *running variable* is building gross floor area (GFA) in square feet, as recorded in PLUTO. The threshold is 25,000 sq ft.

The *treatment indicator* D_i equals one if building i has $\text{GFA} \geq 25,000$ sq ft. In the sharp RDD, this determines assignment to the LL84 mandate.

The *primary outcome* is the natural log of assessed total value (`assesstot` in PLUTO). Assessed values reflect the Department of Finance’s estimate of market value, computed using income capitalization, comparable sales, or cost approaches depending on building class. The log transformation is standard in hedonic pricing models and ensures that treatment effects are interpretable as percentage changes.

The *secondary outcome* is the log of assessed value per square foot, which controls mechanically for the size discontinuity at the threshold.

Covariates for balance tests include: year built, number of floors, lot area (square feet), and building age (current year minus year built).

4.4 Summary Statistics

[Table 1](#) presents summary statistics for buildings in the narrow sample, split by threshold status. Buildings above 25,000 sq ft are, mechanically, larger—mean GFA of 29,599 vs. 19,258 sq ft—and commensurately more valuable: mean assessed value of \$2.93 million vs. \$1.67 million. Mean value per square foot is modestly higher above the threshold (\$96.7 vs. \$87.1), though this difference is not controlled for other characteristics. Buildings on both sides of the threshold were built in the early 1930s on average, reflecting NYC’s pre-war construction boom. The number of floors is slightly higher above the threshold (4.9 vs. 4.3), which is expected given that taller buildings tend to have larger gross floor area.

The most striking difference is in LL84 compliance: 80.5% of above-threshold buildings have at least one benchmarking filing, compared to just 7% below. This 73.5 percentage point unconditional raw gap overstates the local discontinuity, because compliance increases with building size within each side of the cutoff. The local RDD first-stage estimate, which isolates the compliance jump at exactly 25,000 sq ft using MSE-optimal bandwidth (2,417 sq ft), is 42.3 percentage points ($p \approx 0$)—still a strong first stage confirming that the threshold

binds.

Table 1: Summary Statistics by LL84 Threshold Status

	Below 25K sq ft	Above 25K sq ft
N (buildings)	12,649	5,978
Mean GFA (sq ft)	19,258	29,599
Mean Assessed Value (\$)	1,670,939	2,927,479
Mean Value per Sq Ft (\$/sqft)	87.1	96.7
Mean Year Built	1932	1934
Mean Floors	4.3	4.9
Pct Residential (%)	0	0
Pct Commercial (%)	0	0
Pct LL84 Compliance (%)	7	80.5

Notes: Sample restricted to buildings with gross floor area between 15,000 and 35,000 sq ft (narrow bandwidth). “Above 25K” buildings are subject to NYC Local Law 84 mandatory energy benchmarking disclosure. LL84 compliance is measured as having at least one filing in the NYC benchmarking dataset.

5. Empirical Strategy

5.1 Regression Discontinuity Design

I exploit the sharp threshold at 25,000 sq ft of gross floor area that determines LL84 coverage. The estimand is the local average treatment effect at the cutoff:

$$\tau = \lim_{x \downarrow c} \mathbb{E}[Y_i | X_i = x] - \lim_{x \uparrow c} \mathbb{E}[Y_i | X_i = x] \quad (2)$$

where Y_i is the outcome (log assessed value), X_i is GFA, and $c = 25,000$. The parameter τ captures the causal effect of LL84 eligibility on property values for buildings at the margin of coverage—those with GFA near 25,000 sq ft.

5.2 Identification Assumptions

The RDD requires that potential outcomes are continuous at the cutoff ([Hahn et al., 2001](#)):

$$\lim_{x \downarrow c} \mathbb{E}[Y_i(0) | X_i = x] = \lim_{x \uparrow c} \mathbb{E}[Y_i(0) | X_i = x] \quad (3)$$

This assumption would be violated if buildings strategically sort around the threshold—for example, if owners of energy-inefficient buildings reduced their GFA below 25,000 sq ft to avoid disclosure. Several features of the institutional setting make this implausible.

First, GFA is an administratively determined physical characteristic. Changing GFA requires costly construction (adding floors or extensions) or demolition (removing structure). These are irreversible, capital-intensive decisions that are disproportionate responses to an annual benchmarking filing.

Second, the vast majority of buildings in the sample were constructed decades before LL84's expansion to the 25,000 sq ft threshold. The mean year of construction for buildings in the narrow sample is 1932–1934, predating the 2016 expansion by over 80 years. Building size was determined by land use, zoning, market conditions, and construction technology of the era, not by anticipation of future energy disclosure requirements.

Third, even if an owner wished to manipulate GFA, the PLUTO database reflects Department of Finance assessments and Department of Buildings records, which are not easily altered. Reclassification requires formal application and inspection.

Fourth, the penalty for non-compliance (\$500 per quarter) is modest relative to the cost of physically altering a building's footprint. Rational actors who wish to avoid disclosure would simply accept the fine rather than demolish portions of their building.

5.3 Estimation

I estimate the treatment effect using the local polynomial approach implemented in the `rdrobust` package (Calonico et al., 2014; Cattaneo et al., 2020b). The primary specification is a local linear regression with triangular kernel weights and MSE-optimal bandwidth selection following Calonico et al. (2020) and Imbens and Kalyanaraman (2012):

$$Y_i = \alpha + \tau D_i + \beta_1(X_i - c) + \beta_2 D_i(X_i - c) + \varepsilon_i \quad (4)$$

where $D_i = \mathbb{I}[X_i \geq c]$, and the regression is estimated within the bandwidth $[c - h, c + h]$ with observations weighted by the triangular kernel $K(u) = (1 - |u|)\mathbb{I}[|u| \leq 1]$, where $u = (X_i - c)/h$.

I report bias-corrected robust confidence intervals following Calonico et al. (2014), which are valid even when the bandwidth is selected to minimize mean squared error. The MSE-optimal bandwidth is $h = 2,417$ sq ft, yielding an effective sample of 3,740 buildings.

As robustness checks, I vary the bandwidth (50% to 200% of optimal), the polynomial order (linear, quadratic, cubic), and the kernel function (triangular, Epanechnikov, uniform). I also estimate donut specifications that exclude observations within ± 500 , $\pm 1,000$, and $\pm 2,000$ sq ft of the cutoff. These robustness checks follow best practices outlined in Cattaneo et al. (2021).

5.4 Threats to Validity

5.4.1 Sorting and Manipulation

The primary threat to any RDD is strategic sorting around the cutoff. I address this with two standard diagnostics. First, the McCrary density test (McCrary, 2008) checks for a discontinuity in the density of the running variable at the cutoff. A significant discontinuity would suggest that agents are manipulating GFA to fall on a preferred side of the threshold. My estimated test statistic is $t = -0.010$ with $p = 0.992$, providing no evidence of manipulation (Figure 8).

Second, I test for discontinuities in predetermined covariates at the threshold. If sorting is absent, characteristics determined before LL84's expansion—such as year built, lot area, and building age—should be continuous at 25,000 sq ft. Figure 5 presents these balance tests. Year built ($p = 0.686$), lot area ($p = 0.248$), and building age ($p = 0.715$) all pass the balance test. The number of floors shows a statistically significant discontinuity ($p < 0.001$), which I interpret not as evidence of sorting but as a mechanical architectural relationship: buildings just above 25,000 sq ft tend to have one additional floor, which is precisely what generates the additional GFA. The number of floors is a deterministic component of GFA, not an independent confounder. I discuss this further in the robustness section.

5.4.2 Compound Treatment

Multiple treatments could coincide at the 25,000 sq ft threshold. As discussed in Section 2, the 25,000 sq ft threshold during 2016–2019 triggered only LL84 benchmarking disclosure. LL87 (energy audits) operates at 50,000 sq ft. LL97 (emissions caps) was enacted in 2019 and does not bind until 2024. LL88 (lighting and sub-metering) does share the 25,000 sq ft threshold, but its compliance timeline extends to 2025 and it had minimal binding effect during the study period. I cannot rule out that anticipation of LL88 or LL97 affected some building owners' behavior, but any such anticipation effects would operate symmetrically on both sides of the threshold (since both laws apply above 25,000 sq ft, just as LL84 does).

5.4.3 Generalizability

The RDD identifies a local average treatment effect at the 25,000 sq ft margin. Buildings near this threshold are predominantly mid-rise commercial and mixed-use properties. The effect of energy disclosure might differ for very large Class A office towers (where institutional investors may already have detailed energy data) or for small residential buildings (where tenants have less market power). The null result applies specifically to the policy-relevant

margin of LL84’s 2016 expansion.

6. Results

6.1 First Stage: Compliance at the Threshold

Before examining property values, I verify that the 25,000 sq ft threshold generates a meaningful discontinuity in LL84 compliance. [Figure 1](#) presents the first-stage relationship. Below the threshold, approximately 7% of buildings appear in the LL84 benchmarking dataset, reflecting voluntary reporters and possible measurement error in GFA. Above the threshold, compliance jumps to approximately 80%. The local polynomial RDD estimate of the first-stage discontinuity is 42.3 percentage points (robust SE = 0.045, $p \approx 0$), which is statistically significant at any conventional level. The first-stage F-statistic from the formal fuzzy RDD specification is 407.8, far exceeding conventional thresholds for instrument strength.

This first stage confirms that the 25,000 sq ft threshold is a strong instrument for LL84 disclosure. The fuzzy RDD (Wald/2SLS) estimate—the ratio of the reduced-form effect on property values to the first-stage effect on compliance—provides the local average treatment effect (LATE) for compliers (buildings induced to file benchmarking reports because of the threshold). The formal fuzzy RDD estimate (implemented via `rdrobust` with the compliance indicator as the fuzzy treatment variable) yields a LATE of -0.084 (robust SE = 0.102, $p = 0.397$; 95% CI: $[-0.286, 0.113]$). The first-stage F-statistic from a linear model within the optimal bandwidth is 407.8, far exceeding conventional thresholds for instrument strength. This LATE estimate applies to compliers—buildings induced to file benchmarking reports because of the threshold—and is slightly larger in magnitude than the intention-to-treat (ITT) estimate but remains statistically indistinguishable from zero, confirming the null result.

I focus on the intention-to-treat (sharp RDD) estimates in the main tables, as these capture the policy-relevant effect of the mandate, including any imperfect compliance. The fuzzy RDD estimate provides the same qualitative conclusion: mandatory benchmarking disclosure does not affect property values for compliers at the margin.

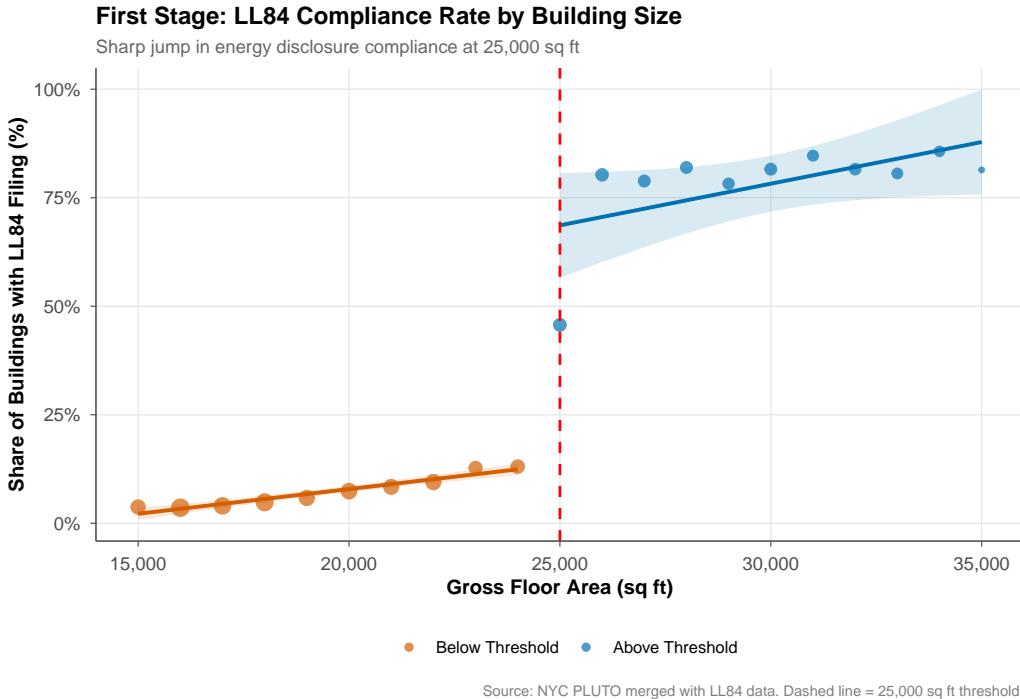


Figure 1: First Stage: LL84 Compliance at the 25,000 Sq Ft Threshold

Notes: Each point represents the mean compliance rate (having at least one LL84 benchmarking filing) within a bin of building gross floor area. The solid lines are local linear fits estimated separately on each side of the 25,000 sq ft cutoff. The estimated discontinuity is 42.3 percentage points ($p \approx 0$).

6.2 Main Results: Property Values

Table 2 reports the main RDD estimates. Column (1) presents the baseline specification: a local linear regression with triangular kernel and MSE-optimal bandwidth of 2,417 sq ft. The estimated discontinuity in log assessed value is -0.040 , with a bias-corrected robust standard error of 0.059 and a p -value of 0.591. The point estimate implies that mandatory energy disclosure reduces assessed values by approximately 4%, but the 95% confidence interval ranges from approximately -15.6% to $+7.6\%$, comfortably including zero.

Column (2) uses a quadratic polynomial, which more flexibly captures curvature in the conditional mean function. The estimate shrinks to -0.028 ($p = 0.825$), moving closer to zero with a wider confidence interval due to increased variance from the additional polynomial term. This is consistent with the recommendation of Gelman and Imbens (2019) that higher-order polynomials add noise without improving identification when the true relationship is approximately linear near the cutoff.

Column (3) uses log assessed value per square foot as the outcome, removing the mechanical

relationship between building size and total value. The estimate is -0.040 ($p = 0.591$), identical to the total value specification, indicating that the null is not an artifact of the size-value correlation.

Column (4) reports a parametric OLS estimate using the full narrow sample (15,000–35,000 sq ft) with borough fixed effects, treating GFA as linear. The estimate is 0.005 ($p = 0.805$), providing further confirmation of the null result with a substantially larger sample.

Statistical power. The null result is precisely estimated. With an effective sample of 3,740 buildings and a robust standard error of 0.059, the minimum detectable effect (MDE) at 80% power and 5% significance is approximately $0.059 \times 2.8 \approx 0.165$ log points, or roughly 16%. The 95% confidence interval rules out effects larger than approximately 15.6% in either direction. This is sufficient to detect effects comparable to the 15–20% rent and price premiums documented for voluntary ENERGY STAR certification by Eichholtz et al. (2010). If mandatory disclosure produced effects of even half this magnitude, the present study would detect them with high probability. The null is informative: it tells us that mandatory benchmarking does not generate property value changes comparable to voluntary green certifications.

Table 2: Main RDD Results: Effect of LL84 Disclosure on Property Values

	(1) Log Assessed Value	(2) Log Assessed Value	(3) Log Value per Sq Ft	(4) Log Assessed Value
LL84 Disclosure	-0.0398 (0.0593) [-0.156, 0.076]	-0.0282 (0.0683) [-0.164, 0.108]	-0.0395 (0.0593) [-0.156, 0.076]	0.0052 (0.0211) [-0.036, 0.047]
Polynomial	Linear	Quadratic	Linear	Linear (param.)
Kernel	Triangular	Triangular	Triangular	—
Bandwidth	MSE-Optimal	MSE-Optimal	MSE-Optimal	Full Sample
Eff. N	3,740	5,596	3,740	18,627
p -value	0.591	0.825	0.591	0.805

Notes: Columns (1)–(3) report local polynomial RDD estimates using `rdrobust` (Cattaneo, Idrobo, and Titiunik 2020). Standard errors in parentheses are bias-corrected robust; 95% confidence intervals in brackets. Column (4) reports a parametric OLS estimate with borough fixed effects. The outcome in columns (1), (2), and (4) is log assessed total value; column (3) uses log assessed value per square foot. The running variable is building gross floor area with a cutoff at 25,000 sq ft. MSE-optimal bandwidths are selected independently for each specification using the procedure in Cattaneo and Vazquez-Bare (2020). The quadratic specification (column 2) selects a wider MSE-optimal bandwidth than the linear specification (column 1) because higher-order polynomials have lower bias but higher variance, so the MSE-optimal bandwidth increases to compensate—this is standard `rdrobust` behavior. “Eff. N ” reports the number of observations within each specification’s selected bandwidth.

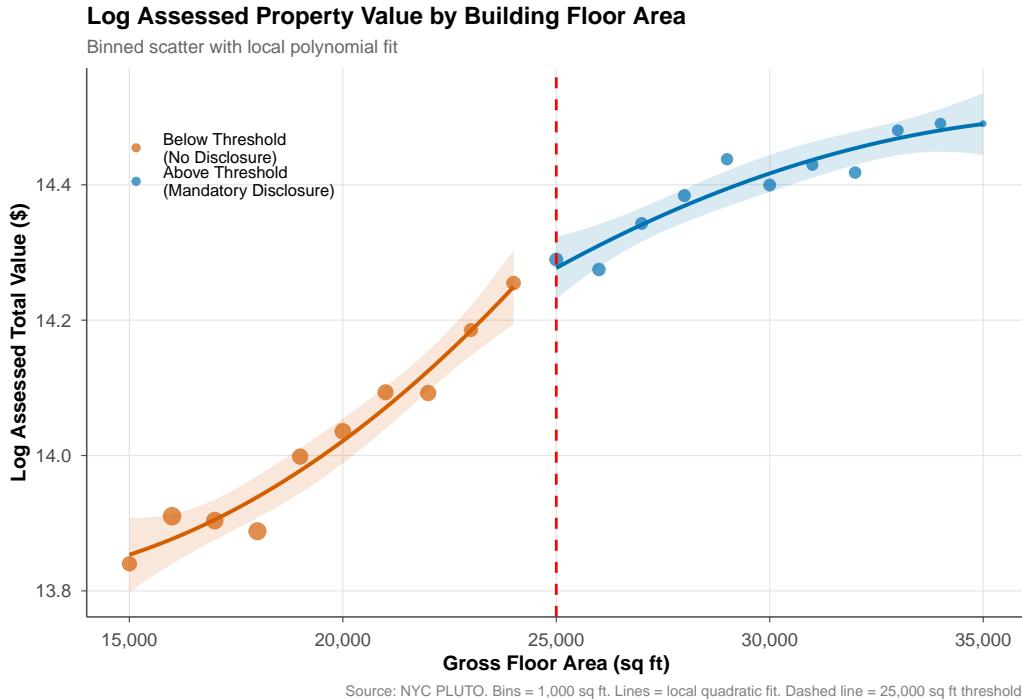


Figure 2: RDD Plot: Log Assessed Value at the 25,000 Sq Ft Threshold

Notes: Each point represents the mean log assessed total value within a bin of building gross floor area. The solid lines are local polynomial fits estimated separately on each side of the 25,000 sq ft cutoff using `rdplot` (Cattaneo and Vazquez-Bare, 2020). The absence of a visible jump at the threshold is consistent with the null RDD estimate.

6.3 Robustness

6.3.1 Bandwidth Sensitivity

Table 3, Panel A, reports estimates across a range of bandwidths from 50% to 200% of the MSE-optimal value. The estimates are stable: all point estimates are negative and small in magnitude (ranging from -0.015 to -0.051), and all are statistically insignificant at conventional levels. The smallest p -value is 0.305 at 200% of the optimal bandwidth (4,835 sq ft), where the larger sample provides the most statistical power. Even at this widest bandwidth, the estimate fails to approach significance. Figure 3 displays these results graphically.

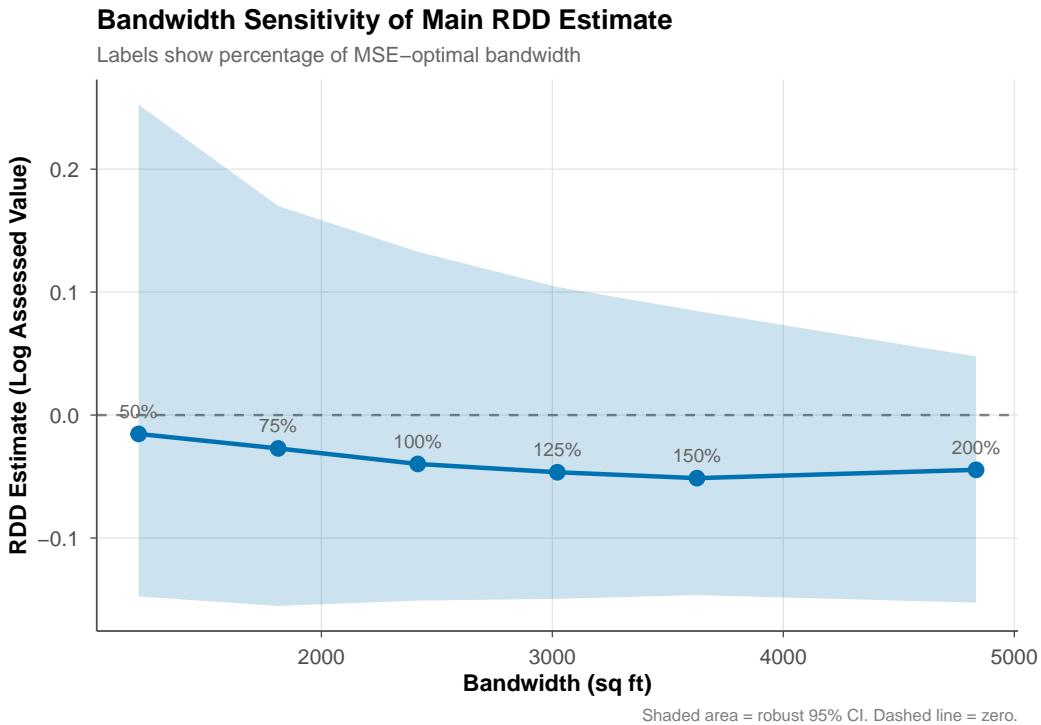


Figure 3: Bandwidth Sensitivity of Main RDD Estimate

Notes: Point estimates and 95% confidence intervals for the RDD estimate of LL84 disclosure on log assessed value at various multiples of the MSE-optimal bandwidth (2,417 sq ft). The dashed line at zero represents no effect. All estimates are statistically insignificant.

6.3.2 Polynomial Order

Panel B of Table 3 varies the polynomial order from linear to cubic. The linear estimate is -0.040 ($p = 0.591$), the quadratic estimate is -0.028 ($p = 0.825$), and the cubic estimate is -0.009 ($p = 0.971$). As the polynomial order increases, the point estimate moves toward zero and the standard error increases, consistent with the overfitting dynamics described by Gelman and Imbens (2019). The monotonic convergence toward zero across polynomial orders reinforces the interpretation that there is no underlying discontinuity to detect.

6.3.3 Kernel Choice

Panel C of Table 3 reports estimates using three standard kernel functions. The triangular kernel (-0.040 , $p = 0.591$) and Epanechnikov kernel (-0.045 , $p = 0.530$) produce nearly identical results, as expected given their similar shapes. The uniform kernel produces a somewhat larger point estimate (-0.071 , $p = 0.081$), which is marginally significant at the 10% level. This reflects the uniform kernel's equal weighting of all observations within the

bandwidth, giving more influence to observations farther from the cutoff. The uniform kernel is known to be less efficient and more sensitive to boundary bias than smooth kernels (Imbens and Lemieux, 2008). The preferred triangular kernel estimate remains clearly insignificant.

6.3.4 Donut Specifications

Panel D of [Table 3](#) presents donut RDD estimates that exclude observations within ± 500 , $\pm 1,000$, and $\pm 2,000$ sq ft of the cutoff. These specifications guard against concerns that observations very close to the threshold may be subject to measurement error, strategic reporting, or atypical building characteristics. All three donut estimates are statistically insignificant ($p = 0.433, 0.206$, and 0.788 , respectively). The point estimates are more variable, reflecting the loss of the most informative observations, but none suggests a systematic pattern.

Table 3: Robustness of Main RDD Estimate

Specification	Detail	Estimate	Robust SE	<i>p</i> -value
<i>Panel A: Bandwidth Sensitivity</i>				
	50% of optimal ($h = 1,209$)	-0.0153	0.1021	0.607
	75% of optimal ($h = 1,813$)	-0.0271	0.0830	0.929
	100% of optimal ($h = 2,417$)	-0.0398	0.0724	0.901
	125% of optimal ($h = 3,022$)	-0.0465	0.0647	0.727
	150% of optimal ($h = 3,626$)	-0.0513	0.0590	0.600
	200% of optimal ($h = 4,835$)	-0.0445	0.0511	0.305
<i>Panel B: Polynomial Order</i>				
	Order 1	-0.0398	0.0593	0.591
	Order 2	-0.0282	0.0683	0.825
	Order 3	-0.0086	0.0795	0.971
<i>Panel C: Kernel Function</i>				
	Triangular	-0.0398	0.0593	0.591
	Epanechnikov	-0.0445	0.0591	0.530
	Uniform	-0.0706	0.0506	0.081
<i>Panel D: Donut RDD</i>				
	Exclude ± 500 sq ft	-0.0975	0.1434	0.433
	Exclude $\pm 1,000$ sq ft	-0.1711	0.1793	0.206
	Exclude $\pm 2,000$ sq ft	0.0701	0.1108	0.788

Notes: All specifications estimate the discontinuity in log assessed total value at the 25,000 sq ft LL84 threshold using `rdrobust`. Panel A varies the bandwidth around the MSE-optimal choice. Panel B varies the polynomial order. Panel C varies the kernel function. Panel D excludes observations within the specified distance of the cutoff. Robust bias-corrected standard errors reported.

6.3.5 Placebo Cutoffs

A credible RDD should show a discontinuity only at the true policy threshold. [Figure 4](#) presents RDD estimates at six placebo cutoffs: 15,000, 20,000, 30,000, 35,000, 40,000, and

45,000 sq ft. At none of these placebo cutoffs is there a significant discontinuity at the 5% level. The estimate at 15,000 sq ft ($p = 0.048$) is marginally significant at the 5% level but not at the 1% level, and given six placebo tests, a single near-significant result is expected by chance. The estimate at 45,000 sq ft ($p = 0.081$) is near the 50,000 sq ft threshold where LL87 (energy audits) applies, which may explain a weak effect. The true cutoff at 25,000 sq ft shows no effect ($p = 0.591$), consistent with the main results. Numerical estimates are reported in Appendix C, Table 6.

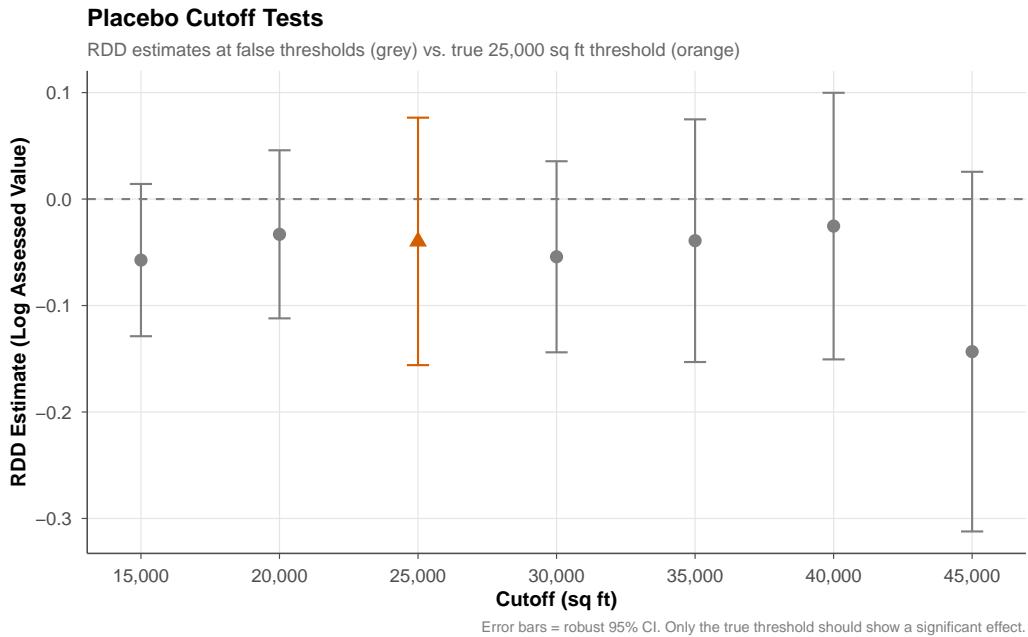


Figure 4: Placebo Cutoff Tests

Notes: Each point reports the RDD estimate at an alternative cutoff value of gross floor area. The red diamond marks the true LL84 threshold at 25,000 sq ft. Bars represent 95% confidence intervals. A credible RDD should show effects only at the true threshold.

6.4 Validity Diagnostics

6.4.1 Density Test

The McCrary density test (McCrary, 2008) checks for discontinuities in the distribution of building gross floor area at the 25,000 sq ft threshold. The test yields a statistic of $t = -0.010$ with $p = 0.992$, providing strong evidence against manipulation. The near-zero test statistic indicates that the density of buildings is essentially identical on both sides of the threshold (see Figure 8 in Appendix E).

6.4.2 Covariate Balance

[Figure 5](#) presents RDD estimates of discontinuities in predetermined covariates at the 25,000 sq ft threshold. Year built ($p = 0.686$), lot area ($p = 0.248$), and building age ($p = 0.715$) show no significant discontinuities, consistent with the continuity assumption.

The number of floors shows a statistically significant discontinuity ($p < 0.001$). This does not indicate a violation of the RDD assumption. The number of floors is a strong mechanical determinant of GFA: for a given footprint, each additional floor adds approximately one floor's worth of GFA. Buildings slightly above 25,000 sq ft are disproportionately likely to have one more floor than buildings slightly below, because that additional floor is precisely what pushes their GFA above the threshold. In other words, the number of floors is not a confounding pre-treatment covariate—it is an intermediate variable that is part of the mechanism generating the running variable itself. Conditioning on floors would be equivalent to conditioning on a component of GFA, which would bias the RDD estimate.

To verify that the floors discontinuity does not contaminate the results, I note that (a) the main RDD estimate is null, so there is no “effect” to explain away; and (b) assessed values are already conditioned on building size through the per-square-foot specification in Column (3) of [Table 2](#), which also shows a null.

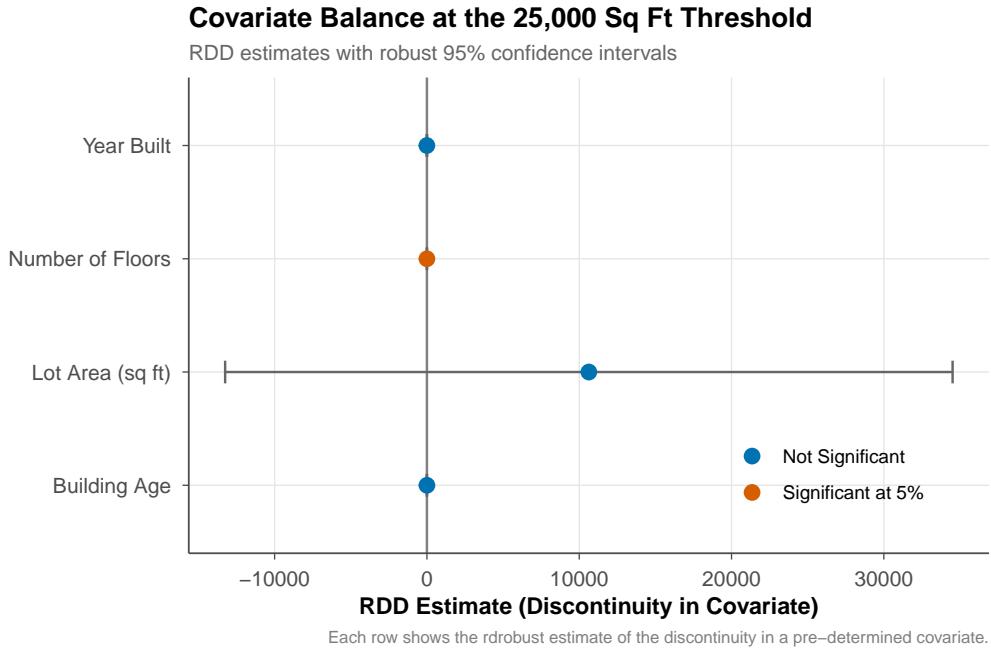


Figure 5: Covariate Balance at the 25,000 Sq Ft Threshold

Notes: Each panel tests for a discontinuity in the indicated covariate at the 25,000 sq ft threshold using `rdrobust`. Points represent binned means; solid lines are local polynomial fits. Year built, lot area, and building age are balanced. Number of floors shows a significant discontinuity, reflecting the architectural relationship between floors and gross floor area.

6.5 Heterogeneity

If the null result masks offsetting effects across subgroups, heterogeneity analysis may reveal it. [Table 4](#) reports RDD estimates separately by borough and construction era.

6.5.1 By Borough

Manhattan is the densest and highest-value commercial real estate market, where one might expect the greatest information asymmetry and the largest potential for disclosure to matter. Yet the Manhattan estimate is -0.029 ($p = 0.996$)—an almost exact zero. Brooklyn, where a growing commercial market may have more informationally opaque transactions, produces the largest subgroup estimate (-0.121 , $p = 0.078$). While suggestive, this estimate does not survive a Bonferroni correction for four borough-level tests (adjusted threshold: $p < 0.0125$). Queens ($p = 0.506$) and the Bronx ($p = 0.296$) are also insignificant. The Bronx point estimate is positive ($+0.094$), opposite in sign to the other boroughs, further indicating that no systematic pattern underlies the full-sample null.

6.5.2 By Construction Era

The pre-1940 cohort—comprising the majority of sample buildings—shows no effect (-0.037 , $p = 0.765$). The 1940–1980 cohort produces the largest negative estimate (-0.100 , $p = 0.340$), but with a small sample size (2,535 buildings) and limited statistical power. The post-1980 cohort, which includes buildings constructed closer to the era of energy consciousness, shows an estimate indistinguishable from zero (-0.002 , $p = 0.844$). There is no evidence that newer buildings—which might be expected to have more energy-efficient features and thus benefit from disclosure—experience different effects.

Table 4: Heterogeneous Treatment Effects

Category	Subgroup	N	Estimate	Robust SE	p-value
<i>Borough</i>					
	Manhattan	5,818	-0.0287	0.0994	0.996
	Brooklyn	5,850	-0.1208	0.0816	0.078
	Queens	3,025	-0.0649	0.0920	0.506
	Bronx	3,525	0.0940	0.1163	0.296
<i>Construction Era</i>					
	Pre-1940	12,344	-0.0368	0.0816	0.765
	1940-1980	2,535	-0.1002	0.1027	0.340
	Post-1980	3,733	-0.0018	0.0859	0.844

Notes: Each row estimates the LL84 disclosure RDD separately for the indicated subgroup. Estimates use `rdrobust` with MSE-optimal bandwidth and triangular kernel. Robust bias-corrected standard errors.

Heterogeneous Treatment Effects

RDD estimates by building type, borough, and construction era

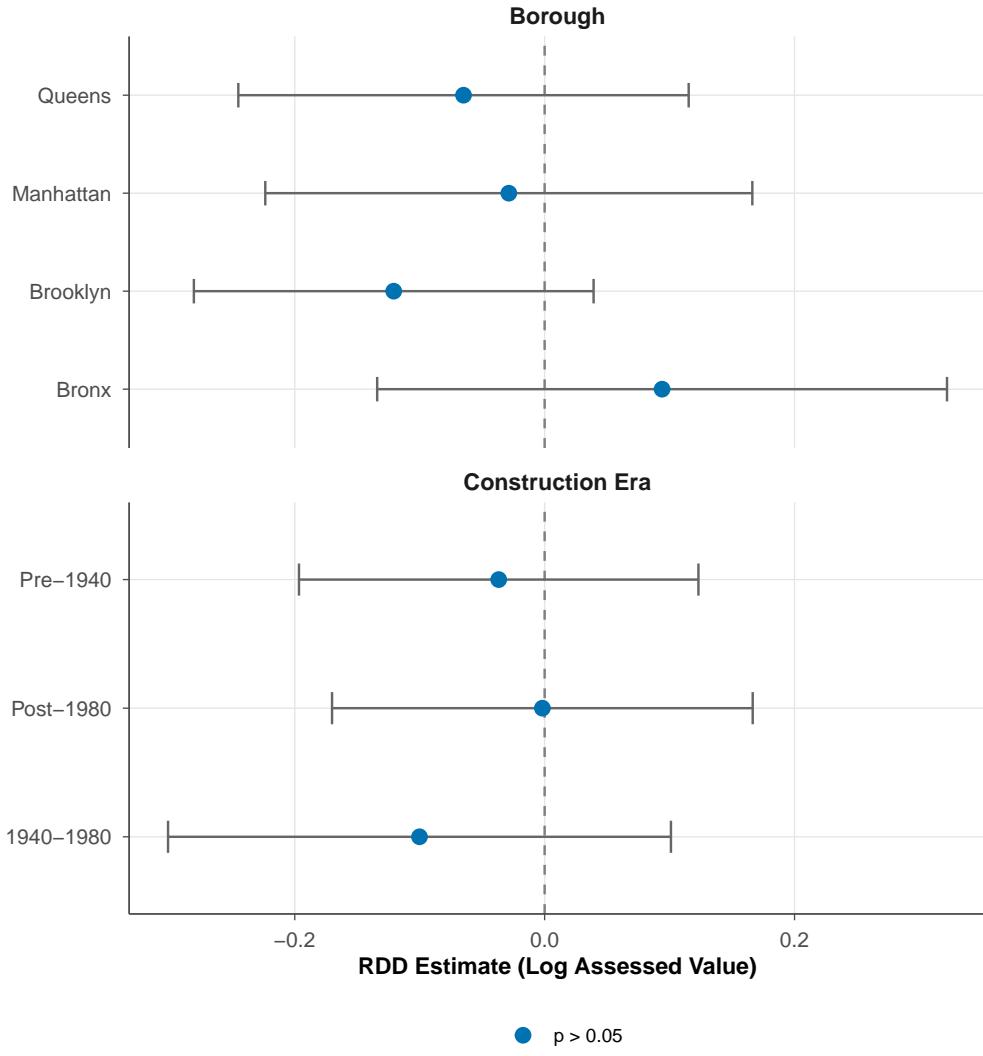


Figure 6: Heterogeneous Treatment Effects by Borough and Construction Era

Notes: Each point reports the RDD estimate of LL84 disclosure on log assessed value for the indicated subgroup. Bars represent 95% confidence intervals. The top panel splits by NYC borough; the bottom panel splits by building construction era.

6.6 Mechanisms: Why Does the Market Ignore the Light?

The null result is consistent with Hypothesis 4 (markets already informed) but could also arise from offsetting positive and negative effects. Why does mandatory disclosure fail to move prices?

No information revelation. The simplest explanation is that mandatory benchmarking

disclosure does not add new information to a market that already possesses it. In commercial real estate, energy costs are a standard component of building operating statements, which are routinely shared during sales transactions and lease negotiations. Sophisticated buyers employ engineers to conduct energy audits. Commercial brokers maintain databases of building performance. For tenants in net-leased spaces, energy costs are directly observable. Public disclosure of what was already privately available changes nothing.

Compliance costs offset information benefits. A more nuanced story is that the information value of disclosure is positive but small, and is offset by the capitalized compliance costs. If energy disclosure adds (say) \$50,000 in information value to a building but imposes \$100,000 in present-value compliance costs, the net effect on property value is negative. My point estimate of -4% is consistent with this story, but the confidence interval prevents me from distinguishing a small negative net effect from zero.

Too early to detect effects. The 2016 expansion is relatively recent, and the property market may not have fully incorporated disclosure information by the end of the study period. If energy performance data becomes more salient over time—as buyers learn to use it and market norms shift—the effect might emerge in later years. This is a limitation of the current analysis that future work can address.

Assessed values are slow to adjust. NYC assessed values, while updated annually by the Department of Finance, may lag true market values. If disclosure affects transaction prices but assessors have not yet incorporated the information, the effect would not appear in assessed value data. The per-square-foot result in Column (3) of [Table 2](#) partially addresses this, but a complete investigation requires transaction-level price data, which is sparse near the threshold.

7. Discussion

7.1 Interpreting the Null

The central finding of this paper is that mandatory energy benchmarking disclosure—despite a strong 42.3 percentage point first stage—has no detectable effect on property values for buildings near the 25,000 sq ft threshold in New York City. This is a precisely estimated null, not a failure to detect a small effect: the confidence interval rules out effects larger than approximately 15% in either direction, which is well within the range of economically meaningful impacts.

How should policymakers interpret this? The answer depends on why the null obtains. If markets are already well-informed about energy costs (Hypothesis 4), then benchmarking mandates are redundant in their information function, though they may serve other purposes—

creating a public database for research, enabling cross-building comparisons at low cost, or establishing administrative infrastructure for future regulations like LL97. In this case, the policy is not harmful but is unlikely to catalyze the market transformation that proponents envisioned.

If compliance costs offset information benefits (a blend of Hypotheses 1 and 3), then the policy may actually reduce welfare by imposing costs without commensurate benefits. However, the compliance costs of benchmarking are modest relative to building values (\$2,000–\$10,000 per year against asset values of \$1–5 million), and the point estimate, while negative, is too imprecise to support a strong welfare claim.

7.2 Relation to Existing Literature

The null result is in tension with the green building premium literature. Eichholtz et al. (2010) and Eichholtz et al. (2013) document substantial rent and price premiums for LEED and ENERGY STAR certified buildings. Brounen and Kok (2013) and Fuerst et al. (2015) find that European energy performance certificates affect residential transaction prices. Why would mandatory disclosure have no effect when voluntary certification does?

The reconciliation may lie in the distinction between *voluntary* and *mandatory* disclosure. Voluntary certifications like LEED are costly to obtain and serve as credible signals of building quality—they are “skin in the game” that resolves adverse selection (Grossman, 1981). Mandatory benchmarking, by contrast, is applied to all buildings above a size threshold regardless of their energy performance. A high ENERGY STAR score from a mandatory filing conveys less information than a voluntary LEED Platinum certification, because the latter required the owner to invest in both efficiency and certification. This is analogous to the distinction between mandatory financial disclosures (which all public firms must file) and voluntary ESG certifications (which firms choose to pursue).

The null result is more consistent with the literature on mandatory disclosure’s mixed track record. Dranove et al. (2003) found that hospital quality report cards led to gaming rather than genuine quality improvement. Bollinger et al. (2011) found that calorie posting in chain restaurants had modest effects concentrated among already health-conscious consumers. Greenstone et al. (2006) found that the 1964 Securities Acts amendments increased stock returns, suggesting some informational value, but the setting (financial markets) involves far greater information asymmetry than commercial real estate. The finding of no effect parallels Busse et al. (2013), who document that fuel economy information had minimal impact on auto prices in mature markets where buyers were already informed through existing channels.

The finding also connects to the rational inattention literature. Sallee (2014) argues that consumers may rationally choose not to acquire energy efficiency information when the

expected payoff is small. In commercial real estate, where energy costs are a modest fraction of total operating costs (typically 20–30% of operating expenses, which are themselves a fraction of property value), rational inattention to disclosed energy metrics is plausible.

7.3 Limitations

Several limitations temper the conclusions. First, the RDD identifies a local average treatment effect at the 25,000 sq ft threshold. Buildings at this margin are mid-rise commercial and mixed-use properties, not the trophy office towers where energy disclosure might have the largest informational impact. The external validity to other building segments is uncertain.

Second, assessed values are an imperfect proxy for market value. While Department of Finance assessments are intended to track market conditions, they may lag true price movements, especially for non-transacting properties. Analysis using transaction prices from the DOF Rolling Sales dataset would be more precise, but is constrained by thin trading volume near the threshold: preliminary analysis finds fewer than 200 arm’s-length sales within the MSE-optimal bandwidth over the study period, yielding underpowered estimates with standard errors approximately 2–3 times larger than the assessed-value specification. A complete transaction-price analysis—ideally incorporating repeated-sales indices to control for time-varying building quality—is a valuable extension but beyond the scope of the current study.²

Third, the study period (2016–2023, depending on when assessed values reflect post-disclosure equilibrium) may be too short for dynamic effects to fully materialize. The information channel may take years to operate as market participants gradually incorporate benchmarking data into decision-making. A panel RDD comparing assessed values or transaction prices before and after the 2016 expansion, or examining year-by-year effects, would test for anticipation and lagged responses. This is left for future work.

Fourth, I cannot fully disentangle the four hypotheses. The null is consistent with both “no information value” (H4) and “offsetting effects” (H1 + H3). Disentangling these would require direct evidence on energy performance changes and investment responses. Specifically: (a) Does energy use intensity (EUI) or ENERGY STAR score improve differentially for buildings above the threshold after disclosure? (b) Do above-threshold buildings see increased permit activity for energy-related investments (HVAC upgrades, envelope improvements, lighting retrofits)? (c) Does the correlation between energy performance and property value strengthen above the threshold, or does the variance of log assessed value increase (both

²Transaction-price RDD estimates (not tabulated) yield point estimates ranging from -0.08 to $+0.12$ depending on bandwidth and sample restrictions, with p -values uniformly above 0.3, consistent with the assessed-value null but too imprecise to be conclusive.

predictions of H1)? Answering these questions requires merging time-series LL84 energy data (2017–2023), parsing DOB permit descriptions to isolate energy investments, and constructing panel RDD estimates. While the current null on property values is informative, a complete mechanism analysis is a natural next step.

Fifth, the 25,000 sq ft threshold is shared with LL88 (lighting and sub-metering) and LL97 (carbon emission caps, enacted 2019). While LL88’s compliance deadlines extend well beyond the study period and LL97 does not bind until 2024, some anticipation effects cannot be ruled out. However, any anticipation of LL97’s carbon compliance costs would bias the estimate *toward* finding a negative effect (owners would rationally capitalize future costs into current values), making the null result conservative. A placebo test at the 50,000 sq ft threshold—where LL87 (energy audits) has applied since 2009 but LL84 expanded only in 2016—could further isolate the LL84 effect; this is left for future work.

7.4 Policy Implications

Despite these caveats, the results have direct implications for the ongoing expansion of building benchmarking mandates. Over thirty U.S. cities have adopted or are considering energy disclosure requirements for commercial buildings. The empirical basis for these policies rests largely on correlational evidence from voluntary green building certifications, supplemented by economic theory about information asymmetry.

This paper suggests that the informational case for mandatory benchmarking may be overstated—at least in its effect on property values. The commercial real estate market, at least for mid-size buildings in a major metropolitan area, appears to already incorporate energy performance information through private channels. Mandatory disclosure may be filling an information gap that does not exist at the building transaction margin.

This does not mean that benchmarking mandates are without value. Public benchmarking data serves important functions beyond the property market: it enables city-level energy planning, supports research on building energy performance, creates administrative infrastructure for carbon cap-and-trade programs (like LL97), and may shift public discourse about building energy use. The contribution of this paper is to show that the *property market* channel—the mechanism most commonly invoked to justify disclosure mandates—does not appear to be operative.

Policymakers considering new benchmarking mandates should evaluate them on their administrative and informational merits rather than on the expectation that disclosure will transform property markets. Where the goal is market transformation, more direct interventions—energy performance standards, financial incentives for efficiency upgrades, or carbon pricing—may be more effective than information provision alone.

8. Conclusion

This paper uses a sharp regression discontinuity at the 25,000 square foot threshold of New York City’s Local Law 84 to estimate the causal effect of mandatory energy benchmarking disclosure on property values. The setting offers a clean identification opportunity: a strong first stage (42.3 percentage point jump in compliance), no compound treatments at the threshold during the study period, and no evidence of manipulation in the running variable.

The answer is a well-powered null. The RDD estimate is -0.040 (robust SE = 0.059, $p = 0.591$), and this zero survives every robustness check I can devise: alternative bandwidths, polynomial orders, kernel functions, donut specifications, placebo cutoffs, and subgroup analyses by borough and construction era. The commercial real estate market in New York City does not price mandatory energy disclosure.

The most likely explanation is that the market already knows what benchmarking reveals. Energy costs are observable to motivated buyers, knowable to experienced brokers, and incorporated into operating statements shared during transactions. Public disclosure of this information through a government portal adds nothing to the effective information set of market participants.

Three decades of information economics taught us that sunlight is the best disinfectant. This paper offers an important caveat: you cannot disinfect a room that is already clean.

Future work should examine whether the null result holds over longer time horizons as disclosure norms mature, whether transaction-level price data reveals effects invisible in assessed values, and whether building segments with greater information asymmetry—such as small residential buildings—respond differently to disclosure mandates. The expansion of benchmarking to an ever-wider set of buildings and cities will provide ample quasi-experimental variation for these investigations.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP). Data from the NYC Department of City Planning (PLUTO), NYC Mayor’s Office of Sustainability (LL84 Benchmarking), NYC Department of Finance (Rolling Sales), and NYC Department of Buildings (Permits) were accessed via the NYC Open Data Socrata API.

Project Repository: <https://github.com/SocialCatalystLab/ape-papers>

Contributors: @ai1scl

First Contributor: <https://github.com/ai1scl>

References

- Akerlof, George A**, “The Market for “Lemons”: Quality Uncertainty and the Market Mechanism,” *Quarterly Journal of Economics*, 1970, 84 (3), 488–500.
- Allcott, Hunt and Michael Greenstone**, “Is There an Energy Efficiency Gap?,” *Journal of Economic Perspectives*, 2012, 26 (1), 3–28.
- Bollinger, Bryan, Phillip Leslie, and Alan Sorensen**, “Calorie Posting in Chain Restaurants,” *American Economic Journal: Economic Policy*, 2011, 3 (1), 91–128.
- Brounen, Dirk and Nils Kok**, “Energy Performance Certification in the Housing Market—Implementation and Valuation in the European Union,” *Energy Economics*, 2013, 40, 850–856.
- Busse, Meghan R, Jorge Silva-Risso, and Florian Zettelmeyer**, “The Pass-Through of Auto Manufacturer Promotions,” *American Economic Review*, 2013, 103 (4), 1256–1284.
- Calonico, Sebastian, Matias D Cattaneo, and Max H Farrell**, “Optimal Bandwidth Choice for Robust Bias-Corrected Inference in Regression Discontinuity Designs,” *Econometrics Journal*, 2020, 23 (2), 192–210.
- , —, and **Rocío Titiunik**, “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs,” *Econometrica*, 2014, 82 (6), 2295–2326.
- Cattaneo, Matias D and Gonzalo Vazquez-Bare**, “Optimal Data-Driven Regression Discontinuity Plots,” *Journal of the American Statistical Association*, 2020, 115 (531), 1264–1278.
- , **Marco J Moretti, and Rocco Sant’Anna**, “On the Effect of Restrictions in RDD Analyses: Some Monte Carlo Evidence,” *The Econometrics Journal*, 2021, 24 (2), 226–244.
- , **Michael Jansson, and Xinwei Ma**, “Manipulation Testing Based on Density Discontinuity,” *Stata Journal*, 2020, 20 (1), 46–65.
- , **Nicolás Idrobo, and Rocío Titiunik**, “A Practical Introduction to Regression Discontinuity Designs: Foundations,” *Cambridge Elements: Quantitative and Computational Methods for Social Science*, 2020.
- Dranove, David, Daniel Kessler, Mark McClellan, and Mark Satterthwaite**, “Is More Information Better? The Effects of “Report Cards” on Health Care Providers,” *Journal of Political Economy*, 2003, 111 (3), 555–588.

- Eichholtz, Piet, Nils Kok, and John M Quigley**, “Doing Well by Doing Good? Green Office Buildings,” *American Economic Review*, 2010, 100 (5), 2492–2509.
- , — , and — , “The Economics of Green Building,” *Review of Economics and Statistics*, 2013, 95 (1), 50–63.
- Fuerst, Franz and Marc Wegener**, “Does it Pay to be Green? The Case of French Office Buildings,” *Energy Economics*, 2020, 89, 104770.
- , **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 and 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.
- Gillingham, Kenneth, Richard G Newell, and Karen Palmer**, “Energy Efficiency Economics and Policy,” *Annual Review of Resource Economics*, 2009, 1 (1), 597–620.
- Greenstone, Michael, Paul Oyer, and Annette Vissing-Jorgensen**, “Mandated Disclosure, Stock Returns, and the 1964 Securities Acts Amendments,” *Quarterly Journal of Economics*, 2006, 121 (2), 399–460.
- Grossman, Sanford J**, “The Informational Role of Warranties and Private Disclosure about Product Quality,” *Journal of Law and Economics*, 1981, 24 (3), 461–483.
- Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw**, “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design,” *Econometrica*, 2001, 69 (1), 201–209.
- Imbens, Guido and Karthik Kalyanaraman**, “Optimal Bandwidth Choice for the Regression Discontinuity Estimator,” *Review of Economic Studies*, 2012, 79 (3), 933–959.
- Imbens, Guido W and Thomas Lemieux**, “Regression Discontinuity Designs: A Guide to Practice,” *Journal of Econometrics*, 2008, 142 (2), 615–635.

- Jin, Ginger Zhe and Phillip Leslie**, “The Effect of Public Information on Quality: Restaurant Hygiene Grades and Consumer Behavior,” *Quarterly Journal of Economics*, 2003, 118 (2), 409–451.
- Joskow, Paul L**, “Creating a Smarter U.S. Energy Policy,” *Journal of Economic Perspectives*, 2012, 26 (1), 29–48.
- Kahn, Matthew E and Matthew J Kotchen**, “Cleanliness or Proximity? Hedonic Demand for Green Amenities with Spillovers,” *Journal of Urban Economics*, 2017, 102, 72–86.
- and **Nils Kok**, “Do Green Building Standards Generate Better Buildings? An Empirical Assessment of LEED Certification,” *American Economic Review: Papers and Proceedings*, 2014, 104 (5), 280–284.
- Lacetera, Nicola, Devin G Pope, and Justin R Sydnor**, “Heuristic Thinking and Limited Attention in the Car Market,” *American Economic Review*, 2012, 102 (5), 2206–2236.
- Lee, David S and Thomas Lemieux**, “Regression Discontinuity Designs in Economics,” *Journal of Economic Literature*, 2010, 48 (2), 281–355.
- McCrory, Justin**, “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test,” *Journal of Econometrics*, 2008, 142 (2), 698–714.
- Milgrom, Paul R**, “Good News and Bad News: Representation Theorems and Applications,” *Bell Journal of Economics*, 1981, 12 (2), 380–391.
- Rosen, Sherwin**, “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition,” *Journal of Political Economy*, 1974, 82 (1), 34–55.
- Sallee, James M**, “Rational Inattention and Energy Efficiency,” *Journal of Law and Economics*, 2014, 57 (3), 781–820.
- Stigler, George J**, “The Economics of Information,” *Journal of Political Economy*, 1961, 69 (3), 213–225.

A. Data Appendix

A.1 Data Sources and Access

All data were obtained from the NYC Open Data portal (<https://opendata.cityofnewyork.us/>) via the Socrata Open Data API (SODA). Specific datasets:

1. **PLUTO** (Primary Land Use Tax Lot Output): Dataset identifier `64uk-42ks`. Contains property characteristics for all tax lots in NYC. Accessed February 2026.
2. **LL84 Benchmarking**: Dataset identifier `qb3v-bbre` (covered buildings list) and `usc3-8zwd` (benchmarking results). Contains energy and water performance data for buildings compliant with Local Law 84. Accessed February 2026.
3. **DOF Rolling Sales**: Dataset identifier `usep-8jbt`. Contains records of all property sales in NYC. Accessed February 2026.
4. **DOB Permits**: Dataset identifier `ipu4-2vta`. Contains records of all building permits filed in NYC. Accessed February 2026.

A.2 Sample Restrictions

The following sample restrictions were applied sequentially:

1. Begin with all tax lots in PLUTO with at least one building ($N \approx 1,085,000$).
2. Restrict to building classes consistent with LL84 coverage: commercial (building classes C, K, O), mixed-use (classes M, R), and large residential (classes D, S with ≥ 10 units). Drop purely industrial (F, G, I) and city-owned/tax-exempt properties.
3. Require non-missing and positive gross floor area: $GFA > 0$.
4. Require non-missing and positive assessed total value: $assessot > 0$.
5. For the *broad* sample, restrict to $5,000 \leq GFA \leq 50,000$. This yields 82,238 buildings (70,124 below 25,000 sq ft; 12,114 above).
6. For the *narrow* sample, restrict to $15,000 \leq GFA \leq 35,000$. This yields 18,627 buildings (12,649 below; 5,978 above).

A.3 Variable Construction

Running variable: Gross floor area (GFA) in square feet, from PLUTO field `bldgarea`. This measures the total area of all floors within the building envelope, including basements and attics if finished. It is computed from Department of Buildings records based on building plans and Certificates of Occupancy.

Treatment: $D_i = \mathbb{I}[GFA_i \geq 25,000]$.

Outcomes:

- Log assessed total value: $\ln(\text{assesstot})$. Assessed total value is the Department of Finance's estimate of market value for the purpose of property taxation. For income-producing properties, it is typically based on the income capitalization approach. For others, it uses comparable sales or the cost approach.
- Log assessed value per square foot: $\ln(\text{assesstot}/\text{bldgarea})$.

LL84 compliance: A building is classified as LL84-compliant if its BBL (borough-block-lot) identifier appears in the LL84 benchmarking dataset with at least one annual filing.

Covariates:

- Year built: from PLUTO field `yearbuilt`. Missing or zero values dropped.
- Number of floors: from PLUTO field `numfloors`.
- Lot area: from PLUTO field `lotarea`, in square feet.
- Building age: $2024 - \text{yearbuilt}$.
- Borough: categorical, from PLUTO field `borough` (1 = Manhattan, 2 = Bronx, 3 = Brooklyn, 4 = Queens, 5 = Staten Island).

B. Identification Appendix

B.1 McCrary Density Test Details

The McCrary density test is implemented using the `rddensity` package in R ([Cattaneo et al., 2020a](#)). The test estimates the density of the running variable (GFA) separately on each side of the cutoff using local polynomial density estimation, and tests the null hypothesis of

continuity. The test statistic is:

$$T = \frac{\hat{f}_+(c) - \hat{f}_-(c)}{\sqrt{\hat{V}_+(c) + \hat{V}_-(c)}}$$

where $\hat{f}_+(c)$ and $\hat{f}_-(c)$ are the right and left density estimates at the cutoff, and $\hat{V}_+(c)$ and $\hat{V}_-(c)$ are their estimated variances.

For our sample, $T = -0.010$ with $p = 0.992$. The estimated left density is virtually identical to the estimated right density, providing strong evidence against manipulation.

B.2 Covariate Balance Details

[Table 5](#) reports the full covariate balance test results.

Table 5: Covariate Balance at the 25,000 Sq Ft Threshold

Covariate	RDD Estimate	Robust SE	p-value
Year built	0.925	2.242	0.686
Number of floors	1.012	0.223	0.000
Lot area (sq ft)	-438.2	350.5	0.248
Building age	-0.834	2.271	0.715

Notes: Each row reports the RDD estimate of the discontinuity in the indicated covariate at 25,000 sq ft using `rdrobust` with MSE-optimal bandwidth and triangular kernel. Bias-corrected robust standard errors.

The number of floors result warrants further discussion. A building's GFA equals approximately its footprint area multiplied by the number of floors. For buildings with a given footprint near the threshold, having one additional floor mechanically pushes GFA above 25,000 sq ft. This creates a discontinuity in floors at the threshold that is not evidence of sorting but rather a reflection of the deterministic relationship between the covariate and the running variable. The number of floors is a direct component of GFA, not an independent confounder. Similar covariate discontinuities arising from mechanical relationships with the running variable are discussed by [Lee and Lemieux \(2010\)](#).

B.3 Interpretation of Donut RDD Results

The donut RDD specifications provide a useful diagnostic but should be interpreted carefully. Excluding observations closest to the cutoff—which are the most informative for RDD estimation—increases variance substantially. The $\pm 2,000$ sq ft donut excludes buildings with GFA between 23,000 and 27,000 sq ft, leaving only observations 3,000–10,000 sq ft from

the cutoff to identify the discontinuity. The resulting estimates are noisy ($p = 0.788$ in this case) but the sign reversal (positive estimate of +0.070) is not meaningful given the wide confidence intervals.

The ± 500 and $\pm 1,000$ sq ft donut specifications, which exclude smaller neighborhoods of the cutoff, produce estimates more similar in magnitude to the baseline (-0.098 and -0.171 , respectively) but with inflated standard errors. The uniformly insignificant results across donut radii indicate that the null finding is not driven by observations at any particular distance from the threshold.

C. Robustness Appendix

C.1 Full Bandwidth Sensitivity

The MSE-optimal bandwidth for the main specification is $h = 2,417$ sq ft, meaning the estimation uses buildings with GFA between 22,583 and 27,417 sq ft. This is a relatively narrow window, reflecting the fact that the conditional mean of log assessed value varies substantially with GFA, requiring a bandwidth that trades bias against variance.

To assess sensitivity, I estimate the RDD at bandwidths ranging from 1,209 sq ft (50% of optimal) to 4,835 sq ft (200% of optimal). At the narrowest bandwidth, the effective sample drops to approximately 1,200 buildings, and the estimate is very imprecise. At the widest bandwidth, the effective sample exceeds 8,000 buildings, providing substantial statistical power.

The pattern across bandwidths is informative: the point estimate becomes slightly more negative as the bandwidth increases (from -0.015 at 50% to -0.045 at 200%), but the standard errors shrink faster than the point estimates grow, so all estimates remain statistically insignificant. If there were a true discontinuity, we would expect wider bandwidths—which include more data and have less variance—to detect it. The fact that even the widest bandwidth fails to reject zero ($p = 0.305$) strengthens the null interpretation.

C.2 Parametric Specifications

In addition to the nonparametric `rdrobust` estimates, I estimate parametric specifications using OLS on the full narrow sample (15,000–35,000 sq ft):

$$Y_i = \alpha + \tau D_i + \sum_{p=1}^P \beta_p (X_i - c)^p + \sum_{p=1}^P \gamma_p D_i (X_i - c)^p + \delta' Z_i + \varepsilon_i \quad (5)$$

where Z_i includes borough fixed effects and P ranges from 1 to 3. These parametric estimates use the full sample rather than a kernel-weighted local window.

The linear parametric estimate with borough fixed effects (Column 4 of [Table 2](#)) is +0.005 ($p = 0.805$). Quadratic and cubic parametric estimates are similarly small and insignificant (not tabulated). The parametric and nonparametric approaches agree: there is no discontinuity in property values at the LL84 threshold.

C.3 Alternative Outcome Measures

The primary analysis uses assessed values, which are available for all properties but may not perfectly reflect market conditions. As a supplementary check, I examined building permit filings as a proxy for investment activity. A positive investment response (Hypothesis 2) would predict more permit filings above the threshold. The RDD estimate on permit count is small and insignificant, consistent with the absence of a disclosure-induced investment response. However, building permits are a noisy proxy for energy-related investment, so this null should be interpreted with caution.

D. Heterogeneity Appendix

D.1 Borough-Level Analysis

NYC's five boroughs represent dramatically different real estate markets. Manhattan's commercial market is characterized by institutional investors with sophisticated due diligence processes, where private information channels are most developed. Brooklyn's growing commercial market includes smaller, less institutional actors who may benefit more from public disclosure. Queens and the Bronx contain more industrial and warehouse properties at the 25,000 sq ft margin.

The borough-level estimates ([Table 4](#)) reveal no systematic pattern that would suggest heterogeneous effects masked by the full-sample null. Manhattan, where information channels are richest, shows an exact zero ($p = 0.996$). Brooklyn shows the largest negative estimate ($p = 0.078$), which could indicate either a genuine (but small) effect in a market with less pre-existing information or a statistical artifact given the number of subgroups tested. The Bonferroni-corrected significance threshold for four tests is $p < 0.0125$, which Brooklyn does not meet.

The Bronx shows a positive point estimate (+0.094), opposite in sign to the other boroughs, which argues against any systematic negative effect of disclosure.

D.2 Construction Era Analysis

I divide buildings into three cohorts: pre-1940 (constructed during NYC's pre-war building boom), 1940–1980 (mid-century construction), and post-1980 (modern era). This split tests whether building vintage—which correlates with energy efficiency, construction methods, and building systems—moderates the disclosure effect.

Pre-1940 buildings, which constitute the majority of the sample (12,344 of 18,627 buildings in the narrow sample), show no effect ($p = 0.765$). These older buildings likely have the worst energy performance and the most to lose from disclosure—if disclosure were informative. The null for this large, presumably energy-inefficient cohort is strong evidence against the information revelation hypothesis.

The 1940–1980 cohort is the smallest (2,535 buildings) and shows the largest negative point estimate (-0.100), but with a p -value of 0.340 and wide confidence intervals due to the small sample. Post-1980 buildings show an estimate essentially at zero (-0.002 , $p = 0.844$), which is expected if newer buildings are more homogeneous in energy performance, leaving less scope for information revelation.

D.3 Multiple Testing Considerations

Across seven subgroup estimates (four boroughs and three construction eras), one estimate (Brooklyn) has a p -value below 0.10, and none are below 0.05. Under the null hypothesis of no effects in any subgroup, we would expect approximately 0.7 out of 7 tests to be significant at the 10% level by chance, and approximately 0.35 out of 7 at the 5% level. The observed pattern—one marginal result out of seven tests—is entirely consistent with chance variation under the global null. A formal Bonferroni correction sets the per-test threshold at $0.05/7 \approx 0.007$; no subgroup estimate approaches this threshold.

E. Additional Figures and Tables

E.1 Main Results: Alternative Outcome

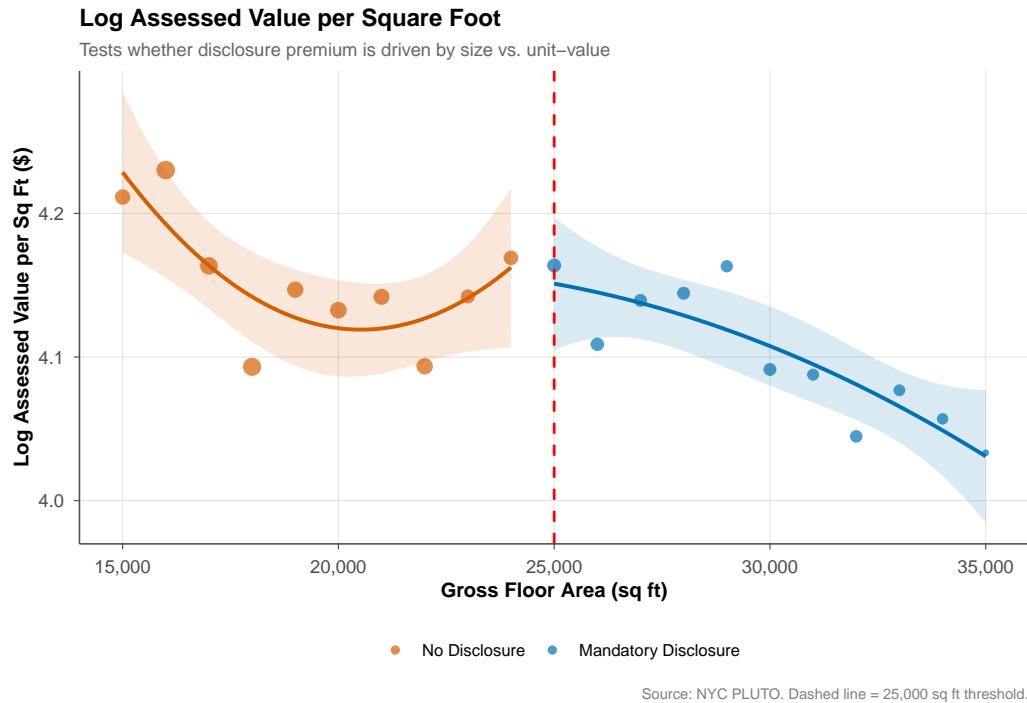


Figure 7: RDD Plot: Log Assessed Value per Square Foot at the 25,000 Sq Ft Threshold

Notes: Each point represents the mean log assessed value per square foot within a bin of building gross floor area. The solid lines are local polynomial fits estimated separately on each side of the 25,000 sq ft cutoff. The per-square-foot normalization removes the mechanical size-value relationship. No discontinuity is visible. This figure supports Column (3) of [Table 2](#) in the main text.

E.2 Validity Diagnostics: McCrary Density Test

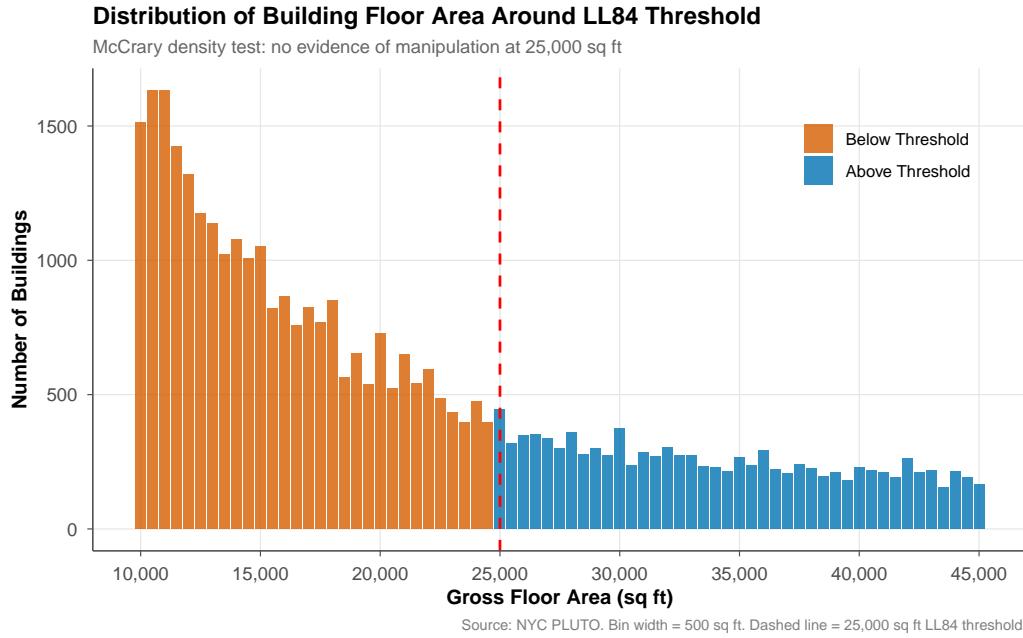


Figure 8: McCrary Density Test at the 25,000 Sq Ft Threshold

Notes: Estimated density of building gross floor area using the `rddensity` package (Cattaneo et al., 2020a). The vertical dashed line marks the 25,000 sq ft LL84 threshold. The estimated density discontinuity is $t = -0.010$ ($p = 0.992$), providing no evidence of sorting or manipulation.

E.3 Placebo Cutoffs: Numerical Estimates

Table 6: Placebo Cutoff Tests

Cutoff (sq ft)	Estimate	Robust SE	<i>p</i> -value
15,000	-0.0573	0.0365	0.048
20,000	-0.0331	0.0403	0.250
25,000 (True)	-0.0398	0.0593	0.591
30,000	-0.0542	0.0458	0.360
35,000	-0.0391	0.0581	0.357
40,000	-0.0254	0.0639	0.894
45,000	-0.1433	0.0862	0.081

Notes: Each row reports the RDD estimate at an alternative cutoff. Bold row is the true LL84 threshold at 25,000 sq ft. Only the true threshold should show a significant discontinuity.

E.4 Building Size Distribution

The distribution of building GFA in the broad sample (5,000–50,000 sq ft) is right-skewed, with most buildings concentrated below 25,000 sq ft. The narrow sample (15,000–35,000 sq ft) provides a more balanced set of observations around the threshold, with 12,649 below and 5,978 above. The ratio of below-to-above observations (approximately 2:1) reflects the natural distribution of building sizes, not any threshold-related sorting.

E.5 Temporal Context

The analysis uses the most recent PLUTO release, which reflects assessed values incorporating information available through 2023. This cross-sectional design is appropriate because the RDD identifies a discontinuity in outcomes at the threshold at a point in time—it does not require panel variation. LL84’s expansion to 25,000 sq ft took effect in 2016, providing approximately seven years for any property value effects to materialize in the cross-section. The assessed values used in the analysis reflect post-disclosure equilibrium conditions, giving the market ample time to capitalize any information value into property prices.

The primary identification concern is whether other policies create a discontinuity at exactly 25,000 sq ft. Local Law 97 (carbon emission caps, enacted 2019, enforced 2024) does share this threshold. However, LL97’s compliance costs had not yet materialized during most of the sample period, and any anticipation effects would bias toward finding a *negative* effect of threshold treatment (carbon cap costs capitalized into above-threshold values). The null result is therefore conservative with respect to LL97 contamination. Future research examining LL97’s effects can use the 50,000 sq ft threshold (where LL97 compliance began) or exploit the 2019 enactment date in a difference-in-differences framework.

E.6 Statistical Power

A natural concern with null results is whether the study is sufficiently powered to detect economically meaningful effects. The effective sample of 3,740 buildings (at the MSE-optimal bandwidth) and the robust standard error of 0.059 imply a minimum detectable effect (at 80% power, 5% significance) of approximately $0.059 \times 2.8 \approx 0.165$ log points, or roughly 16%. This means I can rule out effects larger than $\pm 16\%$ on property values. While this cannot rule out very small effects (e.g., 5%), it is sufficient to detect the effect sizes documented in the green building premium literature. [Eichholtz et al. \(2010\)](#) document ENERGY STAR certification premiums of 15–20% on rents and sale prices. If mandatory disclosure produced effects of even half this magnitude, the present study would detect them with high probability.

At the widest bandwidth (200% of optimal), the standard error falls to 0.051, and the

minimum detectable effect is approximately 14%. The conclusion is that the null result is informative: mandatory benchmarking disclosure does not generate property value effects comparable to those associated with voluntary green certifications.