Greening the Blueprint: Evaluating the Effectiveness of Building Performance Standards in Two American cities

Christopher N. Mesfin

Advised by Professor Robert Stavins

Thursday August 8th, 2024

Abstract

This project presents an early empirical analysis of the effectiveness of Building Performance Standards (BPS) on energy efficiency, focusing on case studies from Boulder, Colorado, and Philadelphia, Pennsylvania. BPSs are policies that mandate existing buildings in an area meet a standard related to greenhouse gas emission or energy efficiency within a given time period. These policies have recently become popular in large municipalities where emissions from buildings make up a sizable proportion of the area's emissions. Because these policies are new, their benefits and costs are largely speculative and based on engineering estimates that have infamously improperly predicted environmental impact in the past. Using a regression discontinuity design (RDD) to exploit the floor area cutoff sizes mandated by the BPS, the study evaluates changes in Energy Use Intensity (EUI) and greenhouse gas (GHG) emissions. The results indicate no significant impact of the BPS compliance deadlines on energy efficiency or emissions in both cities. The empirical analysis highlights potential challenges estimating the effect of a BPS such as data limitations, early compliance variability, and possible manipulation of building size reporting. The findings underscore the need for robust compliance monitoring, comprehensive data collection, and further longitudinal studies to fully assess the effectiveness of BPS policies. This study contributes to the growing body of literature on building energy regulations and provides insights for policymakers aiming to enhance the design of BPS to achieve greater energy efficiency and emission reductions.

1 INTRODUCTION

1.1 Context

2023 marked another year of record high temperatures, underscoring the intensifying effect of global warming on our climate system (Stueck, 2024). Across jurisdictions and over time, scientists, economists, and professionals alike have advocated for dramatic reductions in greenhouse gas (GHG) emissions as the primary means of mitigating rising temperatures and the threats they pose to the Earth.

In the US, the impacts of climate change are increasingly pervasive, with natural disasters becoming more frequent and growing seasons fluctuating. The realization of these impacts may also be driving the pursuit of more aggressive solutions. The current Biden administration has introduced the Inflation Reduction Act (IRA), which may prove to be one of the most important policies for meeting the US's GHG emission goals. This is alongside a suite of regulations, ranging from controlling carbon emissions from power plants to requiring publicly traded companies to disclose emissions data.

1.2 The Role of Building Performance Standards

One policy that has grown in recognition and use in the US is the Building Performance Standard (BPS). BPS, also known as Building Energy Performance Standards (BEPS) or Minimum Energy Performance Standards (MEPS), seeks to address emissions from existing buildings in the regulated area, differing from building codes that set standards for new construction. Building emissions are significant in the US, as the building stock has accounted for about 40% of all energy use (Robinson et al., 2017). On January 21, 2022, the Biden Administration launched the National BPS Coalition in an effort to explore solutions for making the existing building stock more energy efficient. As of July 2024, 45 states and municipalities have joined the coalition, with actionable plans to create or have already created and enacted a BPS.

Additionally, as of the end of 2023, over 50 municipalities and states have implemented Commercial (and in some cases large residential) Energy Benchmarking and Transparency Policies. These policies are well-known predecessors to the BPS but are very distinct, as Benchmarking standards do not have the performance expectation of a BPS. However, they certainly inform a municipality on what performance standards may be feasible.

A BPS mandates performance improvement in existing buildings by requiring them to reach a certain level of performance by a specified date. In the US, there has been considerable variation in how jurisdictions decide on the level of performance and the compliance date. Performance metrics often include GHG emissions (using utility data to estimate emissions from energy use) or energy efficiency through energy use intensity (kbtu/ft^2). Additionally, some BPS use other mechanisms for evaluating performance, such as requiring buildings to perform specific tasks like improving lighting efficiency.

1.3 International Adoption of BPS

BPS have demonstrated substantial energy and emission savings. For instance, Tokyo's BPS has achieved a 27% overall reduction in GHG emissions since its inception. Recognizing their potential, several jurisdictions worldwide have enacted BPS. Tokyo was the earliest adopter in 2010, and since then, several EU member states, such as France and the Netherlands, as well as parts of the UK, have followed suit. In the US, 10 jurisdictions, including Washington DC and New York City, have enacted BPS, with 33 state and local governments pledging to advance BPS legislation by April 2024 as part of the National BPS Coalition.

The design of BPS varies significantly across jurisdictions. In the EU, BPS often target the worst-performing buildings, while in the US, they tend to target a larger proportion of buildings. For example, France's BPS requires residential buildings with the worst EPC ratings (F and G) to upgrade to E levels, affecting around 15% of the housing stock. In contrast, Washington DC's BPS targets the median performance, meaning 50% of buildings will need to improve.

1.4 Challenges and Research Gaps

While BPS policy adoption is accelerating, research to support it lags behind. Most existing studies on BPS provide guidance on policy design and implementation, outlining key components and considerations. However, they lack comprehensive analysis quantifying the costs and benefits of meeting BPS targets. Empirical studies have demonstrated the feasibility of achieving significant energy savings, but achieving BPS targets at scale remains a challenge.

The lack of empirical research into Building Performance Standards may be partly due to the lack of empirical data to analyze. The vast majority of BPS in the US have yet to pass compliance deadlines, which has forced researchers to rely on modeling to draw conclusions about the effectiveness of BPS.

The goal of this paper is to empirically analyze the effectiveness of building performance standards in the US, specifically looking at standards in Boulder, Colorado, and Philadelphia, Pennsylvania, for evidence.

1.5 Boulder, Colorado Building Performance Ordinance

The Boulder Building Performance Ordinance (BPO) requires buildings in Boulder to comply with phased reporting requirements, starting with large buildings (>50,000 sq. ft.), followed by medium-sized buildings (>30,000 sq. ft. but <49,999 sq. ft.), and small buildings (>20,000 sq. ft. but <29,999 sq. ft.). The policy also includes city buildings (>10,000 sq. ft.) and new buildings (>10,000 sq. ft.). Unique standards are set for "Large Industrial Campuses," which are defined as three or more buildings, at least partially used for manufacturing, and served by a central plant or single utility meter.

In addition to meeting reporting requirements, buildings must also undergo several energy efficiency measures. These include completing an ASHRAE Level II-equivalent Energy Assessment, conducting a lighting upgrade, and retrocommissioning the building. Subsequently, they will implement the cost-effective measures recommended in the retrocommissioning report.

The implementation of these measures will follow a phased timeline, as detailed below for the three building classes:

Regulated Buildings (Private)	Rate and Report*	Energy Assessment**	Lighting Upgrade*** and Retrocommissioning**	Cost-Effective Measures**
>=50,000 sf	June 1st, 2016	June 1st, 2019	June 1st, 2022	June 1st, 2024
[30,000 sf – 50,000 sf)	June 1st, 2018	June 1st, 2021	June 1st, 2024	June 1st, 2026
[20,000 sf – 30,000 sf)	June 1st, 2020	June 1st, 2023	June 1st, 2026	June 1st, 2028

* Rating and Reporting happens on an annual basis after compliance deadline.
** Energy Assessment, Retrocommissioning, and Cost-Effective Measures are to be repeated after 10 years.
*** Lighting upgrade is one-time and not repeated and consists of lighting upgrades mentioned in the the ASHRAE level 2 audit.

1.6 Philadelphia, Pennsylvania Building Energy Performance Program

The Philadelphia, Pennsylvania Building Energy Performance Program (BEPP) requires large private buildings (>50,000 sq. ft.) to comply with "tune-up" standards specified in their regulation. The legislation is vague about the exact definition of a tune-up but includes a list of possible actions such as fixing design issues related to energy efficiency, updating lighting, updating equipment controls, and updating ventilation. Individual buildings must submit a report to the Philadelphia Department of Sustainability, which determines if the building's tune-ups comply with BEPP. The legislation allows the Department of Sustainability to create regulations that further define the threshold for tune-ups. Tune-ups consist of two components: an inspection and corrective measures. Below is information on the compliance timeline:

Regulated Buildings (Private)	Reporting Deadline	First Tune-Up Deadline*
>=200,000 sf	September 30th, 2019	September 30th, 2021
[100,000 sf – 200,000 sf)	September 30th, 2019	September 30th, 2022
[70,000 sf – 100,000 sf)	September 30th, 2019	September 30th, 2023
[50,000 sf – 70,000 sf)	September 30th, 2019	September 30th, 2024

*This date is when buildings of the associated size are required to submit reports for the tune-ups they have conducted with 2 years of their deadline.

2 DATA

Data used to empirically evaluate these building performance standards was obtained from the respective city governments in collaboration with a data manager who stores and facilitates the collection of data for the municipality. Boulder's reporting mandate required public disclosure of the dataset but anonymized specific information about the buildings reporting the data. Philadelphia's dataset included information such as a building's address. These two cases are typical for reporting standards in the US.

Both datasets were reported to the cities a yearly basis, but Boulder only reported measures of Energy Use Intensity (EUI), whereas Philadelphia included information on both source EUI and site EUI. Site EUI is the amount of energy consumed at the building site, as reflected in the owner's utility bills. It's calculated by dividing the amount of site energy used in a year by the total square footage of the building. Site EUI is what is most typically required by reporting standards. Source EUI tracks energy use back to the power plant, measuring the amount of raw energy used to operate the building. It accounts for site energy, as well as energy lost during production, transmission, and delivery. The Philadelphia Standard also included total energy use data and total GHG emission data. Both datasets used weather-normalized EUI, which controls for extreme weather events that may impact the energy usage of a building. Below are more specifics about the building stocks of Philadelphia and Boulder, along with summary graphs:

2.1 Philadelphia

There was a total of 1,363 buildings that adhered to the Philadelphia Reporting Standard in 2023. Additionally, 705 buildings were required to comply but did not. The smallest of these buildings was under 1,000 sq. ft., and the largest was 13,000,000 sq. ft. In Figure 1, you can see a collection of the five most common use types for reporting buildings.

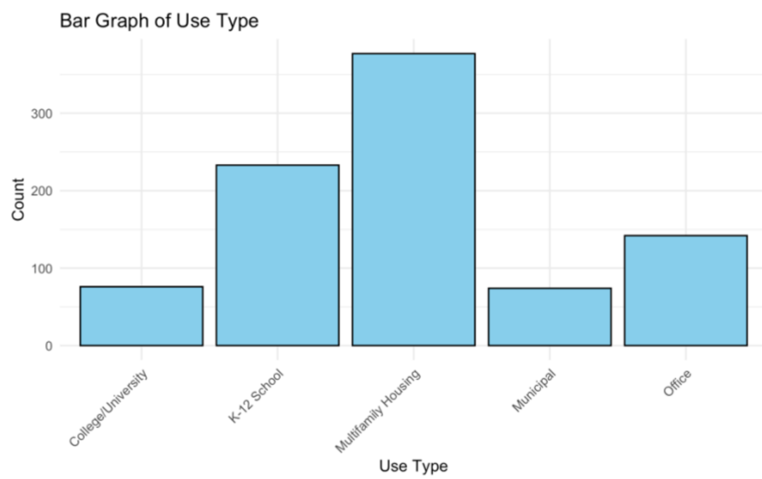

Figure 1

Figure 2 shows the average EUI across the reported building stock in Philadelphia. The dotted vertical lines indicate when required reporting began (red) and when the first compliance date was set. The graph shows an increase in EUI or a decrease in energy efficiency, which has multiple explanations that I will discuss. GHG emissions trends and electricity use trends look very similar to the curve below.

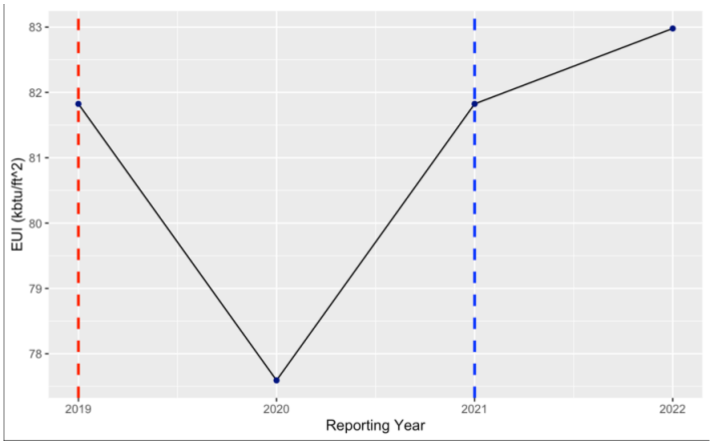

Figure 2

2.2 Boulder

There were total of 474 buildings that adhered to the Boulder Reporting Standard in 2023. The smallest of these buildings was 1,958 sq. ft., and the largest was 737,421 sq. ft. In Graph 3, you can see a collection of the five most common use types for reporting buildings.

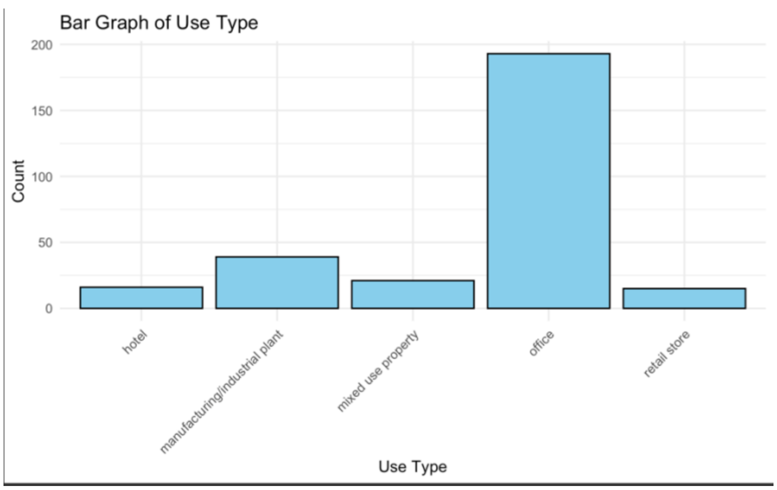

Figure 3

Figure 4 for Boulder is similar to Figure 3 for Philadelphia and looks at the average EUI across the reported building stock in Boulder. The dotted vertical lines indicate when required reporting began (red) and when the first compliance date was set. On average, buildings in Boulder, CO are less energy efficient than buildings in Philadelphia, PA.

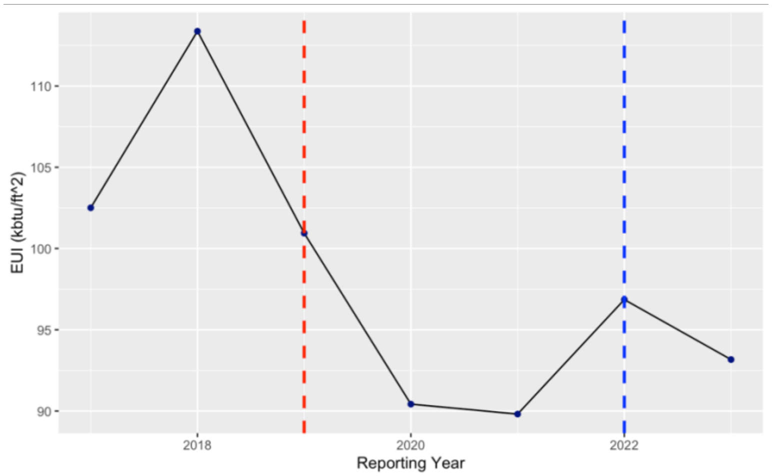

Figure 4

3 METHODS

To evaluate the effectiveness of the BPS, I decided to use a regression discontinuity design (RDD) by exploiting the floor area cutoff size for the RDD. This method estimates the local average treatment effect (LATE) using the RDD. The main estimation equation controls for building fixed effects, incorporating a smooth control function for the running variable (building size).

For this analysis, I used cross-sectional portions of the datasets based on compliance year, which was 2022 for both Boulder and Philadelphia.

The validity of the RDD was tested by checking for discontinuities in observable characteristics across building size, primarily building type, to ensure that there wasn't a specific type of building over or below the threshold that would skew the RDD. This analysis was primarily done graphically using scatterplots.

Given the data, this method was the most viable, as the low frequency of sampling rendered any method related to using the available time series data infeasible. RDD was the most feasible because of the establishment of a clear cutoff in the BPS. That said, RDD is not without its drawbacks. The method requires as much data immediately surrounding the threshold as possible to find significant effects, which is a complication in this case due to the low quantity of available data.

Other issues may be related to early compliance with the BPS. The RDD assumes that compliance is confined to the date the BPS mandates a building of a certain size to make a change. Buildings have the autonomy to comply with this regulation earlier, but, at least in the cases of Boulder and Philadelphia, there is no measurement of exactly when a building complies; there is only a measurement of whether they did by the time they were required. This means that all of the following regression results must be analyzed with the caveat that we are assuming buildings make changes in the year they are required. This makes the RDD fuzzy, as the treatment effect (complying with RDD regulations) is not completely determined by the threshold of building performance.

4 RESULTS

4.1 Philadelphia

Figure 5 (EUI) and Figure 6 (GHG) are the regression tables from the regression discontinuity design.

Figure 5 consists of four regressions in columns 1-4, analyzing the effect of regulation (big_100) on Energy Use Intensity (EUI). Big_100 is a dummy variable indicating whether a particular building is regulated. A building is considered regulated if it is larger than 200,000 ft^2, according to the Philadelphia BPS. The variable Center is used to center the running variable (building size) so that a1 may represent a true treatment effect. Use_type is a control variable that allows us to account for the type of building (e.g., Office, University, Bank).

$$Y_i = a_0 + a_1\ big\_100 + a_2\ center + a_3\ use\_type + a_4\ (big\_100 \times center)$$

The same regression equation is used across each regression in Table 1 and Table 2. The difference between each regression is its bandwidth. Regression (1) has a bandwidth of 50,000 ft^2, which allows more observations, increasing power but deviating from the threshold. Regression (2) has a bandwidth of 25,000 ft^2, (3) has 10,000 ft^2, and (4) has 5,000 ft^2. I saw no need to include variation in the order of the control functions, as the scatterplot of the data indicates that a higher-order control function would not change the lack of discontinuity at the threshold.

The a_1 coefficients in Figure 5 offer little besides correlation, as shown by large standard errors, T-statistics < 1.96, p-values > 0.05, and 95% confidence intervals that include 0. This indicates that the compliance date of the building performance standard had little to no effect on energy efficiency or that there was not enough data available to make a conclusion.

	(1)	(2)	(3)	(4)
VARIABLES	weather_site_eui	weather_site_eui	weather_site_eui	weather_site_eui
big_100	8.39	-4.1	32.19	76.07
Standard Error	(20.11)	(32.99)	(56.30)	(55.67)
T-Statistics	0.42	-0.12	0.57	1.37
Pvalue	0.67	0.901	0.572	0.199
Confidence Interval	[-31.26, 48.04]	[-69.66, 61.46]	[-82.78, 147.1]	[-46.47, 198.6]
Bandwidth	50,000	25,000	10,000	5,000
Building Use Controls	Yes	Yes	Yes	Yes
Control Function	Linear	Linear	Linear	Linear
Observations	202	92	35	16
R-squared	0.009	0.034	0.050	0.082
Robust standard errors in parentheses				
*** p<0.01, ** p<0.05, * p<0.1				

Figure 5

The same regression equation is used for Figure 6 with a different dependent variable being GHG emissions (which was measured in $MTCO2_e$). This regression yielded very similar results in terms of causality as the bias and power tradeoff is demonstrated by increases in bias following smaller sample sizes with relatively greater significant results.

	(1)	(2)	(3)	(4)
VARIABLES	total_ghg	total_ghg	total_ghg	total_ghg
big_100	9.017	-56.40	268.3	1,059
Standard Error	(268.5)	(431.3)	(825.8)	(813.7)
T-Statistics	0.03	-0.13	0.32	1.3
P-value	0.973	0.89	0.747	0.22
Confidence Interval	[-520.5, 538.5]	[-913.6, 800.8]	[-1418, 1954]	[-732.3, 2849]
Bandwidth	50,000	25,000	10,000	5,000
Building Use Controls	Yes	Yes	Yes	Yes
Control Fucntion	Linear	Linear	Linear	Linear
Observations	202	92	35	16
R-squared	0.090	0.078	0.060	0.084
Robust standard errors in parentheses				
*** p<0.01, ** p<0.05, * p<0.1				

Figure 6

4.2 Boulder

Figure 7 is the regression table from the regression discontinuity design.

Figure 7 consists of four regressions in columns 1-4, analyzing the effect of regulation (big_50) on Energy Use Intensity (EUI). A building is considered regulated if it is larger than 50,000 ft^2, according to the Boulder, CO Building Performance Ordinance (BPO). The regression equation used is the same as for Philadelphia with the same controls.

$$Y_i = a_0 + a_1\ big\_50 + a_2\ center + a_3\ use\_type + a_4\ (big\_50\ x\ center)$$

Regression (1) has a bandwidth of 20,000 ft^2, which allows more observations, increasing power but deviating from the threshold. Regression (2) has a bandwidth of 10,000 ft^2, (3) has 5,000 ft^2, and (4) has 2,500 ft^2.

The a_1 coefficients in Figure 7, in a similar fashion as before, show little besides correlation, as shown by large standard errors, T-statistics < 1.96, p-values > 0.05, and 95% confidence intervals that include 0. This indicates that the compliance date of the building performance standard had little to no effect on energy efficiency or that there was not enough data available to make a conclusion.

	(1)	(2)	(3)	(4)
Dependent Variables	EUI_2021 (kbtu/ft)	EUI_2021 (kbtu/ft)	EUI_2021 (kbtu/ft)	EUI_2021 (kbtu/ft)
big_50	15.00	49.36	62.22	100.2
Standard Error	(21.24)	(31.69)	(47.13)	(71.99)
T-statistic	0.71	1.56	1.32	1.39
Pvalue	0.481	0.125	0.198	0.207
Confidence Interval	[-26.95, 56.94]	[-14.01, 112.7]	[-34.66, 159.1]	[-70.08, 270.3]
Bandwidth (sqft)	20,000	10,000	5,000	2,500
Building Use Controls	Yes	Yes	Yes	Yes
Control Function	Linear	Linear	Linear	Linear
Observations	161	65	31	12
R-squared	0.010	0.062	0.152	0.280

Figure 7

4.3 Additional Considerations

Figure 8 shows a McCrary test conducted on the Boulder data to identify if there could have been manipulation of the running variable (building size). This test analyzes the density of running variable observations on either side of the RDD threshold and identifies if it is possible that entities are able to manipulate which side of the threshold they fall into (McCrary, 2006). As you can see in the McCrary test, there is a discontinuity at the 50,000 ft^2 threshold. Specifically, the coefficient for the discontinuity was -0.9458 with a standard error of 0.1834. Dividing these two numbers, we get -5.157, which has a much greater magnitude than the typical used 1.96 critical value, indicating that we cannot reject the hypothesis that there is manipulation of the running variable in our case.

In application, this means that it is possible for buildings to alter their reported size and go from regulated to unregulated. While this isn't technically possible, because building sizes are self-reported by building owners and they knew the cutoffs for regulation before reporting, it is reasonable to assume that some building owners may lie about the size of their building to avoid regulation. If this were true, it would jeopardize the pseudo-randomized experiment created by the RDD, as the treatment (regulation) would not be randomly assigned.

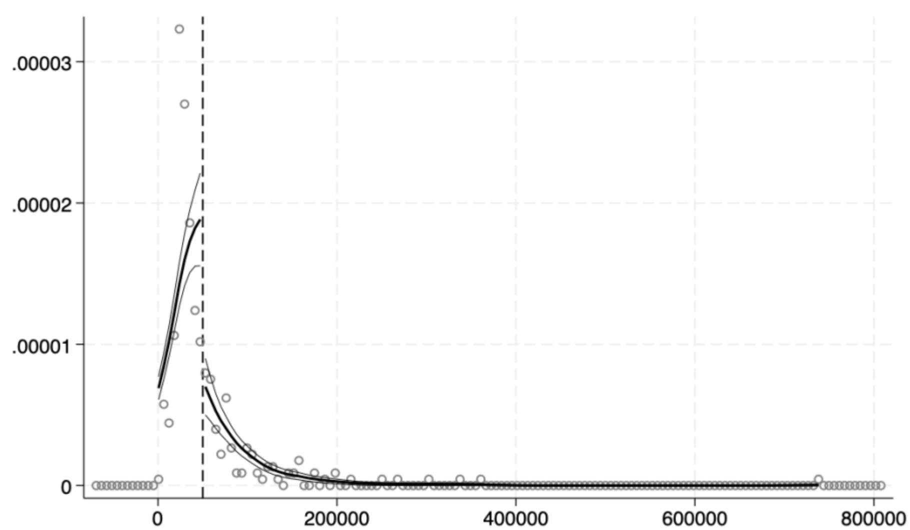

Figure 8

Beyond manipulation of the running variable is it possible that a lack of enforcement or lax standards are causing the lack of treatment effect. Both the Boulder and Philadelphia standards do not possess standards that mandate buildings in the city reach a certain level of GHG production or EUI (i.e. Buildings greater than 50,000 ft^2 must reduce emissions by x% from 2012 levels by 2030). Rather these standards require the implementations of specific technologies or specific actions (i.e. Perform retro commissioning by 2022). These "technology-based" standards do not make up a majority of active BPSs as the largest municipalities across the US that have a BPS (Boston, New York City, Washington DC) all have standards that decrease allowed levels of GHG emission or increase required energy efficiency.

5 DISCUSSION

5.1 Summary of Findings

The empirical analysis of the Building Performance Standards (BPS) in Boulder, Colorado, and Philadelphia, Pennsylvania, reveals mixed results concerning the effectiveness of these regulations in enhancing energy efficiency. The regression discontinuity design (RDD) employed in this study did not find significant evidence to suggest that the BPS compliance deadlines had a marked impact on Energy Use Intensity (EUI) or greenhouse gas (GHG) emissions in both cities.

In Philadelphia, the regression results indicate no significant treatment effect on EUI or GHG emissions at the threshold of 200,000 square feet. This suggests that either the BPS compliance date had no substantial impact on energy efficiency, or there was insufficient data to draw definitive conclusions. Similarly, in Boulder, the RDD analysis revealed no significant treatment effect at the 50,000 square feet threshold, and the McCrary test suggested possible manipulation of the building size variable, indicating that some buildings may have reported sizes strategically to avoid regulation.

5. 2 Implications

The lack of significant findings in this study points to several important implications for policymakers and researchers:

1. **Compliance and Enforcement Challenges**: The potential manipulation of building size data in Boulder highlights a critical challenge in the enforcement of BPS. Effective compliance monitoring and verification mechanisms are essential to ensure that buildings adhere to the regulations and report accurate data.
2. **Early Compliance and Data Limitations**: The assumption that buildings comply precisely by the deadline may not hold true in practice, as some buildings might comply earlier or later. This variability can obscure the true impact of BPS. Additionally, the limited amount of available data immediately surrounding the compliance thresholds complicates the ability to detect significant effects.
3. **Need for Comprehensive Data**: The differences in data availability and reporting standards between Boulder and Philadelphia underscore the need for standardized and comprehensive data collection practices. Consistent reporting of both site and source EUI, along with detailed GHG emissions data, would enhance the ability to conduct robust empirical analyses.

5. 3 Future Research Directions

Given the limitations and findings of this study, several avenues for future research can be pursued:

1. **Longitudinal Studies**: Future research could benefit from longitudinal studies that track buildings over extended periods, capturing the effects of BPS over time rather than focusing solely on compliance deadlines. This approach would provide a more comprehensive understanding of the long-term impacts of BPS on energy efficiency and emissions.
2. **Behavioral Insights**: Understanding the behavioral responses of building owners and managers to BPS is crucial. Research into the motivations, barriers, and incentives for compliance can inform the development of more effective policy instruments.
3. **Similar Analysis**: The analysis conducted in this paper was a very early look at the impact of only a few BPS. Waiting for more data from larger municipalities could allow a better conclusion to be formed about the effectiveness of the BPS.

5. 4 Conclusion

The analysis of BPS in Boulder and Philadelphia presents a nuanced picture of the challenges and opportunities associated with regulating building energy performance. While the study did not find significant evidence of the effectiveness of BPS in the examined cities, it highlights critical areas for improvement in policy design, enforcement, and research methodologies. As jurisdictions worldwide continue to adopt and refine BPS, ongoing empirical research will be vital to ensuring these standards achieve their intended goals of enhancing energy efficiency and reducing emissions.

References

1. Stueck, W. (2024). With 2023 the hottest year on record, world on track to eclipse 1.5 C threshold; “Persistently” high temperatures set off string of climate milestones. *Globe & Mail (Toronto, Canada)*, A1–A1.
2. Robinson, C., Dilkina, B., Hubbs, J., Zhang, W., Guhathakurta, S., Brown, M. A., & Pendyala, R. M. (2017). Machine learning approaches for estimating commercial building energy consumption. Applied Energy, 208, 889–904. https://doi.org/10.1016/j.apenergy.2017.09.060
3. Ito, K. (2015). Asymmetric Incentives in Subsidies: Evidence from a Large-Scale Electricity Rebate Program. *American Economic Journal: Economic Policy*, *7*(3), 209–237. https://doi.org/10.1257/pol.20130397
4. Andrews, A., & Jain, R. K. (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. https://doi.org/10.1016/j.jobe.2023.107375
5. Bergfeld, K., Mathew, P., Duer-Balkind, M., Perakis, J., Noorikhah, P., Walter, T., & Held, A. (2020). *Making Data-Driven Policy Decisions for the Nation’s First Building Energy Performance Standards*. https://doi.org/10.20357/B7831V
6. Nadel, S., & Hinge, A. (n.d.). *Mandatory Building Performance Standards: A Key Policy for Achieving Climate Goals*.
7. McCrary, J. (n.d.). *Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test*.