The Data Sources for
Harming Tenants: The Impact of the 2019 Housing Stability and Tenant Protection Act on New York's Affordable Housing

Jason M. Barr July 30, 2025

Costs vs. Rent Increases

The figure creates an operating cost index from the Rent Guidelines Board (RGB) from 2012 to 2024, which estimates the percentage change in rental building operating expenses each year. The RGB gives percent increases. To make an index, I assume that year 2012 is set to 1 and then each year the index is determined by the formula:

$$Cost\ Index_t = (1 + Cost\ Growth\ Rate_t) \times Cost\ Index_{t-1}$$

Similarly, the RGB designates rent increases each year. They give increases for those who sign one-year and two-year leases, respectively. For the one-year lease, I create a rental index in a similar fashion to the cost index:

$$Rent\ Index_t = (1 + Rent\ Increase\ Rate_t) \times Rent\ Index_{t-1},$$

where for t=2012 $Rent\ Index_t = 1$.

For two-year leases, I assume a person signs a lease in 2012 and then apply the two-year rent increases from 2013 onward, where increases come every other year (or applied annually where the increases are specified separately for one-year intervals). However, given the two-year increases are generally double the one-year increase but apply for two years, the net effect on housing costs over the long run is similar.

The sources for the data are here: https://rentguidelinesboard.cityofnewyork.us/research/.

Statistical Analysis Data

Class C Violations. From the Building Violations database from the NYC Department of Housing Preservation and Development (HPD), which contains (as best I can tell) a complete list of building violations issued in the city. It can be downloaded from here: https://data.cityofnewyork.us/Housing-Development/Housing-Maintenance-Code-Violations/wvxf-dwi5/about_data.

After downloading the data, I kept only Class C violations, the most extreme type of violations. Class C violations are classified as "immediately hazardous" violations. They include inadequate heat in the winter months, the presence of rodents, peeling lead paint in homes with small children, no hot water, mold and mildew. Property owners must correct these issues within 24 hours of receiving the Notice of Violation.

Since HSTPA went into effect on June 14, 2019, I look at violations based on the initial inspection dates from January 1, 2015 to December 31, 2024.

My unit of analysis is the BBL (Borough-Block-Lot number). The BBL assigns a unique ID to every tax lot in the city. Given that other data sets have data at the BBL level, this variable was chosen over the building level. In most cases, the tax lot (BBL) has only a single structure on it, but in some cases, there might be multiple buildings, if the lot contains, for example, a housing complex.

For each BBL, I added up the number of Class C violations for each quarter of each year. Call this the *Class C-Year_Q* data set.

In a new file, I created a list of all BBLs in NYC as of 2016 (see PLUTO File from below). For each BBL, I created a panel of BBLs for each year quarter, from Q1 2015 to Q4 2024. I then merged the *Class C-Year_Q* file with the list of BBLs for each year quarter. For BBLs in a year-quarter without any violations I coded as 0. Call this the BBL-Year-Q data the *BBL Panel.*

The PLUTO File: Next, I downloaded the 2016 Primary Land Use Tax Output (PLUTO) file, which contains detailed data on the structure of every BBL in NYC. From this data set, I retained for each BBL, the year the structure was built, its current building use category (see below), the community district number, the residential floor area, the number of floors, and the number of residential units.

I then merged the PLUTO data to the *BBL Panel*. This created a balanced panel of number of Class C violations along with building, lot, and location data

Rent Stabilized Buildings: New York does not keep a comprehensive list of which units are rent stabilized (or rent controlled). Rather it keeps a list of which buildings have at least one rent stabilized unit. I used the downloaded list of stabilized rental buildings as of 2017. By BBL, I created a list of which buildings were rent stabilized (=1) versus those that were not (=0). See: https://rentguidelinesboard.cityofnewyork.us/resources/rent-stabilized-building-lists/ for the current list of buildings. Note the assumption here is that the building remains always with at least one rent stabilized unit, which seems reasonable.

The Econometric Model

The econometric model is an event study/difference-in-difference specification. That is to say, we compare two groups of buildings the treatment (=1) versus the control (=0). I used two sets of controls: rent stabilized buildings vs. non-rent stabilized apartment buildings, and rental apartment buildings versus condos and coops, and. The data set only contains buildings with more than five residential units per BBL.

The first model is a standard D-i-D specification

$$V_{it} = \alpha_0 + \alpha_1 RS_i + \beta(RS_i \times HSTPA) + \theta(Rental_i \times HSTPA) + \sum_{q=2}^{4} \rho_q [Q = 1]_{\tau q} + \sum_{\tau=2014.1}^{2024.4} \delta_\tau [t = 1]_\tau + \sum_{j=1}^{59} \sum_{\tau=2014.1}^{2024.4} \gamma_{jt} [CD_j = 1][t = 1]_\tau + \sum_{j=1}^{59} \rho_j [CD_j = 1] + X_i + \mu_{it},$$

where V_{it} is the number of Class C violations for BBL i at period t. α_0 is the constant. α_1 is the shift in the constant for the rent-stabilized treatment group (RS_i), which is a dummy that takes on the value of 1 if the BBL is in the respective treatment group, 0 otherwise. β is the variable of interest, it gives the average number of additional violations in the rent stabilized group after HSTPA went into effect (2019.2) (i.e. HSTPA=1 after Q2 2019, 0 otherwise). CD_j, are dummies for each of the 59 Community Districts

$\theta$is the effect of a rental building after HSTPA goes into effect. This is included in a second regression for all multifamily buildings, not just rentals. In this case β is the effect of rent stabilized buildings vs. all types of multifamily buildings. δ_τ, for $\tau = 2015.1 \ldots 2024.4$, are the coefficients for the impact of each year-q (i.e., dummies for each period to account for general trends on number of Class C violations.). $\sum_{q=2}^{4} \rho_q [Q = 1]_{\tau q}$ are quarter dummies to account for seasonal differences.

Further, I interact Community District dummies with year-q dummies to account for differing trends across the city neighborhoods. X_i are BBL-specific variables. In my preferred specification, I use block dummies (fixed effects) that control for fixed block characteristics. X_i includes number of units, the number of floors, the buildings age, a dummy for 2020 and 2021 for the COVID Pandemic.

The second type of model is an event study, that takes the standard D-i-D model but splits up $\beta(RS_{i\tau} \times HSTPA)$ into treatment x year-q dummies for each period. If HSTPA is having a causal impact on Class C violations, we should expect to see no trend in the betas before Q2 2029, but a trend after that. The full specification is given by:

$$V_{it} = \alpha_0 + \alpha_1 RS_i + \sum_{\tau=2015.1}^{\tau=2024.4} \beta_\tau (RS_i \times [t=1]_\tau) + \sum_{\tau=2015.1}^{\tau=2024.4} \theta_\tau (Rental_i \times [t=1]_\tau) + \sum_{q=2}^{4} \rho_q [Q=1]_{\tau q} + \sum_{\tau=2015.1}^{2024.4} \delta_\tau [t=1]_\tau + \sum_{j=1}^{59} \sum_{\tau=2015.1}^{2024.4} \gamma_{jt} [CD_j = 1][t=1]_\tau + \rho_j \sum_{j=1}^{59} CD_j + X_i + \mu_{it}$$

Descriptive Statistics and Graphs

The final data set is a balanced panel, where each year has 62,778 and the total data set from 2015.1 to 2024.4 has 2,448,342 observations. From my data set, I determined that fraction of multifamily buildings (here set as 6 or more units per BBL) that have at least one Class C violation. The results are given in Figure 1, which shows a rise in violations in 2019, a drop during COVID and then a sharp rise after that.

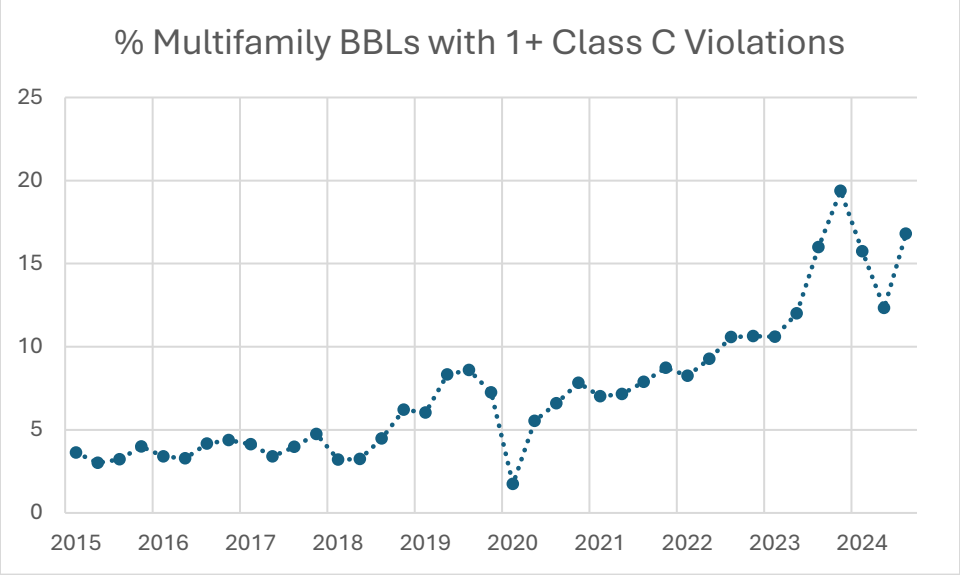
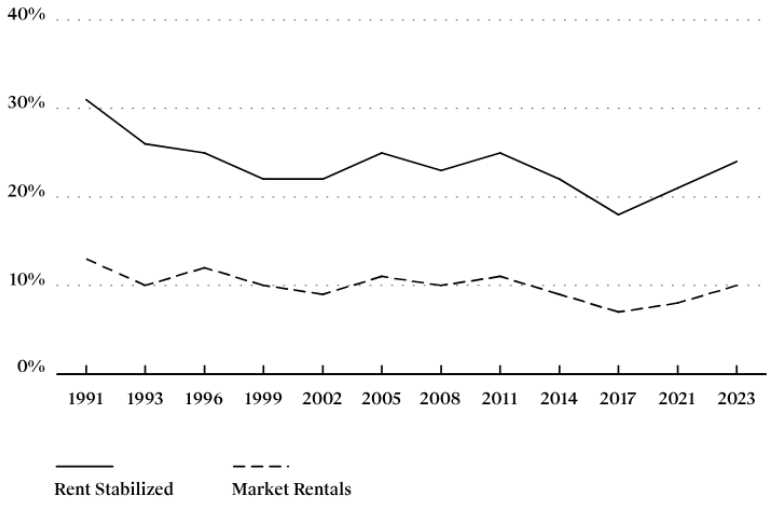


Figure 1: All Residential Buildings

Also note this graph correlations with the findings in the 2023 NYC Housing & Vacancy Survey (https://www.nyc.gov/assets/hpd/downloads/pdfs/about/2023-nychvs-selected-initial-findings.pdf). In the report's Figure 13, we see a rise housing problems particularly since 2017.

Figure 13. Units Reporting 3+ Housing Problems for
Rent Stabilized Units and Market Rentals: 1991 – 2023



Source: New York City Housing and Vacancy Survey (NYCHVS), 1991 – 2023.
Please refer to the technical appendices at the end of this report for details on the NYCHVS design, definitions, and methodology.

Descriptive Statistics

Table 1 provides descriptive statistics for the dataset.

Variable	Obs	Mean	Std. dev.	Min	Max
# Class viols	2,448,342	.0735236	1.11431	0	374
covid dummmy	2,448,342	.6923077	.4615386	0	1
q2	2,448,342	.2564103	.436651	0	1
q3	2,448,342	.2564103	.436651	0	1
q4	2,448,342	.2564103	.436651	0	1
Units res	2,448,342	36.27946	139.163	6	13248
# floors	2,448,342	5.330781	4.34586	0	90
Rental apt	2,448,342	.882634	.3218559	0	1
year built	2,444,481	1933.154	30.6134	1800	2016

Note: Obs. are at the tax lot level, which may include more than one building.

Regressions Results

Table 2 gives the D-i-D Results for several specifications.

Table 2: D-i-D Regression Results. Dep. Var. # of Class C Violations per Quarter, 2015 to 2024. All specifications include year-quarter (YQ) dummies, community district CD) dummies, and YQ-CD interaction terms to control for differing trends across neighborhoods. Not regressions only include BBLs with less than 200 violations in a given year-quarter. Standard errors below estimates. RS=rent stabilized building dummy. Rental=Multifamily rental building dummy.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Post 2019 x RS	0.026***	0.026***	0.026***	0.022***	0.022***	0.022***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
RS Building	0.022***	0.016***		0.022***	0.017***	
	(0.004)	(0.004)		(0.004)	(0.004)	
Post 2019 x Rental				0.023***	0.023***	0.023***
				(0.004)	(0.004)	(0.004)
Rental buildings only	Yes	Yes	Yes	No	No	No
All Multifamily	No	No	No	Yes	Yes	Yes
BBL FE's	No	No	Yes	No	No	Yes
Block Fes	No	Yes	No	No	Yes	No
F-Stat P-value for FEs		0.00	1.00		0.00	1.00
Additional Controls	Yes	Yes	N/A	Yes	Yes	N/A
R-sq	0.01	0.01	0.02	0.01	0.01	0.02
Nobs.	2,158,804	2,158,804	2,158,804	2,444,479	2,444,479	2,444,479

Equations (1) – (3) compare rent stabilized buildings to non-rent stabilized multifamily apartment buildings (6+ units) per BBL. Eq. (1) regresses Class C violations on several controls, including a COVID dummy (2020-2021), quarterly dummies, number of residential units, number of floors, the year built, and building type dummies (From PLUTO Building Class Categories). Eq. (2) is same as (1) but includes block level dummies (fixed effects). Eq. (3) includes BBL dummies, so building level variables drop out.

The D-i-D regressions suggest that after 2019Q2, there was a rise in 0.026 violations per BBL (recall that most buildings had zero violations in a given year-quarter.)

Equations (4) – (5) employ the same variables. The only difference is that the control group is all multifamily buildings (BBLs) and an additional set of controls is added: The interaction of a rental building dummy x post-2019Q2. The results are quite similar.

Also, note that all results present unadjusted standard errors. While I suspect there is heteroskedasticity, my computer is unable to compute robust standard errors in Stata with >2 million observations—given the great need for computing power to do this. Having said this, I do not think it would materially affect the conclusions if some type of robust or clustered standard errors were used.

Finally, another set of regression was run using year-q x rent stabilized building dummies to perform a event study. The coefficients and standard errors (using equation (2) and (4), respectively) are given in the Blog Post. The results suggest that after 2022 there was a strong rise in Class C violations driven by HSTPA.