Building Energy Performance Standards: Impacts on Building Energy Efficiency and GHG Emissions in Washington, DC

Tosin Kolajo Gbadegesin

Ph.D. Economics Candidate

Primary Fields: Energy and Environmental Economics

Secondary Fields: Urban and Labor Economics

tosin.gbadegesin@bison.howard.edu | tosingbade05@gmail.com

Job market website: tosinsdgs.github.io

Department of Economics, Howard University, Washington DC

This paper is Chapter 3 of my Ph.D. dissertation. It has been submitted to **Energy Policy** *and is currently under review.*

Abstract

This study evaluates the causal impact of Washington, DC's Building Energy Performance Standards (BEPS) on building energy efficiency and GHG emissions using a continuous DiD framework. Leveraging panel data on large public and private buildings from 2013–2023, the analysis exploits variation in treatment intensity measured by each building's pre-policy compliance gap relative to BEPS thresholds. Results show that a one-unit increase in the compliance gap reduces Site EUI by 0.33 kBtu/ft^2 and Source EUI by 0.45 kBtu/ft^2, increases Energy Star scores by 0.41 points, and lowers total GHG emissions by 0.35 metric tons of CO_2e and emissions intensity by 0.65 kgCO_2e/ft^2. Event-study evidence confirms parallel trends and shows that effects intensify after the first binding compliance cycle in 2021. The findings indicate that BEPS delivers meaningful efficiency gains and emissions reductions, driven primarily by baseline noncompliance rather than building ownership status.

Keywords: Compliance Gap; Continuous DiD; Energy Efficiency; GHG Emissions; Energy Use Intensity

1.0 Introduction

The accelerating urgency of climate change and the need to reduce carbon emissions have intensified policy efforts to improve energy efficiency and sustainability in the built environment. Buildings are central to this challenge, accounting for over 40% of total energy use and greenhouse gas (GHG) emissions in Europe and roughly 40% of total energy consumption and 30% of emissions in the United States (European Commission 2024; U.S. Department of Energy 2021). In response, governments worldwide have increasingly adopted performance-based standards and building codes aimed at decarbonizing the sector. Yet despite their growing prevalence, credible ex post evidence on the effectiveness of such policies, particularly for existing buildings, which account for the bulk of urban energy use and emissions, remains limited.

This study evaluates the causal impacts of Washington, DC's Building Energy Performance Standards (BEPS) on building-level energy efficiency and environmental outcomes. It addresses three central questions. First, what is the causal effect of BEPS on building-level energy efficiency outcomes? Second, to what extent does BEPS reduce total GHG emissions and emissions intensity among covered buildings? Third, how do these impacts vary across public and private buildings? Together, these questions assess whether mandatory performance-based regulation delivers measurable energy and emissions reductions in existing buildings and whether building ownership shapes policy effectiveness.

To situate BEPS within the broader policy landscape, many jurisdictions have pursued performance-based approaches to building decarbonization. In the European Union, the Energy Performance of Buildings Directive mandates nearly-zero energy buildings for new construction and is supported by Minimum Energy Performance Standards targeting the worst-performing structures for renovation (Sunderland and Santini 2020). Similar initiatives, including Energiesprong in the Netherlands, the Building Research Establishment Environmental Assessment Method in the United Kingdom, and the Global Sustainability Assessment System in Qatar, illustrate the global shift toward enforceable performance standards (Madigan 2025).

In the United States, early policy efforts focused primarily on benchmarking and disclosure rather than mandatory performance thresholds. Benchmarking programs such as New York City's Local Law 84 and Chicago's energy rating system have been associated with modest reductions in energy consumption and improvements in market valuation for efficient buildings (Kontokosta 2013; Hsu 2014a). However, evidence suggests that while informational policies can yield energy savings of 3–8% over two to four years, they often fail to overcome structural and behavioral barriers to deeper efficiency improvements (Hsu 2014b; Mims et al. 2017). These limitations have motivated a transition toward mandatory performance-based regulation.

Washington, DC has been at the forefront of this transition. The Clean and Affordable Energy Act (CAEA) of 2008 introduced mandatory benchmarking and disclosure for large buildings but did not impose enforceable efficiency requirements (District of Columbia Council 2008). Recognizing these constraints, the Clean Energy DC Omnibus Act (CEOA) of 2018 established the Building Energy Performance Standards (BEPS), requiring existing buildings to meet minimum performance thresholds or pursue prescriptive compliance pathways within multi-year compliance cycles (District of Columbia Council 2018; DOEE 2019; DOEE 2021a). BEPS is implemented in three sequential cycles that progressively expand coverage by building size, ultimately encompassing nearly the entire large-building stock in the District (DOEE 2021a). By mandating measurable performance outcomes, BEPS represents a fundamental shift from voluntary transparency to enforceable regulation designed to accelerate retrofits, promote clean technologies, and reduce emissions (DOEE 2021b).

Against this policy backdrop, this study adopts a continuous-treatment difference-in-differences (DiD) framework in which treatment intensity is defined by buildings' pre-policy compliance gaps rather than a binary treated–control distinction. The empirical strategy accounts for potential non-random attrition using inverse probability weighting and assesses robustness across alternative gap constructions based on property-type means, property-type percentiles, and citywide benchmarks. While prior research has largely relied on modeling and simulation to assess building decarbonization pathways (e.g., Andrews and Jain 2023; Webb and McConnell 2023; Palmer and Walls 2017; Asensio and Delmas 2017), this study provides rare ex post causal evidence on the realized impacts of mandatory building performance standards. Event-study models are used to test the parallel trends in intensity assumption and to trace the dynamic evolution of BEPS effects over time, while heterogeneity analyses examine differences across public and private buildings.

The remainder of the paper is organized as follows. Section 2 reviews the theoretical and empirical literature on benchmarking and performance standards in the building sector. Section 3 describes the data sources and empirical methodology, outlining the continuous DiD framework and identification strategy. Section 4 presents the empirical findings, including baseline estimates, dynamic event-study results, heterogeneity analyses, and robustness checks. Section 5 concludes by summarizing the key results and discussing their implications for building energy policy and the design of performance-based standards in urban settings.

2.0 Literature Review

2.1 Theoretical Foundation of Benchmarking and BEPS

A core motivation for benchmarking policies lies in closing the energy efficiency gap, defined as the under-adoption of cost-effective energy-saving measures (Palmer & Walls, 2015). Several factors contribute to this gap, including split incentives (where building owners do not benefit from tenants' reduced utility bills) and information asymmetry (where energy consumption data are not transparent to prospective tenants, buyers, or investors). By requiring building owners to collect and disclose energy performance information, benchmarking policies aim to correct these market failures (Palmer & Walls 2017).

The theoretical foundation rests on information economics and behavioral response theory. When credible information on building performance becomes publicly available, market participants adjust their choices, generating reputational and financial incentives for energy efficiency (Allcott and Greenstone 2012). Benchmarking thus serves both as a transparency mechanism and a behavioral nudge, encouraging building owners to invest in energy-saving measures even without direct mandates. Over time, greater transparency can also promote competitive differentiation, allowing efficient buildings to command higher rents, improved occupancy rates, and stronger investor interest (Palmer and Walls 2015).

Such transparency can theoretically lead to market-based rewards for efficient buildings through higher occupancy rates, increased property values, and stronger investor interest (Palmer & Walls, 2015). In parallel, policy frameworks such as building codes, labeling programs, and performance standards help ensure that minimum efficiency requirements are met (Laustsen 2008). Building codes typically apply to new construction, but older structures remain a challenge because of their high energy consumption and the lack of enforceable requirements for retrofits (EPA 2014). The introduction of Building Energy Performance Standards (BEPS) addresses this gap by mandating improvements in underperforming existing buildings, effectively complementing the informational function of benchmarking with a regulatory "stick" (Palmer & Walls 2017).

Conceptually, BEPS can be viewed as a hybrid policy instrument that bridges informational and command-and-control approaches. It internalizes energy externalities by imposing quantifiable performance thresholds while preserving flexibility in compliance—allowing owners to choose between operational upgrades and prescriptive pathways (Gillingham and Palmer 2014). This framework aligns with broader insights from building energy performance research emphasizing the need for robust assessments that account for dynamic thermal properties and regional climates (Vollaro et al. 2015; Lam et al. 2008). Ultimately, benchmarking and BEPS policies enhance social welfare by mitigating informational inefficiencies, stimulating retrofit investments, and aligning private incentives with public climate and decarbonization goals (Aldy and Stavins 2012).

2.2 Empirical Evidence on Benchmarking Policies and BEPS

A number of studies provide empirical support for the effectiveness of benchmarking and disclosure laws in driving moderate energy savings. For example, Palmer and Walls (2015) note that U.S. cities with benchmarking ordinances observe 3–8% reductions in building energy use over a two- to four-year period, attributable to increased awareness and the reputational effects of public disclosure. New York City's Local Law 84 offers a notable case study, where buildings subject to benchmarking requirements realized 5.7% lower weather-normalized source energy use within three years, coupled with an 8.3% reduction in GHG emissions (Kontokosta 2014).

Despite these gains, studies caution that benchmarking alone may not deliver the deep energy savings necessary to meet aggressive climate goals (Palmer & Walls, 2017). Buildings often exhibit complex, climate-dependent energy demands (Lam et al. 2008), and a voluntary or informational approach may not overcome persistent financial, technical, or behavioral barriers. As a result, the shift toward BEPS in Washington, DC and other jurisdictions represents a more directive policy mechanism that can spur retrofits in the least-efficient segment of the building stock (Palmer & Walls, 2015). The expectation is that by coupling benchmarking data with mandatory performance thresholds, BEPS policies can achieve significantly larger energy reductions compared to transparency measures alone (Palmer & Walls, 2017).

BEPS programs extend the logic of benchmarking by requiring buildings that fall below a designated energy performance threshold to undertake upgrades or face penalties (Palmer & Walls, 2017). In Washington, DC, for instance, buildings that fail to meet these standards are placed on a compliance pathway, during which they must implement efficiency measures or otherwise demonstrate improvement. This approach seeks to accelerate the rate of energy retrofits across a large share of the building stock, addressing the older, more energy-intensive structures that often dominate urban environments (EPA 2014; CBI 2012).

Comparative studies emphasize that building energy performance is influenced by a variety of factors, including construction practices, climatic conditions, and building age (Lam et al. 2008; Vollaro et al., 2015). Hence, successful BEPS implementation often requires flexibility to accommodate different property types and local conditions. In line with this, the Clean Energy Omnibus Act of 2018 includes compliance pathways such as prescriptive upgrades or a target percentage reduction in energy use to account for the unique circumstances of each building (Palmer & Walls, 2017). By integrating benchmarking data with dynamic performance requirements, Washington, DC's BEPS framework exemplifies a policy design that is responsive to diverse building conditions, while ensuring tangible progress toward efficiency goals.

Recent research highlights the growing interplay between performance standards and building energy labeling programs worldwide. In the European Union, Energy Performance Certificate (EPC) schemes have become instrumental in driving demand for energy-efficient properties, although challenges persist regarding data quality and uniform implementation (Li et al. 2019). Similar labeling efforts in Singapore emphasize a rigorous benchmarking database and independent audits by accredited Energy Service Companies, resulting in the Energy Smart Office Label for top-performing buildings (Lee & Rajagopalan 2008). In Brazil, voluntary labeling schemes were introduced for residential, commercial, and service buildings, aiming to inform consumer choice and encourage more efficient design (Fossati et al. 2016).

Alongside these national programs, studies including Goldstein & Eley (2014) have examined how performance indices (e.g., asset versus operational ratings) can better inform building owners, operators, and policymakers about both the intrinsic efficiency of a structure and its real-world operational management. These initiatives often operate within a broader policy environment that mixes regulatory mandates and voluntary measures, where cost-effectiveness, enforcement, and stakeholder engagement remain central considerations (Lee & Yik 2004; Sun et al. 2016). Furthermore, the advent of big data approaches, such as the consolidation of large-scale building energy datasets, is enhancing peer group analysis and empirical methods for evaluating retrofit impacts and performance outcomes (Mathew et al. 2015).

Beyond demonstrating the potential for energy savings in the short term, recent literature has underscored the importance of standardizing methods for data collection, reporting, and evaluation to strengthen the long-term efficacy of BEPS. In a review of 24 state and local jurisdictions, Mims et al. (2017) find that most benchmarking and transparency programs yield energy reductions between 3% and 8% over a two- to four-year period, yet the diversity of data collection practices and analytical methods complicates definitive comparisons of policy outcomes. Similar challenges appear in jurisdictions like China and Europe, where divergent building standards and limited data accessibility hamper rigorous cross-study evaluations (Zhang et al. 2017). Additionally, the question of how much information is necessary to spur meaningful energy improvements remains pivotal, with Hsu (2014b) arguing that building-level benchmarking data alone often outperforms more detailed engineering audits in predicting energy use intensity.

Recent analyses also highlight the emerging role of emissions-based performance standards, showing that combining annual GHG targets with peak-load flexibility requirements can drive 89% overall reductions in building emissions for certain U.S. cities (Andrews & Jain, 2023). Studies of building energy data further demonstrate that performance improvements vary by building size, type, and operational patterns (Papadopoulos et al. 2018), underscoring the necessity for tailored compliance pathways and robust enforcement to achieve substantial and enduring emissions cuts (Hicks & Clough 1998; Webb & McConnell 2023). Asensio and Delmas (2017) also show that even high-profile labeling and certification programs may fail to capture significant savings in small and medium buildings, underscoring a gap that BEPS policies must address through carefully structured mandates and incentives. Lastly, Cohen and Bordass (2015) advocate for operational ratings that focus on actual in-use performance rather than purely asset-based assessments, a perspective that aligns with the push toward standardized operational data (Mims et al. 2017) and highlights how BEPS can evolve from static benchmarks to dynamic, outcome-focused regulation.

Scholarly evidence underscores the critical role of occupant behavior and actual operating conditions in achieving modeled energy savings, suggesting that BEPS must account for these real-world dynamics. McCoy et al. (2018) find that simulated energy usage often overestimates actual consumption in newly constructed green homes, while being less accurate for renovated properties, highlighting the complexity of existing building stock retrofits and the need for more occupant-centric modeling. Similarly, Li et al. (2014) demonstrate that high-performance buildings do not always deliver low EUIs in practice, due in part to occupant-driven loads and operational factors. These findings align with Parker (2009) observation that very low energy designs can achieve near net-zero outcomes only when users engage in energy-conscious practices. Indeed, occupant heterogeneity and building usage patterns contribute to wide variance in measured outcomes, even for buildings employing similar technologies (Wang et al. 2012; Chung et al. 2006).

In addition to these operational and occupant-driven dynamics, another strand of research highlights the informational and behavioral channels through which building performance policies can generate impact. Stavins et al. (2013) emphasize that labeling, scoring, and benchmarking policies serve a similar role to consumer product efficiency labels, providing transparent information to buyers, renters, and investors and thereby shifting market demand toward higher-performing buildings. This informational effect complements physical retrofits by shaping expectations and investment behavior. Evidence from residential energy conservation programs reinforces the importance of behavioral responses: Allcott (2011) shows that peer comparison reports reduce household electricity use by an average of 2%, with much larger effects for high-use households, while Costa and Kahn (2013) find that ideological orientation conditions the effectiveness of such "nudges," with liberals more responsive than conservatives.

Building on this, Papadopoulos and Kontokosta (2019) show that machine learning approaches can enhance building grading by incorporating occupancy and operational data, underscoring the need for BEPS to integrate more granular metrics beyond static design parameters. Collectively, this body of work echoes Ruparathna et al. (2016) and Foroushani et al. (2022) in suggesting that performance-based regulations must adopt flexible, context-specific strategies, addressing occupant behavior, building typology, and local climate to maximize real-world energy savings and effectively drive the net-zero carbon transition.

While the theoretical rationale for benchmarking and BEPS is well-established, empirical evidence on their realized effectiveness remains comparatively limited. The literature shows that benchmarking and performance-based standards can generate measurable but heterogeneous improvements in building energy efficiency and emissions. Benchmarking programs promote transparency and modest voluntary reductions, while BEPS and related performance mandates offer greater potential for deep decarbonization when supported by robust enforcement, flexible compliance pathways, and high-quality data. Yet, key empirical gaps persist. Most studies remain ex ante or descriptive, rely on limited or short time-series data, and rarely quantify realized efficiency or emissions outcomes at the building level. Moreover, behavioral, operational, and climatic heterogeneity remain underexplored—particularly regarding how policy stringency and compliance shortfalls shape responses across building types and ownership categories. Addressing these limitations, the present study provides an ex post causal evaluation of Washington, DC's BEPS, using building-level longitudinal data to assess its realized energy efficiency and environmental impacts.

3.0 Methodology

3.1 Data and Cleaning Procedures

The data used in this study are drawn from Open Data DC, which provides building-level benchmarking records reported under the District of Columbia's Building Energy Performance Standards (BEPS). These records constitute an unbalanced panel spanning 2013–2023, capturing year-to-year variation in energy consumption, emissions, and compliance behavior across buildings. Climate-related variables, cooling degree days (CDD) and heating degree days (HDD), calculated at a 65°F base temperature is obtained from Bizee Degree Days (degreedays.net) and mapped by year since they vary over time but not across buildings. Together, these sources form the empirical foundation for assessing the impacts of BEPS on building energy efficiency and greenhouse gas performance in the District of Columbia.

Extensive data cleaning procedures are implemented to ensure accuracy, consistency, and reliability. First, only properties with a reporting status of "In Compliance" are retained, since compliance indicates adherence to DC's benchmarking standards. Second, buildings with missing or zero values for electricity consumption or weather-normalized site energy use intensity are excluded to avoid incomplete records. Third, only standalone buildings or the primary property within a campus are retained, eliminating partial or nested structures. Fourth, outliers are identified and removed by excluding properties whose log-transformed, weather-normalized site energy intensity deviate by more than two standard deviations from their property-type mean.

Further refinements improve classification consistency: property type categories are collapsed by grouping all "Other" labels into a single "Others" category, and targeted recoding harmonizes related categories (e.g., "Food Sales" mapped to "Supermarket/Grocery Store," "Warehouse (Unrefrigerated)" mapped to "Non-Refrigerated Warehouse," "Vocational School" mapped to "Adult Education"), ensuring comparability while reducing noise from inconsistent classifications.

To identify publicly owned buildings, benchmarking records are matched to the District Government Owned Structures dataset using normalized SSL identifiers. Properties that match are coded as "Public," while others are assigned to "Private" ownership, enabling replication of the analysis on ownership-based subsets. Additional cleaning steps address measurement and missingness. A dummy variable flags property with missing Energy Star scores for robustness checks, and buildings reporting zero GHG emissions are removed, as such entries likely reflect reporting errors or incomparable structures. After all cleaning steps, the dataset is reduced from 26,689 building-year observations to a final sample of 15,828 observations.

3.2 Descriptive Statistics

Table 1 reports descriptive statistics for the main variables used in the analysis. The average Site Energy Use Intensity (EUI) is 63.8 kBtu/ft^2, while Source EUI averages 142.8 kBtu/ft^2, both exhibiting substantial variation across buildings. The mean Energy Star score is 65.6 with a coverage spanning the full 1–100 range. Average annual GHG emissions amount to approximately 1,041 metric tons of CO_2e per building, with a highly skewed distribution reflected in the large maximum value of nearly 67,800 metric tons. On an intensity basis, GHG emissions average 5.4

$kgCO_2e/ft^2$. Buildings in the sample are large on average, with a mean floor area of about 182,740 ft^2, though considerable variation exists, ranging from structures just over 10,000 ft^2 to large complexes exceeding 5.6 million ft^2. Climate controls indicate limited interannual variation, with average annual cooling degree days of 1,825 and heating degree days of 3,723 over the study period.

Table 1: Summary statistics of key variables

Variable	Count	Mean	Std	Min	25%	50%	75%	Max
Site EUI ($kBtu/ft^2$)	15415	63.80	25.69	19.40	45.60	59.50	76.30	174.40
Source EUI ($kBtu/ft^2$)	15415	142.77	59.37	22.60	100.40	129.00	172.80	519.30
Energy Star Score	13514	65.56	24.16	1.00	51.00	72.00	84.00	100.00
GHG Emissions (MTCO2e)	15415	1041.01	1893.27	0.00	319.05	615.90	1236.35	67780.00
GHG Emissions Intensity ($KgCO2e/ft^2$)	15415	5.41	2.56	0.10	3.70	4.80	6.50	30.60
Building Size (floor area, ft^2)	15415	182739.67	216096.16	10071.00	68709.00	119549.00	233173.00	5634890.00
Cooling degree days	15415	1825.26	110.94	1680.60	1728.80	1835.70	1951.30	1993.70
Heating degree days	15415	3722.61	278.44	3216.70	3498.20	3759.30	3925.00	4163.20

3.3 Outcome Variables of Interest

The empirical analysis focuses on five outcome variables that capture the energy efficiency and environmental performance of buildings, each of which is directly tied to the objectives of the BEPS. Together, these measures allow the study to assess whether BEPS has reduced energy use and emissions while improving performance relative to industry benchmarks.

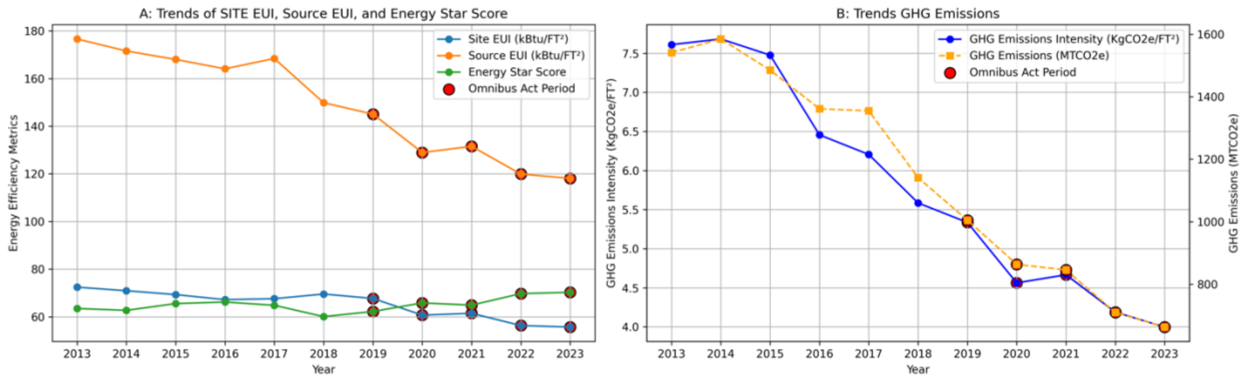
The first two outcomes are Weather-Normalized Site Energy Use Intensity (Site EUI, $kBtu/ft^2$) and Weather-Normalized Source Energy Use Intensity (Source EUI, $kBtu/ft^2$). Site EUI measures the amount of energy consumed per square foot of floor area at the property level, reflecting operational efficiency. Source EUI expands this measure to account for the total upstream energy required to deliver energy to the building, including generation and transmission losses. Both Site and Source EUI are normalized for weather, meaning they are adjusted to reflect what energy consumption would have been under 30-year average climate conditions. This adjustment allows for more consistent year-to-year comparisons within buildings and across regions by correcting for unusually hot or cold years, while not altering differences between distinct climate zones (NEEP 2020; DOEE 2021a).

The third outcome is the Energy Star Score (1–100), a standardized performance rating calculated by the Energy Star Portfolio Manager. This metric evaluates how efficiently a building operates relative to comparable properties nationwide, adjusting for climate and operational characteristics. A score of 50 represents the national median, while a score of 75 or higher indicates high efficiency and potential eligibility for Energy Star Certification. This outcome provides an intuitive benchmark of relative performance and allows evaluation of whether BEPS has shifted buildings toward higher levels of efficiency recognized in national certification programs (U.S. EPA 2022).

The final two outcomes capture the environmental dimension of BEPS. GHG Emissions ($MTCO_2e$) measure the aggregate amount of carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O) released into the atmosphere due to a building's energy consumption. This measure accounts for the varying global warming potentials of these gases and includes both direct emissions from on-site fuel use and indirect emissions from purchased energy produced off-site. To standardize across buildings of different sizes, GHG Emissions Intensity ($kgCO_2e/ft^2$) divides total emissions by floor area, providing a comparable measure of how efficiently a building manages emissions relative to its operational scale (NEEP 2020; DOEE 2021a).

Figure 1 illustrates temporal trends in building energy performance and emissions outcomes from 2013 to 2023. Panel A shows a steady decline in both weather-normalized Site EUI and Source EUI over the period, alongside a gradual increase in Energy Star scores, indicating broad improvements in energy efficiency and benchmarking performance across buildings. Panel B highlight the notable reductions in greenhouse gas outcomes, with both GHG emissions and emissions intensity ($kgCO_2e/ft^2$) declining substantially over time. Notably, the downward trajectories in emissions measures become steeper after 2019, coinciding with the enactment of the Clean Energy Omnibus Act and the lead-up to the first binding BEPS compliance cycle.

Figure 1: Trends in Building Energy Efficiency Metrics and GHG Emissions



3.4 Policy Variable Construction

The treatment variable in this study is defined through the concept of a 'gap', which measures the difference between each building's baseline energy performance and the applicable BEPS threshold. The baseline is calculated as the building's average performance during the pre-policy period (2013–2018), prior to the implementation of the Clean Energy Omnibus Act. The gap represents the degree of underperformance relative to the BEPS standard: buildings performing below the threshold have a positive gap, reflecting the magnitude of required compliance, while buildings at or above the threshold have a gap of zero. Importantly, among the outcome variables, only the Energy Star score is a higher-is-better metric; all others (Site EUI, Source EUI, total GHG emissions, and GHG intensity) are lower-is-better. This distinction ensures that a positive gap always represents underperformance, regardless of the outcome.

Defining treatment intensity through a performance gap follows a well-established literature such as Ryan (2012); Greenstone et al. (2012); and Aghion et al. (2016) that models regulatory exposure as a continuous function of baseline noncompliance or distance from a policy standard. Buildings that are further below the BEPS threshold therefore face greater regulatory stringency and stronger incentives to undertake efficiency improvements. Anchoring the gap to pre-policy performance captures ex ante regulatory pressure while avoiding post-treatment endogeneity, consistent with approaches in the environmental and energy economics literature, including Holland et al. (2009) and Bushnell et al. (2008).

Formally, for building i and outcome v, under Method A (type-average threshold), the gap is defined as:

$$Gap_{i,v}^{A} = \begin{cases} \max\{0,\ Baseline_{i,v} - Threshold_{s,v}^{A}\}, if\ lower\ is\ better \\ \max\{0, Threshold_{s,v}^{A} - Baseline_{i,v}\}, if\ higher\ is\ better \end{cases}$$

where $Baseline_{i,v}$ is the 2013–2018 pre-policy mean for building i, outcome v, and $Threshold_{s,v}^{A}$ is the pre-policy mean for property type s.

For Method B (type-percentile benchmark), the gap is instead defined relative to the efficient end of the property type distribution:

$$Gap_{i,v}^{B} = \begin{cases} max\{0,\ Baseline_{i,v} - Q_{0.25}(Threshold_{j,v} |\ j\ \in s)\}, if\ lower\ is\ better \\ max\{0, Q_{0.75}(Threshold_{j,v} |\ j \in s\ ) - Baseline_{i,v}\}, if\ higher\ is\ better \end{cases}$$

where $Q_{0.25}$ and $Q_{0.75}$ denote the 25th and 75th percentiles of pre-policy performance within property type s.

For Method C (citywide benchmark), the gap is constructed relative to the citywide average:

$$Gap_{i,v}^{c} = \begin{cases} max\{0,\ Baseline_{i,v} - Threshold_{s,v}^{c}\}, if\ lower\ is\ better \\ max\{0, Threshold_{s,v}^{c} - Baseline_{i,v}\}, if\ higher\ is\ better \end{cases}$$

where $Threshold_{s,v}^{c}$ is the overall pre-policy mean across all buildings for outcome v.

Table 2 summarizes descriptive statistics for the three alternative compliance gap measures. Under Gap A, most buildings exhibit zero or small compliance shortfalls, with median gaps of zero across all outcomes, but the distributions are highly right-skewed; for example, the mean Site EUI gap is 8.3 kBtu/ft², 12.1 kBtu/ft² at 75th percentile and exceeds 105 kBtu/ft² at the upper tail, while average GHG emissions gaps are 426 metric tons of CO_2e with maxima near 60,000 metric tons. Gap B produces larger and more dispersed shortfalls: mean gaps rise to 18.1 kBtu/ft² for Site EUI and 36.1 kBtu/ft² for Source EUI, with corresponding medians of 12.3 and 23.4 kBtu/ft², and mean GHG emissions and intensity gaps of 838 metric tons and 1.6 kgCO_2e/ft², respectively. Gap C yields intermediate magnitudes, with mean Site EUI and Source EUI gaps of 9.2 and 22.2 kBtu/ft² and average GHG emissions and intensity gaps of 491 metric tons and 0.87 kgCO_2e/ft². Despite these differences, Table 3 shows that the three gap measures are highly correlated across all outcomes, with pairwise correlations ranging from 0.82 to 0.97, indicating that the they capture a common underlying dimension of pre-policy noncompliance.

Table 2: Descriptive Statistics across the Three Compliance Gaps

Outcome Variables	Count	Mean	Std	Min	25%	50%	75%	Max
A: Property-Type Mean Compliance Gaps								
Site EUI (kBtu/ft²)	11937	8.29	14.74	0.00	0.00	0.00	12.14	105.81
Source EUI (kBtu/ft²)	11937	16.32	31.45	0.00	0.00	0.00	20.82	332.90
Energy Star Score	11261	8.50	13.90	0.00	0.00	0.00	13.26	67.08
GHG Emissions (MTCO2e)	11937	425.63	1651.87	0.00	0.00	0.00	301.28	59876.80
GHG Emissions Intensity (KgCO2e/ft²)	11937	0.68	1.48	0.00	0.00	0.00	0.77	13.50
B: Property-Type Percentile Compliance Gaps								
Site EUI (kBtu/ft²)	11937	18.05	20.20	0.00	0.27	12.30	28.41	124.50
Source EUI (kBtu/ft²)	11937	36.07	41.79	0.00	1.50	23.35	53.73	361.50
Energy Star Score	11261	19.28	20.32	0.00	1.33	12.33	31.80	84.00
GHG Emissions (MTCO2e)	11937	838.15	1922.29	0.00	9.77	332.62	1068.58	63896.38
GHG Emissions Intensity (KgCO2e/ft²)	11937	1.59	1.93	0.00	0.10	1.07	2.28	15.27
C: Citywide Mean Compliance Gaps								
Site EUI (kBtu/ft²)	11937	9.16	17.38	0.00	0.00	0.00	11.54	104.97
Source EUI (kBtu/ft²)	11937	22.15	40.09	0.00	0.00	0.00	31.47	354.55
Energy Star Score	11281	9.58	15.51	0.00	0.00	0.00	14.69	62.49
GHG Emissions (MTCO2e)	11937	491.20	1909.40	0.00	0.00	0.00	305.86	63004.33
GHG Emissions Intensity (KgCO2e/ft²)	11937	0.87	1.78	0.00	0.00	0.00	1.04	14.54

Table 3: Compliance Gap Correlations

Outcome Variables	Corr (A, B)	Corr (A, C)	Corr (B, C)
Site EUI (kBtu/ft²)	0.94	0.86	0.84
Source EUI (kBtu/ft²)	0.95	0.82	0.84
Energy Star Score	0.93	0.91	0.89
GHG Emissions (MTCO2e)	0.97	0.95	0.96
GHG Emissions Intensity (KgCO2e/ft²)	0.94	0.87	0.89

This table reports pairwise correlations between the three alternative compliance gap definitions (A, B, and C) across the outcome variables.

Since the dataset is unbalanced, attrition may bias estimates if buildings that exit the sample differ systematically from those that remain, a well-established concern in panel data analysis highlighted by Gottschalk (1998) and Wooldridge (2010). Following practice developed by Robins et al. (1995) and formalized by Wooldridge (2007), the analysis addresses potential attrition bias by estimating a logit model for the probability of exit, defined as a building having no observations at or after 2021.

Formally, let R_i^{max} denote the maximum reporting year observed for building i. The stayer and exit buildings are defined as:

$$Stayer_i = \begin{cases} 1, if R_i^{max} \geq 2021, \\ 0, if R_i^{max} < 2021, \end{cases} \qquad Exit_i = 1 - Stayer_i$$

Thus, buildings with observations in or after 2021 are classified as *stayers* ($Stayer_i = 1$), while those with no post-2020 records are classified as *exiters* ($Exit_i = 1$). This variable ($Exit_i = 1$) is used as the dependent variable in the attrition logit model used to generate inverse probability weights.

The model uses the last available pre-policy (≤ 2018) record for each building and includes baseline performance measures—Site EUI, Source EUI, Energy Star score, total GHG emissions, and GHG emissions per sqft—alongside building size, ownership type (public vs. private), and climate variables (HDD, CDD). The specification is:

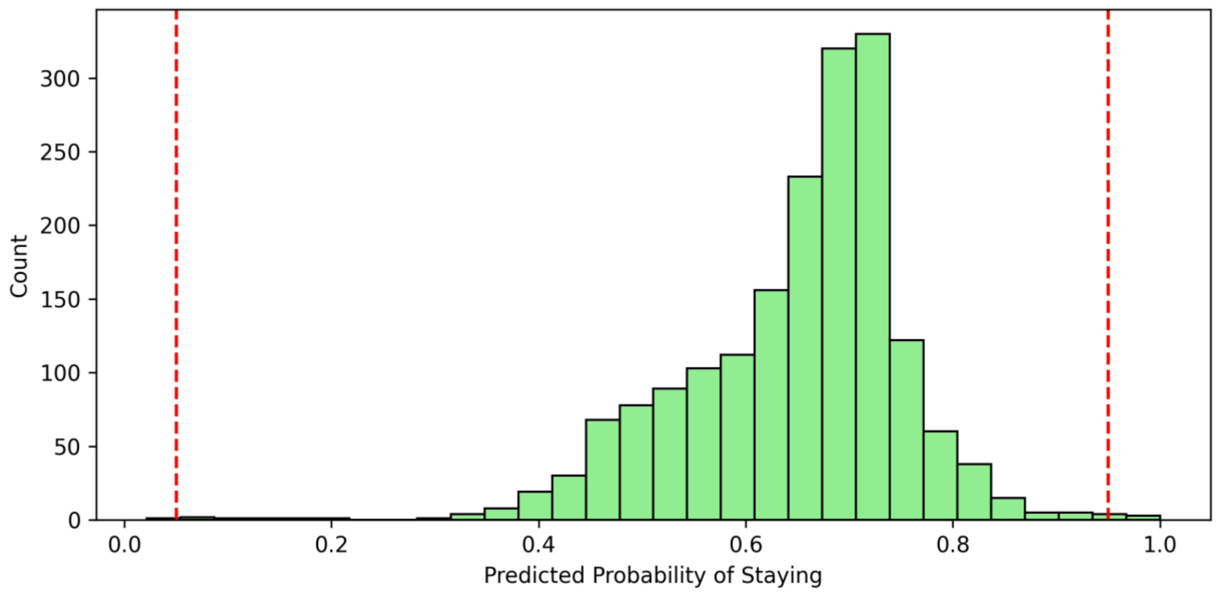
$$Pr(Exit_i = 1|X_i) = \Lambda(\gamma_0 + \gamma' X_i)$$

where $\Lambda(\cdot)$ is the logistic function and X_i is the vector of baseline predictors. To ensure a well-specified model, near-zero variance predictors and collinear variables are dropped, and clustered standard errors are calculated at the building level.

From this model, predicted probabilities of staying, $\widehat{p_i} = 1 - Pr(Exit_i = 1)$, are used to construct inverse probability weights, $w_i = {}^{1}/_{\widehat{p_i}}$.

These weights are applied in robustness checks of the continuous DiD models to mitigate potential attrition bias. The overlap (positivity) assumption is examined to ensure the validity of inverse probability weighting. This assumption requires that all buildings, regardless of their covariate profiles, retain a non-zero probability of staying in the sample. Figure 3 displays the distribution of predicted probabilities of staying from the attrition logit model, with red dashed lines marking the [0.05, 0.95] range typically used to assess common support. The distribution is concentrated between 0.4 and 0.8, with most buildings clustering around 0.6–0.7. Only 0.2% of observations fall below 0.05 and 0.1% exceed 0.95, leaving 99.8% of the sample. These results indicate substantial overlap between exiters and stayers, with no subset of buildings deterministically predicted to remain or exit. The overlap condition is therefore satisfied, supporting the application of inverse probability weighting without concerns of instability from extreme weights.

Figure 3: Distribution of predicted probabilities of staying



3.5 Identification Strategy and Estimation of Causal Effects

The primary empirical challenge in evaluating the impact of BEPS is that nearly all large public and private buildings in the dataset are subject to the policy, leaving no natural untreated comparison group. This institutional feature rules out a conventional binary-treatment DiD design. Instead, treatment exposure varies continuously according to the magnitude of each building's compliance gap—the distance between its pre-policy baseline performance and the property-type threshold. Buildings with larger gaps face stricter compliance obligations, while those with small or zero gaps face weaker or no obligations. To accommodate this structure, the analysis employs a generalized Difference-in-Differences (DiD) design with continuous treatment intensity. Because treatment is synchronized across units, the effect is estimated via a Two-Way Fixed Effects (TWFE) model where the coefficient of interest is the interaction between a unit-level dosage measure and a post-treatment indicator.

Formally, the baseline specification is expressed as:

$$Y_{it} = \beta_0\left(Post_t\ x\ Gap_{i,v}\right) + \alpha_i + \tau_t + \varepsilon_{it}$$

where Y_{it} denotes the outcome of interest for building i in year t (Site EUI, Source EUI, Energy Star Score, GHG emissions, or GHG emissions intensity). The variable $Gap_{i,v}$ measures the pre-policy shortfall relative to the threshold, and $Post_t$ is the policy indicator, building fixed effects (α_i) absorb time-invariant heterogeneity while year fixed effects (τ_t) account for common shocks across all buildings in a given year.

The $Post_t$ indicator is coded as 1 beginning in 2021, consistent with the implementation of BEPS 1, but only for buildings meeting the statutory eligibility thresholds: public buildings with a reported floor area of at least 10,000 square feet and private buildings with a reported floor area of at least 50,000 square feet. For all other cases, including buildings below the size cutoffs or observations prior to 2021, the Post indicator equals 0.

$$Post_t = \begin{cases} 1,\ if\ t \geq 2021\ and \begin{cases} public\ building\ with\ floor\ area_i \geq 10{,}000\ ft^2, \\ private\ building\ with\ floor\ area_i \geq 50{,}000\ ft^2, \end{cases} \\ 0,\ otherwise \end{cases}$$

The interaction term, $Post_t\ x\ Gap_{i,v}$, captures whether buildings with larger pre-policy shortfalls relative to the threshold experienced greater changes in energy efficiency and emissions outcomes once the policy became binding. The coefficient, β_0, therefore measures the average causal response to treatment—the marginal effect of a one-unit increase in the pre-policy compliance gap on post-policy outcomes.

In addition to BEPS, the study period overlaps with other energy-related policy interventions that may affect building performance, particularly for public buildings. Most notably, the federal Infrastructure Investment and Jobs Act (IIJA) became effective in 2021, and DC expanded local energy retrofit and financing programs beginning in 2020. Although these policies are not directly tied to BEPS compliance, they may influence investment incentives and retrofit activity in ways that differ systematically by ownership type. To ensure that estimated BEPS effects are not confounded by these contemporaneous interventions, the empirical strategy explicitly accounts for overlapping federal and local policies in augmented specifications by allowing their effects to vary with public ownership.

$$Y_{it} = \beta_0\left(Post_t\ x\ Gap_{i,v}\right) + \gamma_1(Fed_t\ x\ Public_i) + \ \gamma_2(Local_t\ x\ Public_i) + \alpha_i + \tau_t + \varepsilon_{it}$$

where Fed_t dummy indicates the post-2021 federal Infrastructure Investment and Jobs Act period, $Local_t$ dummy captures local energy retrofit programs introduced in 2020, and $Public_i$ dummy identifies public buildings.

To further examine whether BEPS induces heterogeneous responses across building ownership types, the specification is extended to include a triple interaction between the post-policy indicator, the compliance gap, and public ownership:

$$\begin{aligned} Y_{it} = \beta_0\left(Post_t\ x\ Gap_{i,v}\right) + \beta_p\left(Post_t\ x\ Gap_{i,v}\ x\ Public_i\right) + \gamma_1(Fed_t\ x\ Public_i) \\ + \ \gamma_2(Local_t\ x\ Public_i) + \alpha_i + \tau_t + \varepsilon_{it} \end{aligned}$$

where β_0 captures the post-policy marginal response to treatment intensity for private buildings, while β_p measures the differential marginal response for public buildings relative to private buildings.

Importantly, several concerns that motivate recent critiques of TWFE estimators are not relevant in this setting. First, the negative-weighting problem emphasized in staggered binary-treatment designs does not arise here, because all buildings that meet the statutory criteria become subject to BEPS at the same time, and identification relies on continuous variation in treatment intensity rather than staggered adoption across units. Second, the treatment variable—the compliance gap—is predetermined using pre-policy outcomes, eliminating concerns about endogenous treatment timing or selection into treatment intensity. Third, the interpretation of β_0 does not rely on implicit comparisons across heterogeneous treatment cohorts with non-intuitive weighting schemes; instead, it captures a marginal response to treatment intensity in the post-policy period.

This empirical strategy is grounded in the continuous DiD literature that extends classical DiD designs to settings with non-binary, multi-valued, and continuous treatments (Roth et al. 2023;

Callaway and Sant'Anna 2021; De Chaisemartin and d'Haultfoeuille 2024; Baker et al. 2022). While much of this literature focuses on binary treatment adoption, a growing body of work explicitly studies continuous or ordered treatment intensity (e.g., De Chaisemartin and d'Haultfoeuille 2018, 2020, 2024). In this context, the analysis adopts a sharp continuous-treatment DiD design in which policy exposure varies at the unit level according to the magnitude of each building's compliance gap. Estimation is implemented within a fixed-effects panel framework that differences out time-invariant building characteristics and common time shocks, and recent studies such as Wooldridge (2021); Callaway et al. (2021); and de Chaisemartin et al. (2023) show that in multi-period settings, the coefficient on the interaction between post-policy exposure and treatment intensity identifies an average causal response to treatment. This parameter summarizes the marginal effect of a one-unit increase in pre-policy policy on post-policy building outcomes, providing a dose–response interpretation of BEPS impacts.

3.6 Event-Study Analysis

To complement the baseline estimations, an event-study framework is implemented to assess the identifying assumptions and to examine the dynamic effects of BEPS over time. The event-study serves two primary purposes. First, it provides a diagnostic test of the parallel trends in intensity assumption, by examining whether buildings with different pre-policy compliance gaps exhibited different outcome trajectories prior to BEPS becoming binding. Second, it allows for the estimation of the dynamic causal response to treatment intensity, showing whether policy effects emerge immediately or evolve gradually as compliance investments are undertaken. The event-study specification is given by:

$$Y_{it} = \sum_{k \neq -1} \delta_k 1\{t - t_0 = k\} \times Gap_{i,v} + +\theta_1(Fed_t\ x\ Public_i) + \ \theta_2(Local_t\ x\ Public_i) + \alpha_i + \tau_t + \mathcal{E}_{it}$$

where t_0 denotes the final pre-policy year (2018), and k indexes the number of years relative to that baseline. The omitted category is $k = -1$, corresponding to the year immediately preceding policy implementation, so all coefficients are interpreted relative to 2018. The coefficient δ_k trace the dynamic relationship between treatment intensity and outcomes k years before or after BEPS took effect.

The estimated pre-policy coefficients ($\delta_k\ for\ k < 0$) serve as a direct test of the identifying assumption. Flat and statistically insignificant pre-trends provide evidence that, absent BEPS, outcomes would have evolved similarly across buildings with different compliance gaps. Conversely, significant pre-trends would undermine identification, suggesting that results could be driven by differential trajectories rather than the policy itself. To formally assess this condition, joint significance test (F-tests) is conducted for all pre-policy coefficients. Failure to reject the joint null hypothesis supports the validity of the parallel trends in intensity assumption. Post-policy coefficients ($\delta_k\ for\ k \geq 0$) capture the dynamic causal response of outcomes to the compliance gap in each year following BEPS implementation. These effects show whether improvements in energy efficiency and emissions occur immediately in the first compliance cycle or accumulate more gradually as building owners undertake investments and operational adjustments.

The event-study analysis is conducted using Method A compliance gaps (defined relative to property-type average thresholds) and applied to the full sample of eligible buildings. In the corresponding figures, the pre-policy period is expected to display coefficients fluctuating narrowly around zero, consistent with the absence of differential pre-trends. In contrast, the post-policy period illustrates the timing and persistence of BEPS effects, with confidence intervals conveying both statistical significance and uncertainty in the dynamic response.

3.7 Estimation and Robustness Analyses

Estimation proceeds using the compliance gap derived from Method A (property-type average threshold). The baseline analysis is implemented in three steps. First, the core model is estimated without overlapping policy controls to establish the primary relationship between BEPS treatment intensity and building outcomes. Second, the baseline specification is augmented with controls for contemporaneous federal and local energy policies through interactions with the public-building indicator, accounting for potential confounding policy shocks. Third, the augmented specification is re-estimated using inverse probability weights derived from the attrition model. Across all baseline specifications, heterogeneity in policy effects is examined through interaction-based models, including a triple interaction between the post-policy indicator, the compliance gap, and the public-building indicator, which allows the marginal response to treatment intensity to differ between public and private buildings while maintaining a unified identification framework.

A series of robustness checks is then conducted to assess the sensitivity of the results to alternative modeling choices. First, the baseline analysis is replicated using alternative compliance gap definitions. Method B, based on property-type percentile benchmarks, emphasizes relative performance within building categories, while Method C, based on citywide average benchmarks, applies a uniform standard across all buildings. For each alternative gap definition, the model including the use of overlapping policy controls is estimated in both unweighted and IPW-weighted forms, allowing assessment of robustness to alternative constructions of treatment intensity.

Robustness to policy timing and contemporaneous shocks is then examined using alternative post-policy definitions and sample restrictions. To assess potential anticipatory responses following the passage of the Clean Energy Omnibus Act, two alternative post indicators are constructed. In the first, the post-policy period begins in 2019, and in the second, it begins in 2020 as indicated below.

$$Post2019_t = \begin{cases} 1,\ if\ t \geq 2019\ and \begin{cases} public\ building\ with\ floor\ area_i \geq 10{,}000\ ft^2, \\ private\ building\ with\ floor\ area_i \geq 50{,}000\ ft^2, \end{cases} \\ 0,\ otherwise \end{cases}$$

$$Post2020_t = \begin{cases} 1,\ if\ t \geq 2020\ and \begin{cases} public\ building\ with\ floor\ area_i \geq 10{,}000\ ft^2, \\ private\ building\ with\ floor\ area_i \geq 50{,}000\ ft^2, \end{cases} \\ 0,\ otherwise \end{cases}$$

These alternative definitions allow the analysis to distinguish effects driven by formal enforcement beginning in 2021 from earlier responses that may have occurred in anticipation of compliance obligations. To further isolate BEPS effects from pandemic-related disruptions, a donut specification excludes observations from 2020, the year most affected by COVID-related shocks and transitional implementation dynamics:

$$Y_{it} = \beta_0^D \left(Post_t\ x\ Gap_{i,v}\right) + \gamma_1 (Fed_t\ x\ Public_i) + \gamma_2 (Local_t\ x\ Public_i) + \alpha_i + \tau_t + \varepsilon_{it},\quad t \neq 2020$$

This specification ensures that estimated treatment effects are not mechanically driven by pandemic-era volatility or transitional policy effects. As with the main analysis, the donut specification is estimated in both unweighted and IPW-weighted forms.

Finally, the interaction-based heterogeneity analysis is extended across alternative compliance gap definitions. The triple interaction specification is replicated for Gaps B and C, and estimated under both unweighted and IPW-weighted specifications with overlapping policy controls. Following this, the analysis is complemented by ownership-specific subsample estimations for public buildings and private buildings, respectively. These subsample results provide a validation of the interaction-based findings by allowing the response to treatment intensity to be estimated separately within each ownership group.

4.0 Empirical Findings

4.1 Causal Impacts of BEPS on Energy Efficiency and Emissions

Table 4 reports the estimates of the impact of BEPS using the property-type mean compliance gap as the treatment intensity. Panel A presents baseline unweighted estimates. The results indicate that BEPS led to significant improvements in building energy performance and reductions in greenhouse gas emissions, with larger effects for buildings facing greater pre-policy compliance shortfalls. A one-unit increase in the compliance gap leads to a reduction of 0.33 kBtu/ft^2 in Site EUI and 0.45 kBtu/ft^2 in Source EUI, alongside a 0.41-point increase in Energy Star scores. On the environmental margin, total GHG emissions decline by 0.35 metric tons of CO_2e and emissions intensity falls by 0.65 kgCO_2e/ft^2 per additional unit of gap. All estimates are statistically significant at the 1 percent level.

Panel B augments the baseline specification by controlling for overlapping federal (2021) and local (2020) energy retrofit policies that coincide with the BEPS compliance period. The estimated coefficients remain virtually unchanged in both magnitude and statistical significance, suggesting that the observed improvements in energy efficiency and emissions outcomes are not driven by concurrent policy interventions. Panel C reports the weighted estimates that correct for potential non-random attrition of buildings over time. The results remain robust and, for emissions outcomes, increase in magnitude. In particular, the estimated reductions in total GHG emissions and emissions intensity rise to 0.54 metric tons of CO_2e and 0.76 kgCO_2e/ft^2 per unit of compliance gap, respectively, while the estimated effects on energy use intensity and Energy Star scores remain closely aligned with the unweighted specifications. These findings suggest that sample attrition does not drive the main results and that emissions responses among initially underperforming buildings may be understated in the unweighted models.

In substantive terms, the estimates imply that a building performing 10 percent worse than its property-type mean benchmark prior to BEPS experienced sizable post-policy improvements. Such a building reduced its Site EUI by approximately 3.3 kBtu/ft² and its Source EUI by 4.5 kBtu/ft², while improving its Energy Star score by about 4.1 points. At the same time, total greenhouse gas emissions declined by roughly 3.5 metric tons of CO_2e (or 5.4 metric tons under the IPW specification), and emissions intensity fell by 6.5–7.6 $kgCO_2e/ft^2$. The findings demonstrate that BEPS has led to meaningful improvements in energy efficiency and GHG emissions in buildings.

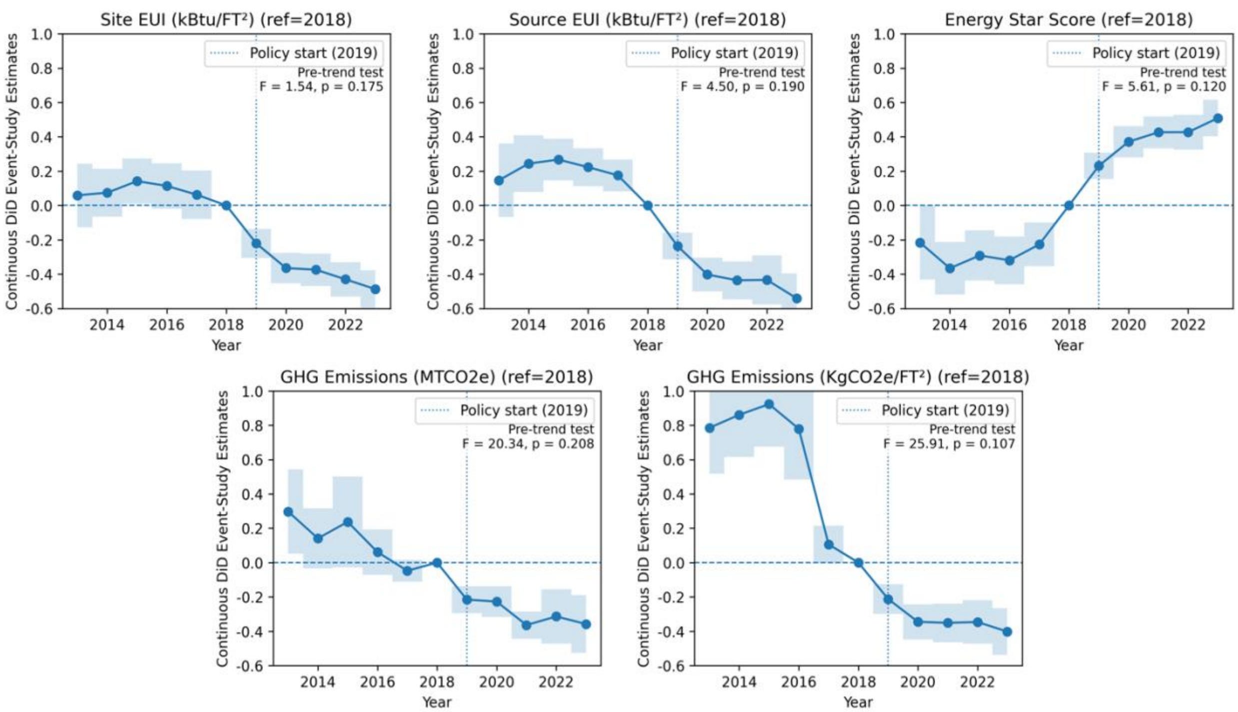
Table 4: BEPS Impacts Estimated with Mean Property-Type Compliance Gaps

	Site EUI (kBtu/ft²)	**Source EUI (kBtu/ft²)**	**Energy Star Score**	**GHG Emissions ($MTCO_2e$)**	**GHG Intensity ($KgCO_2e/ft^2$)**
A: Base Model Estimates					
Post × Compliance Gap	-0.33*** (0.04)	-0.45*** (0.06)	0.41*** (0.03)	-0.35*** (0.07)	-0.65*** (0.04)
Observations	11,937	11,937	10,686	11,937	11,937
Entities	1997	1997	1810	1997	1997
R^2 (within)	0.106	0.21	0.054	0.21	0.34
B: Estimates with overlapping policies					
Post × Compliance Gap	-0.33*** (0.04)	-0.45*** (0.06)	0.41*** (0.04)	-0.35*** (0.08)	-0.65*** (0.04)
Observations	11,937	11,937	10,686	11,937	11,937
Entities	1997	1997	1810	1997	1997
R^2 (within)	0.11	0.22	0.05	0.22	0.34
C: IPW-Weighted Estimates					
Post × Compliance Gap	-0.32*** (0.04)	-0.44*** (0.06)	0.42*** (0.04)	-0.54*** (0.06)	-0.76*** (0.02)
Observations	11,281	11,281	10,686	11,281	11,281
Entities	1810	1810	1810	1810	1810
R^2 (within)	0.10	0.22	0.05	0.21	0.24

This table reports continuous DID estimates of the impact of BEPS using the mean compliance gap. Panel A reports unweighted estimates. Panels B and C report inverse probability weighted (IPW) estimates that correct for non-random building attrition and control for overlapping federal (2021) and local (2020) energy subsidy programs. All models include building and year fixed effects. Standard errors are clustered at the building level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure 4 illustrates the dynamic effects of BEPS on building energy performance outcomes. Across all five measures, pre-policy coefficients are small and show no systematic trends prior to implementation, and joint tests fail to reject the null of no differential pre-trends, supporting the parallel trends in intensity assumption. While modest adjustments appear following the policy announcement in 2019, the largest and most persistent effects emerge after 2021, when compliance obligations became binding. Buildings with larger pre-policy compliance gaps experience steady post-2021 declines in Site EUI of roughly 0.4–0.6 kBtu/ft² and in Source EUI of about 0.5–0.7 kBtu/ft² by 2022–2023. Energy Star scores increase markedly over the same period, rising by approximately 0.5 to 0.6 points, indicating improved benchmarking performance. Environmental outcomes show parallel dynamics. GHG emissions decline progressively after the policy becomes binding, with cumulative reductions on the order of 0.4–0.5 metric tons of CO_2e relative to pre-policy levels. GHG emissions intensity exhibits an even sharper post-2021 response, falling by approximately 0.4–0.5 $kgCO_2e/ft^2$ and stabilizing at lower levels through the end of the sample period. Taken together, the results indicate that BEPS effects were not immediate at the time of policy passage but intensified once compliance obligations became binding.

Figure 4: Event-Study Plots of BEPS Impacts



This figure plots event-study estimates where coefficients represent interactions between event-time indicators and the pre-policy mean compliance gap (Gap A). The omitted category is the year 2018 the law was passed. Vertical dashed lines indicate the first binding compliance year (2021). Shaded areas denote 95% confidence intervals based on building-clustered standard errors. Joint tests indicate that pre-policy coefficients are jointly insignificant, supporting the parallel trends assumption.

Table 5 provide the estimates on whether the impact of BEPS varies systematically between public and private buildings by interacting the post-policy indicator and compliance gap with a public-building indicator. Across all outcomes, the first coefficient of interest (Post × Compliance Gap) is similar in magnitude to the baseline model and statistically significant in both the unweighted and IPW-weighted specifications, indicating that the overall performance-based response to BEPS is robust and consistent with the baseline results in Table 3.1. Larger pre-policy compliance gaps continue to lead to sizable post-policy improvements in energy efficiency and reductions in GHG emissions.

The second coefficient of interest (Post × Compliance Gap × Public) is designed to capture whether the marginal response to BEPS differs between publicly and privately-owned buildings, conditional on the size of the compliance gap. Across all outcomes and specifications, whether weighted or not, the estimated coefficients are small and statistically insignificant. This finding indicates that public buildings do not exhibit systematically different post-BEPS responses relative to private buildings. The absence of statistically significant differential effects suggests that BEPS induced broadly similar performance-based adjustments across ownership types, with improvements primarily driven by the intensity of pre-policy noncompliance gap rather than by public versus private building ownership.

Table 5: BEPS Impacts on Private and Public Buildings

	Site EUI (kBtu/ft²)	**Source EUI (kBtu/ft²)**	**Energy Star Score**	**GHG Emissions ($MTCO_2e$)**	**GHG Intensity ($KgCO_2e/ft^2$)**
		A: Unweighted Estimates			
Post × Compliance Gap	-0.34*** (0.04)	-0.44***(0.06)	0.41***(0.04)	-0.34***(0.08)	-0.65***(0.04)
Post × Compliance Gap × Public	0.04(0.13)	-0.06(0.24)	-0.04(0.10)	-0.15(0.13)	0.01(0.19)
Observations	11,937	11,937	10,686	11,937	11,937
Entities	1,997	1,997	1,810	1,997	1,997
R^2 (within)	0.11	0.22	0.05	0.22	0.34
		B: IPW-Weighted Estimates			
Post × Compliance Gap	-0.33*** (0.04)	-0.45***(0.06)	0.43***(0.05)	-0.55***(0.07)	-0.76***(0.02)
Post × Gap × Public	0.13(0.07)	0.15(0.13)	-0.07(0.10)	0.01(0.12)	0.12(0.17)
Observations	11,281	11,281	10,686	11,281	11,281
Entities	1,810	1,810	1,810	1,810	1,810
R^2 (within)	0.10	0.22	0.05	0.21	0.24

This table report heterogeneous BEPS effects across public and private buildings using a triple interaction between the post-policy indicator, the compliance gap, and a public-building indicator. The coefficient on the triple interaction captures the differential post-policy response of public buildings relative to private buildings per unit of compliance gap. Panel A reports unweighted estimates and Panel B reports IPW-weighted estimates. All estimates include building and year fixed effects and control for overlapping federal (2021) and local (2020) energy subsidy programs. Standard errors are clustered at the building level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

4.2 Robustness Estimates of Impacts of BEPS on Energy Efficiency and Emissions

Robustness checks (Table 6) confirm the main findings when alternative definitions of the compliance gap are used. Using property-type percentile compliance gaps (Gap B), the result indicate that each additional unit pre-policy gap reduces Site EUI by about 0.23–0.24 kBtu/ft^2 and Source EUI by approximately 0.36 kBtu/ft^2, while increasing Energy Star scores by roughly 0.32–0.33 points. Environmental outcomes also improve substantially, with GHG emissions declining by 0.28–0.43 metric tons of CO_2e per unit gap and emissions intensity falling by 0.51–0.66 kgCO_2e/ft^2. Results using citywide mean compliance gaps (Gap C) yield a similar pattern. A one-unit increase in the gap leads to reductions of 0.24–0.25 kBtu/ft^2 in Site EUI and about 0.37 kBtu/ft^2 in Source EUI, alongside increases of 0.38–0.39 points in Energy Star scores. GHG emissions fall by 0.26–0.43 metric tons of CO_2e per unit gap, while emissions intensity declines by 0.57–0.69 kgCO_2e/ft^2. These effects remain highly stable across the baseline, overlapping-policy, and IPW-weighted specifications, indicating that the estimated BEPS effects are not sensitive to how compliance gaps are constructed.

Table 6: Robustness estimates with property-type percentile and citywide mean compliance gaps

	Site EUI ($kBtu/ft^2$)	**Source EUI ($kBtu/ft^2$)**	**Energy Star Score**	**GHG Emissions ($MTCO_2e$)**	**GHG Intensity ($KgCO_2e/ft^2$)**
Property-Type Percentile Compliance Gaps					
A: Base Model Estimates					
Post × Compliance Gap	-0.23*** (0.02)	-0.36*** (0.04)	0.33*** (0.03)	-0.43*** (0.05)	-0.66*** (0.04)
B: Estimates with overlapping policies					
Post × Compliance Gap	-0.24*** (0.02)	-0.36*** (0.04)	0.32*** (0.02)	-0.28*** (0.06)	-0.51*** (0.03)
C: IPW-Weighted Estimates					
Post × Compliance Gap	-0.23*** (0.02)	-0.36*** (0.04)	0.33*** (0.03)	-0.43*** (0.05)	-0.66*** (0.04)
Citywide Mean Compliance Gaps					
A: Base Model Estimates					
Post × Compliance Gap	-0.24*** (0.03)	-0.37*** (0.04)	0.39*** (0.04)	-0.43*** (0.09)	-0.69*** (0.03)
B: Estimates with overlapping policies					
Post × Compliance Gap	-0.25*** (0.03)	-0.37*** (0.04)	0.38*** (0.03)	-0.26*** (0.07)	-0.57*** (0.04)
C: IPW-Weighted Estimates					
Post × Compliance Gap	-0.24*** (0.03)	-0.37*** (0.04)	0.39*** (0.04)	-0.43*** (0.09)	-0.69*** (0.03)

This table replicates the baseline analysis using alternative compliance gap definitions. All specifications follow the continuous DiD framework and include building and year fixed effects and control for overlapping federal (2021) and local (2020) energy subsidy programs. Standard errors are clustered at the building level. Unweighted and IPW-weighted estimates are reported. Standard errors are clustered at the building level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 7 provide the robustness of the baseline results to alternative assumptions about BEPS timing and to a donut-year specification that excludes 2020. When the post-policy indicator is redefined to begin in 2019, the estimated effects remain similar and statistically significant. A one-unit increase in the compliance gap reduces Site EUI by approximately 0.40–0.42 $kBtu/ft^2$ and Source EUI by about 0.54–0.55 $kBtu/ft^2$, while increasing Energy Star scores by roughly 0.55 points. Environmental outcomes also improve substantially, with GHG emissions declining by 0.32–0.71 metric tons of CO_2e and emissions intensity falling by 0.86–1.02 $kgCO_2e/ft^2$ per unit gap.

Results are similarly robust when the post-policy period is shifted to 2020. Both unweighted and IPW-weighted estimates show consistent reductions in Site and Source EUI (0.37–0.39 and 0.50–0.51 $kBtu/ft^2$, respectively) and increases in Energy Star scores of about 0.48 points per unit gap. Excluding the year 2020 entirely yields nearly identical estimates, indicating that the main findings are not driven by pandemic-related disruptions or anticipatory behavior. Parallel robustness checks using property-type percentile and citywide mean compliance gaps (Appendix Table A1) produce comparable magnitudes and statistical significance across all outcomes, confirming that the timing and donut-year results are not sensitive to the choice of compliance gap definition.

Table 7: Anticipatory and Donut-Year Robustness Checks Using Mean Compliance Gap

	Site EUI ($kBtu/ft^2$)	**Source EUI ($kBtu/ft^2$)**	**Energy Star Score**	**GHG Emissions ($MTCO_2e$)**	**GHG Intensity ($KgCO_2e/ft^2$)**
A: Unweighted Estimates (2019 Policy Timing)					
Post2019 × Compliance Gap	-0.42*** (0.04)	-0.55*** (0.05)	0.55*** (0.05)	-0.32*** (0.09)	-0.86*** (0.05)
B: IPW-Weighted Estimates (2019 Policy Timing)					
Post2019 × Compliance Gap	-0.40*** (0.05)	-0.54*** (0.06)	0.55*** (0.05)	-0.71*** (0.08)	-1.02*** (0.03)
C: Unweighted Estimates (2020 Policy Timing)					
Post2020 × Compliance Gap	-0.39*** (0.04)	-0.51*** (0.05)	0.48*** (0.04)	-0.31*** (0.08)	-0.75*** (0.05)
D: IPW-Weighted Estimates (2020 Policy Timing)					
Post2020 × Compliance Gap	-0.37*** (0.04)	-0.50*** (0.06)	0.48*** (0.04)	-0.63*** (0.07)	-0.88*** (0.02)
E: Unweighted Estimates (2020 Donut-Year Robustness)					
Post × Compliance Gap	-0.41*** (0.04)	-0.53*** (0.06)	0.50*** (0.05)	-0.41*** (0.10)	-0.77*** (0.05)
F: IPW-Weighted Estimates (2020 Donut-Year Robustness)					
Post × Compliance Gap	-0.39*** (0.04)	-0.51*** (0.06)	0.52*** (0.05)	-0.64*** (0.07)	-0.88*** (0.02)

This table report robustness to alternative policy timing assumptions using the mean compliance gap. Panel A redefines the post-policy indicator as $t \geq 2019$, Panel B as $t \geq 2020$, and Panel C excludes the year 2020 to account for pandemic-related disruptions and early policy activity. All estimates include building and year fixed effects and control for overlapping federal (2021) and local (2020) energy subsidy programs. Unweighted and IPW-weighted estimates are reported. Standard errors are clustered at the building level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 8 assesses the robustness of the public–private heterogeneity results using alternative compliance gap definitions based on property-type percentiles and citywide mean benchmarks. Across both gap measures and specifications, the first coefficient of interest remains statistically significant for all outcomes, confirming that larger pre-policy gaps lead to greater post-BEPS improvements in energy efficiency and emissions reductions. In contrast, the second coefficient of interest is generally small and statistically insignificant across outcomes. These findings are further supported by the separate public and private subsample analyses reported in Appendix Table A2, which show qualitatively similar BEPS effects across ownership types and all three compliance gap definitions. These findings also reinforce the conclusion that BEPS effects are primarily driven by baseline noncompliance rather than building ownership status, and that the absence of public–private differences is robust to alternative compliance gap definitions.

Table 8: BEPS impacts on private and public buildings under alternate compliance gaps

	Site EUI (kBtu/ft²)	Source EUI (kBtu/ft²)	Energy Star Score	GHG Emissions ($MTCO_2e$)	GHG Intensity ($KgCO_2e/ft^2$)
Property-Type Compliance Gaps					
A: Unweighted Estimates					
Post × Compliance Gap	-0.24*** (0.02)	-0.35***(0.04)	0.33***(0.03)	-0.28***(0.06)	-0.51***(0.04)
Post × Compliance Gap × Public	0.00(0.09)	-0.04(0.15)	-0.03(0.06)	-0.06(0.09)	0.03(0.11)
B: IPW-Weighted Estimates					
Post × Compliance Gap	-0.24*** (0.03)	-0.36***(0.04)	0.34***(0.03)	-0.44***(0.05)	-0.67***(0.03)
Post × Compliance Gap × Public	0.06(0.05)	0.09(0.09)	-0.06(0.06)	0.05(0.10)	0.19*(0.11)
Citywide Mean Compliance Gaps					
A: Unweighted Estimates					
Post × Compliance Gap	-0.24*** (0.03)	-0.37***(0.04)	0.39***(0.04)	-0.25***(0.08)	-0.57***(0.04)
Post × Compliance Gap × Public	-0.05(0.12)	-0.07(0.17)	-0.10(0.08)	-0.01(0.11)	-0.01(0.14)
B: IPW-Weighted Estimates					
Post × Compliance Gap	-0.25*** (0.04)	-0.37***(0.05)	0.41***(0.04)	-0.43***(0.10)	-0.69***(0.02)
Post × Compliance Gap × Public	0.03(0.07)	0.09(0.11)	-0.13*(0.09)	0.02(0.14)	0.17(0.14)

This table examines whether public–private heterogeneity in BEPS effects persists under alternative compliance gap definitions. The specification mirrors Table 5 and includes building and year fixed effects and control for overlapping federal (2021) and local (2020) energy subsidy programs. Unweighted and IPW-weighted estimates are reported. All models include building and year fixed effects. Standard errors are clustered at the building level. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

5.0 Conclusion

This study evaluates the causal impacts of Washington, DC's BEPS program on building energy efficiency and greenhouse gas emissions. Using a panel of large public and private buildings observed from 2013 to 2023 and drawn from Open Data DC, the analysis adopts a modern continuous DiD framework suited to settings with universal policy exposure and heterogeneous treatment intensity. The empirical strategy centers on a compliance gap measure that captures the distance between a building's pre-policy performance and the applicable BEPS threshold, constructed using three alternative benchmarks: property-type means, property-type percentiles, and citywide means. By interacting these continuous gaps with a post-policy indicator corresponding to the first binding compliance cycle, the analysis estimates how outcomes respond differentially to the stringency of the standard. Event-study designs further assess the validity of the identifying assumptions and trace the dynamic evolution of policy effects, providing evidence on both the timing and persistence of BEPS impacts. Methodologically, this study contributes to the policy evaluation literature by demonstrating how continuous treatment DiD designs can be applied to performance-based regulations with universal coverage but heterogeneous compliance incentives.

The empirical findings demonstrate that BEPS led to significant improvements in energy efficiency and emissions outcomes, with larger effects for buildings facing greater pre-policy shortfalls. A one-unit increase in the compliance gap reduces Site EUI by 0.33 $kBtu/ft^2$ and Source EUI by 0.45 $kBtu/ft^2$, while increasing Energy Star scores by 0.41 points. Total GHG emissions fall by 0.35 metric tons of CO_2e and emissions intensity declines by 0.65 $kgCO_2e/ft^2$. These effects are robust to controls for overlapping federal and local energy programs and under IPW weighting. In substantive terms, a building performing 10 percent worse than its property-type benchmark prior to BEPS reduced its Site EUI by about 3.3 $kBtu/ft^2$, Source EUI by 4.5 $kBtu/ft^2$, increased its Energy Star score by roughly 4.1 points, and reduced total emissions by 3.5–5.4 metric tons of CO_2e following policy implementation. These magnitudes indicate that BEPS delivers economically meaningful improvements rather than marginal compliance effects.

The dynamic analysis indicates that BEPS effects were not immediate at the time of policy passage but intensified once compliance obligations became legally binding. Pre-policy estimates show no evidence of differential trends across buildings with varying compliance gaps, supporting the validity of the identification strategy. While modest adjustments appear following the policy announcement, the strongest and most persistent effects emerge after 2021, coinciding with the first binding compliance cycle. Post-2021, buildings with larger compliance gaps experience steady declines in energy use intensity on the order of 0.4–0.7 $kBtu/ft^2$, sustained increases in Energy Star scores of about 0.5–0.6 points, and pronounced reductions in both GHG emissions and emissions intensity. This pattern highlights the central role of enforcement and regulatory deadlines, rather than informational signaling alone, in driving observed outcomes.

The results show little evidence that BEPS effects differ systematically between public and private buildings once baseline compliance gaps are accounted for. While larger pre-policy gaps consistently translate into greater post-policy improvements across all outcomes, the marginal response to BEPS is broadly similar across building ownership types. These findings are robust to alternative definitions of the compliance gap, alternative assumptions about policy timing, and specifications that exclude the pandemic year. Taken together, the evidence indicates that BEPS operates as a uniform, performance-based regulation, with its impacts driven primarily by baseline noncompliance and strengthened by enforcement rather than ownership status or concurrent policy interventions. A limitation of the analysis is that it captures effects through the first compliance cycle only, leaving longer-run adjustment costs, technology adoption dynamics, and potential strategic behavior across future cycles for subsequent research.

The policy implications of this research extend beyond Washington, DC. By demonstrating that enforceable performance standards with binding deadlines generate verifiable efficiency and emissions gains, the findings strengthen the case for adopting BEPS-style regulation in other U.S. cities and internationally, particularly where disclosure-only regimes have reached diminishing returns. The results suggest that well-designed standards can deliver substantial energy and environmental benefits when compliance is mandatory and time-bound. A key limitation of the analysis is that it captures impacts through the first compliance cycle only, leaving longer-run adjustment dynamics, technology adoption responses, and potential strategic behavior across subsequent cycles for future research. Future research should examine longer-term adjustments across compliance cycles, interactions with complementary policies, and potential spillovers to market behavior, further informing the design of scalable building decarbonization strategies.

References

Aghion, Philippe, Antoine Dechezleprêtre, David Hémous, Ralf Martin, and John Van Reenen. 2016. "Carbon Taxes, Path Dependency, and Directed Technical Change: Evidence from the Auto Industry." *Journal of Political Economy* 124 (1): 1–51.

Allcott, Hunt. 2011. "Social norms and energy conservation." *Journal of public Economics* 95, no. 9-10: 1082-1095.

Allcott, Hunt, and Michael Greenstone. 2012. "Is there an energy efficiency gap?" *Journal of Economic perspectives* 26(1): 3-28.

Andrews, Abigail, and Rishee Jain. 2023. "Evaluating Building Decarbonization Potential in U.S. Cities under Emissions-Based Building Performance Standards and Load Flexibility Requirements." *Journal of Building Engineering,* 76: 107375.

Angrist, Joshua, and Guido Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association,* 90(430): 431–442. https://doi.org/10.1080/01621459.1995.10476535.

Angrist, Joshua, and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Asensio, Omar Isaac, and Magali Delmas. 2017. "The effectiveness of US energy efficiency building labels." *Nature Energy,* 2(4): 1-9.

Baker, Andrew C., David F. Larcker, and Charles CY Wang. 2022. "How much should we trust staggered difference-in-differences estimates?" *Journal of Financial Economics,* 144(2): 370-395.

Building Innovation Hub. 2025. *DC Energy Benchmarking Trends Part III: Zooming in on Low-Performing Buildings and BEPS Compliance.* Washington, DC: District of Columbia Building Innovation Hub, 2025.

Bushnell, James B., Carla Peterman, and Catherine Wolfram. 2008. "Local Solutions to Global Problems: Climate Change Policies and Regulatory Jurisdiction." *Review of Environmental Economics and Policy* 2 (2): 175–193.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro Sant'Anna. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics,* 225(2): 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of econometrics* 225(2): 200-230.

Cameron, Colin, and Pravin Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

CBI (Commercial Building Inventory). 2012. *The Age of US Commercial Buildings.*

Chung, William, Y. V. Hui, and Y. Miu Lam. 2006. "Benchmarking the energy efficiency of commercial buildings." *Applied energy,* 83(1): 1-14.

Cohen, Robert, and Bill Bordass. 2015. "Mandating transparency about building energy performance in use." *Building research & information,* 43(4): 534-552.

Costa, Dora, and Matthew Kahn. 2013. "Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment." *Journal of the European Economic Association,* 11(3): 680-702.

District of Columbia Council. 2008. *Clean and Affordable Energy Act of 2008*. Washington, DC: DC Council. https://code.dccouncil.gov/us/dc/council/laws/17-250

District of Columbia Council. 2018. *Clean Energy DC Omnibus Amendment Act of 2018*. Washington, DC: DC Council. https://code.dccouncil.gov/us/dc/council/laws/22-257

District Department of Energy & Environment (DOEE). 2019. *Building Energy Performance Standards (BEPS) Program Guidebook.* Washington, DC: DOEE.

District of Columbia Department of Energy & Environment (DOEE). 2021a. *Building Energy Performance Standards (BEPS) Compliance Guidebook*. Washington, DC: Government of the District of Columbia. https://doee.dc.gov/service/building-energy-performance-standards-beps.

DC Department of Energy & Environment (DOEE). 2021b. *BEPS Task Force Report.* Washington, DC: Government of the District of Columbia: DOEE.

De Chaisemartin, Clément, Xavier D'Haultfœuille, Laure Pasquier, and Gonzalo Vazquez-Bare. 2023. "Difference-in-Differences Estimators with Continuous Treatments and Instrumental Variables." *arXiv preprint* arXiv:2305.19312. https://arxiv.org/abs/2305.19312.

De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2018. "Fuzzy differences-in-differences." *The Review of Economic Studies,* 85(2): 999-1028.

De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American economic review* 110(9): 2964-2996.

De Chaisemartin, Clément, and Xavier d'Haultfoeuille. "Difference-in-differences estimators of intertemporal treatment effects." *Review of Economics and Statistics* (2024): 1-45.

EPA (US Environmental Protection Agency). 2014. *Draft Inventory of US Greenhouse Gas Emissions and Sinks 1990–2013.*

European Commission. 2024. "Energy Performance of Buildings Directive." *Energy, Climate Change, Environment*. European Commission. https://energy.ec.europa.eu/topics/energy-efficiency/energy-performance-buildings/energy-performance-buildings-directive_en

Foroushani, Sepehr, Rob Bernhardt, and Mark Bernhardt. 2022. "On the use of the reference building approach in modern building energy codes." *Energy and Buildings,* 256: 111726.

Fossati, Michele, Veridiana Atanasio Scalco, Vinícius Cesar Cadena Linczuk, and Roberto Lamberts. 2016. "Building energy efficiency: An overview of the Brazilian residential labeling scheme." *Renewable and Sustainable Energy Reviews,* 65: 1216-1231.

Goldstein, David B., and Charles Eley. 2014. "A classification of building energy performance indices." *Energy Efficiency,* 7(2): 353-375.

Gottschalk, Peter. 1998. "An Analysis of Sample Attrition in Panel Data." *Journal of Human Resources* 33 (2): 251–299.

Greenstone, Michael, John A. List, and Chad Syverson. 2012. "The Effects of Environmental Regulation on the Competitiveness of U.S. Manufacturing." *Quarterly Journal of Economics* 127 (1): 387–425.

Hicks, Thomas, and Dennis Clough. 1998. "The ENERGY STAR building label: Building performance through benchmarking and recognition." In *Proceedings of the ACEEE 1998 Summer Study on Energy Efficiency in Buildings*, vol. 4, pp. 205-210..

Holland, Stephen P., Jonathan E. Hughes, and Christopher R. Knittel. 2009. "Greenhouse Gas Reductions under Low Carbon Fuel Standards?" *American Economic Journal: Economic Policy* 1 (1): 106–146.

Hsu, David. 2014a. Improving energy benchmarking with self-reported data. *Building Research & Information*, *42*(5), 641–656.

Hsu, David. 2014b. “How Much Information Disclosure of Building Energy Performance Is Necessary?” *Energy Policy* 64: 263–272.

Kontokosta, Constantine. 2013. “Energy Disclosure, Climate Behavior, and the Building Data Ecosystem.” *Annals of the American Academy of Political and Social Science,* 669(1): 53–73.

Kontokosta, Constantine. 2014. A market-specific methodology for a commercial building performance index. *Journal of Real Estate Finance and Economics, 1–29.*

Lam, Joseph, Kevin KW Wan, C. L. Tsang, and Liu Yang. 2008. "Building energy efficiency in different climates." *Energy Conversion and Management,* 49(8): 2354-2366.

Laustsen, Jens. 2008. *Energy Efficiency Requirements in Building Codes, Energy Efficiency Policies for New Buildings.* IEA Information Paper. Sweden.

Lee, Siew Eang, and Priyadarsini Rajagopalan. 2008. "Building energy efficiency labeling programme in Singapore." *Energy Policy,* 36(10): 3982-3992.

Lee, Wai Ling, and F. W. H. Yik. 2004 "Regulatory and voluntary approaches for enhancing building energy efficiency." *Progress in energy and combustion science,* 30(5): 477-499.

Li, Cheng, Tianzhen Hong, and Da Yan. 2014. "An insight into actual energy use and its drivers in high-performance buildings." *Applied energy,* 131: 394-410.

Li, Yu, S. Kubicki, Guerriero, and Y. Rezgui. 2019. "Review of building energy performance certification schemes towards future improvement." *Renewable and Sustainable Energy Reviews,* 113: 109244.

Madigan, Shane. 2025. “The BREEAM Rating System Explained.” *GBRI Online*.

Mathew, Paul A, Laurel Dunn, Michael Sohn, Andrea Mercado, Claudine Custudio, and Travis Walter. 2015. "Big-data for building energy performance: Lessons from assembling a very large national database of building energy use." *Applied Energy,* 140: 85-93.

McCoy, Andrew, Dong Zhao, Teni Ladipo, Philip Agee, and Yunjeong Mo. 2018. "Comparison of green home energy performance between simulation and observation: A case of Virginia, United States." *Journal of Green Building,* 13(3): 70-88.

Mims, Natalie, Steven R. Schiller, Elizabeth Stuart, Lisa Schwartz, Chris Kramer, and Richard Faesy. 2017. *Evaluation of U.S. Building Energy Benchmarking and Transparency Programs:*

Attributes, Impacts, and Best Practices. Energy Futures Group and Lawrence Berkeley National Laboratory, Electricity Markets and Policy Group.

Northeast Energy Efficiency Partnerships (NEEP). 2020. *DC Building Energy Performance Standards: Regional Policy Brief*. Lexington, MA: NEEP, 2020. https://neep.org.

Palmer, Karen L., and Margaret Walls. 2015. "Can benchmarking and disclosure laws provide incentives for energy efficiency improvements in buildings?" *Resources for the Future Discussion Paper,* 15-09.

Palmer, Karen, and Margaret Walls. 2017. "Using information to close the energy efficiency gap: a review of benchmarking and disclosure ordinances." *Energy Efficiency,* 10(3): 673-691.

Palmer, Karen, and Margaret Walls. 2015. "Limited attention and the residential energy efficiency gap." *American Economic Review,* 105(5): 192-195.

Papadopoulos, Sokratis, Bartosz Bonczak, and Constantine E. Kontokosta. 2018. "Pattern recognition in building energy performance over time using energy benchmarking data." *Applied Energy,* 221: 576-586.

Papadopoulos, Sokratis, and Constantine Kontokosta. 2019. "Grading buildings on energy performance using city benchmarking data." *Applied energy,* 233: 244-253.

Parker, Danny. 2009. "Very low energy homes in the United States: Perspectives on performance from measured data." *Energy and buildings,* 41(5): 512-520.

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao. 1995. "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data." *Journal of the American Statistical Association,* 90 (429): 106–121.

Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. 2023. "What's trending in difference-in-differences? A synthesis of the recent econometrics literature." *Journal of Econometrics,* 235(2): 2218-2244.

Ruparathna, Rajeev, Kasun Hewage, and Rehan Sadiq. 2016. "Improving the energy efficiency of the existing building stock: A critical review of commercial and institutional buildings." *Renewable and sustainable energy reviews,* 53: 1032-1045.

Ryan, Stephen P. 2012. "The Costs of Environmental Regulation in a Concentrated Industry." *Econometrica* 80 (3): 1019–1061.

Stavins, Robert, Todd Schatzki, and Jonathan Borck. 2013. "An economic perspective on building labeling policies." *Analysis Group Inc*.

Sun, Xiaojing, Marilyn A. Brown, Matt Cox, and Roderick Jackson. "Mandating better buildings: a global review of building codes and prospects for improvement in the United States." *Wiley Interdisciplinary Reviews: Energy and Environment* 5, no. 2 (2016): 188-215.

Sunderland, Louise, and Marion Santini. 2020. *Filling the Policy Gap: Minimum Energy Performance Standards for European Buildings.* Brussels: Regulatory Assistance Project, June 2020.

U.S. Department of Energy. 2021. *Annual Energy Outlook 2021*.
https://www.eia.gov/outlooks/aeo/

U.S. Environmental Protection Agency (U.S. EPA). 2022. *ENERGY STAR Portfolio Manager Technical Reference: Weather Normalization*. Washington, DC: U.S. EPA. https://www.energystar.gov/buildings/portfoliomanager/technicalreference.

Vollaro, Roberto De Lieto, Claudia Guattari, Luca Evangelisti, Gabriele Battista, Emiliano Carnielo, and Paola Gori. 2015. "Building energy performance analysis: A case study." *Energy and Buildings,* 87: 87-94.

Wang, Shengwei, Chengchu Yan, and Fu Xiao. 2012. "Quantitative energy performance assessment methods for existing buildings." *Energy and buildings,* 55: 873-888.

Webb, Amanda, and Colby McConnell. 2023. "Evaluating the feasibility of achieving building performance standards targets." *Energy and Buildings,* 288: 112989.

Wooldridge, Jeffrey M. 2007. "Inverse Probability Weighted Estimation for General Missing Data Problems." *Journal of Econometrics* 141 (2): 1281–1301.

Wooldridge, Jeffrey. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Wooldridge, Jeffrey. 2021. "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators." NBER Working Paper No. 30821.

Zhang, Yurong, Jingjing Wang, Fangfang Hu, and Yuanfeng Wang. 2017. "Comparison of evaluation standards for green building in China, Britain, United States." *Renewable and sustainable energy reviews,* 68: 262-271.

Appendix

Table A1: Anticipatory and Donut-Year Robustness Checks Using Alternative Compliance Gap Definitions

	Site EUI (kBtu/ft²)	Source EUI (kBtu/ft²)	Energy Star Score	GHG Emissions ($MTCO_2e$)	GHG Intensity ($KgCO_2e/ft^2$)
Property-Type Percentile Compliance Gaps					
A: Unweighted Estimates (2019 Policy Timing)					
Post2019 × Compliance Gap	-0.29***(0.03)	-0.44***(0.03)	0.42***(0.03)	-0.30***(0.07)	-0.68***(0.05)
B: IPW-Weighted Estimates (2019 Policy Timing)					
Post2019 × Compliance Gap	-0.28***(0.03)	-0.44***(0.04)	0.42***(0.03)	-0.57***(0.06)	-0.90***(0.05)
C: Unweighted Estimates (2020 Policy Timing)					
Post2020 × Compliance Gap	-0.27***(0.03)	-0.41***(0.03)	0.36***(0.03)	-0.28***(0.06)	-0.60***(0.04)
D: IPW-Weighted Estimates (2020 Policy Timing)					
Post2020 × Compliance Gap	-0.26***(0.03)	-0.42***(0.04)	0.37***(0.03)	-0.51***(0.05)	-0.78***(0.04)
E: Unweighted Estimates (2020 Donut-Year Robustness)					
Post × Compliance Gap	-0.29***(0.03)	-0.43***(0.04)	0.39***(0.03)	-0.34***(0.07)	-0.61***(0.04)
F: IPW-Weighted Estimates (2020 Donut-Year Robustness)					
Post × Compliance Gap	-0.28***(0.03)	-0.42***(0.04)	0.40***(0.03)	-0.51***(0.05)	-0.77***(0.04)
Citywide Mean Compliance Gaps					
A: Unweighted Estimates (2019 Policy Timing)					
Post2019 × Compliance Gap	-0.35***(0.04)	-0.47***(0.04)	0.53***(0.04)	-0.29***(0.07)	-0.75***(0.05)
B: IPW-Weighted Estimates (2019 Policy Timing)					
Post2019 × Compliance Gap	-0.34***(0.05)	-0.47***(0.05)	0.54***(0.05)	-0.60***(0.10)	-0.93***(0.04)
C: Unweighted Estimates (2020 Policy Timing)					
Post2020 × Compliance Gap	-0.34***(0.04)	-0.46***(0.04)	0.46***(0.04)	-0.27***(0.06)	-0.67***(0.04)
D: IPW-Weighted Estimates (2020 Policy Timing)					
Post2020 × Compliance Gap	-0.33***(0.04)	-0.46***(0.04)	0.47***(0.04)	-0.51***(0.10)	-0.81***(0.03)
E: Unweighted Estimates (2020 Donut-Year Robustness)					
Post × Compliance Gap	-0.33***(0.04)	-0.45***(0.05)	0.48***(0.04)	-0.31***(0.09)	-0.67***(0.04)
F: IPW-Weighted Estimates (2020 Donut-Year Robustness)					
Post × Compliance Gap	-0.32***(0.04)	-0.45***(0.05)	0.48***(0.04)	-0.51***(0.10)	-0.80***(0.03)

This table report supplementary robustness checks corresponding to Table 7, including timing robustness and donut specifications. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A2: Public-Private Subset Analysis across Compliance Gaps

	Site EUI ($kBtu/ft^2$)	**Source EUI ($kBtu/ft^2$)**	**Energy Star Score**	**GHG Emissions ($MTCO_2e$)**	**GHG Intensity ($KgCO_2e/ft^2$)**
BEPS Impacts on Private Buildings					
Property-Type Mean Compliance Gaps					
A: Unweighted Estimates					
Post × Compliance Gap	-0.34*** (0.04)	-0.44*** (0.06)	0.42***(0.04)	-0.34***(0.08)	-0.66***(0.04)
B: IPW-Weighted Estimates					
Post × Compliance Gap	-0.33*** (0.04)	-0.45***(0.06)	0.44***(0.05)	-0.55***(0.07)	-0.76***(0.02)
Property-Type Percentile Compliance Gaps					
A: Unweighted Estimates					
Post × Compliance Gap	-0.24*** (0.02)	-0.35***(0.04)	0.33***(0.03)	-0.28***(0.06)	-0.51***(0.04)
B: IPW-Weighted Estimates					
Post × Compliance Gap	-0.24*** (0.03)	-0.36***(0.04)	0.34***(0.03)	-0.44***(0.05)	-0.67***(0.03)
Citywide Mean Compliance Gaps					
A: Unweighted Estimates					
Post × Compliance Gap	-0.25*** (0.03)	-0.37***(0.04)	0.40***(0.04)	-0.25***(0.08)	-0.57***(0.04)
B: IPW-Weighted Estimates					
Post × Compliance Gap	-0.25*** (0.04)	-0.37***(0.05)	0.41***(0.04)	-0.43***(0.10)	-0.69***(0.02)
BEPS Impacts on Public Buildings					
Property-Type Mean Compliance Gaps					
A: Unweighted Estimates					
Post × Compliance Gap	-0.28*** (0.13)	-0.46***(0.27)	0.35***(0.09)	-0.42***(0.08)	-0.56***(0.22)
B: IPW-Weighted Estimates					
Post × Compliance Gap	-0.20*** (0.05)	-0.26***(0.10)	0.34***(0.09)	-0.49***(0.10)	-0.57***(0.23)
Property-Type Percentile Compliance Gaps					
A: Unweighted Estimates					
Post × Compliance Gap	-0.22*** (0.09)	-0.35*** (0.17)	0.27***(0.06)	-0.28***(0.06)	-0.42***(0.13)
B: IPW-Weighted Estimates					
Post × Compliance Gap	-0.18*** (0.04)	-0.24***(0.07)	0.27***(0.06)	-0.36***(0.08)	-0.44***(0.14)
Citywide Mean Compliance Gaps					
A: Unweighted Estimates					
Post × Compliance Gap	-0.27*** (0.12)	-0.38***(0.20)	0.26***(0.08)	-0.22***(0.06)	-0.49***(0.17)
B: IPW-Weighted Estimates					
Post × Compliance Gap	-0.21*** (0.06)	-0.23**(0.08)	0.25***(0.08)	-0.37***(0.10)	-0.49***(0.16)

This table reports continuous DiD estimates separately for public and private buildings using the three compliance gaps, with building and year fixed effects included. IPW estimates correct for non-random building attrition where indicated; overlapping federal and local energy subsidy controls are not included, and standard errors are clustered at the building level.