Regression Results for New York City Building Height and Green House Gas Emissions

Jason Barr
Rutgers University-Newark
May 30, 2018

1. The Data

The first data set comes from http://www.nyc.gov/html/gbee/html/plan/ll84_scores.shtml, which contains information on the total greenhouse gases (GHGs) for each building tracked under the guidelines of New York City's Local Law 84. I downloaded the data sets for 2014, 2015, and 2016 and then "stacked" them together to make one large data set. This data set also contains some basic building information, including its gross building area, its main use (such as "multifamily housing"), its address, its unique borough-block-lot (BBL) tax id number, the year the structure was built, and other information not used for the analysis in this blog post. For ease of discussion call this the GHG data set.

Using the BBL number for each property, I then merged information from the 2016 PLUTO file. The PLUTO data set contains more information about the buildings, including the year completed, the number of floors, the total building area, the total building area used for residential purposes, commercial, manufacturing, or garages, etc. See the data dictionary here.

The combined data set contains 33,606 observations, with nearly all buildings having recorded data for at least two years (i.e., a panel data set). The data set, however, appears to have some anomalies and outliers, which might have been recording mistakes. For this reason, the statistical analysis was performed after the following restrictions were imposed. Observations were **excluded** if:

1. The year built in the GHG data set was not the same as that in the PLUTO file.
2. If the gross floor area given in the GHG data set was 10% different than of that given in the PLUTO file.
3. If the total greenhouse gas emissions per 10,000 square feet of building floor area (greenhouse gas intensity) was less than 0.4 or greater than 200. This represented a removal of observations that were in the 1st and 99th percentile, respectively.
4. If the property had multiple buildings; this was done for clarity of analysis. If a property had multiple structures, it would be hard to determine which structure was responsible for the emissions.
5. If the property was not a multifamily dwelling. That is, only apartment buildings were included in analysis here.
6. If the floor count or lot area given was zero.

These exclusions reduced the data set to 13,903 observations over the three-year period. The biggest reduction from the initial data set to the final data set came from the removal of properties with multiple structures and from unequal year-built values across the two data sets; those two alone reduced the data set by roughly half.

Table 1 gives the descriptive statistics for the variables used in the analysis. These are total greenhouse gas emissions (metric tons per year), building area (square feet), building floor count, total lot area (square feet), total building area devoted to garage space (square feet), the year built, and the slenderness ratio.

The slenderness ratio is a measure of how tall the building is relative to its width. This variable has gained significant attention in the last decade because of the rise of the supertall, superslim luxury condos being built on or near 57th Street ("Billionaires Row"). The slenderness ratio is typically measured as the total height divided by the width. Because I don't have data on building height, I calculate the slenderness ratio as *(12xfloorcount)/(building width)*, both in feet, on the assumption that the average floor height is 12 feet. A superslim is very tall building with a slenderness ratio of at least 10. In one of the regressions, the slenderness ratio was included to see if it had an effect on emissions.

Variable	Mean	Std. Dev.	Min.	Max.	Nobs.
Greenhouse Gas Emissions (metric tons/year)	659.8	960.2	4.4	27410	13,903
GHGs per 1000 sq. ft. of Building Area	5.8	2.4	0.06	20.0	13,903
Floors	9.0	6.3	1	59	13,903
Building Front (feet)	141.3	150.7	0	7538	13,903
Building Area (per 10,000 sq. ft.)	11.3	12.0	3.8	276.9	13,903
Slenderness Ratio	1.01	1.03	0.015	16.4	13,621
Garage Area (per 10,000 sq. ft.)	0.22	1.65	0	61.4	13,903
Year Built	1942	24.0	1850	2015	13,903
Lot Area (per 10,000 sq. ft.)	2.50	3.28	0.28	82.9	13,903

Table 1: Descriptive Statistics for 2014-2016.

2. Basic Regression Results

Table 2 gives the basic regressions for the effect of building area and floor count on the natural log of total greenhouse gas emissions for 2016 only. Equations (1) only has the floor count. Equation (2) just has *ln(Building Area)*. Equation (3) is with both floor count and *ln(Building Area)*; Equation (4) is the same as (3) but with census tract fixed effects. Equation (5) is the same as (3) but with *ln(Floor Count*). *A priori*, it's not clear whether levels or logs should be included. The adjusted-R^2, AIC, and BIC actually provide little guidance in this matter. Finally equation (6) includes *ln(Floor Count)*, *ln(Building Area)*, and *ln(Floor Count) x ln(Building Area)*, an interaction term. Equation (6) also includes census tract fixed effects.

In short, we see in equation (5) a floor elasticity of 0.09 and a building area elasticity of 0.89. Equation (6) suggests that an interaction term is significant. While the interaction term is important it switches the sign of the floor count coefficient. In short, it suggests a diminishing marginal "return" to GHGs with respect to height.

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Floor	0.061		0.009	0.010		
	0.00		0.00	0.00		
ln(Floor)					0.09	-0.69
					0.00	0.03
ln(Building Area)		0.95	0.88	0.89	0.89	0.73
		0.00	0.00	0.00	0.00	0.00
ln(Floor) x ln(Building Area)						0.068
						0.01
Constant	5.6	-4.6	-4.0	-4.1	-4.2	-2.4
	0.00	0.00	0.00	0.00	0.00	0.00
Nobs	5027	5027	5027	5027	5027	5027
R^2	0.28	0.55	0.55	0.67	0.55	0.67
Adj. R^2	0.28	0.55	0.55	0.59	0.55	0.59
AIC	7746.4	5463.5	5426.1	**3788.8**	5433.2	3790.3
BIC	7759.5	5476.5	5445.7	**3801.9**	5452.8	3809.8
Census Tract Fixed Effects	No	No	No	Yes	No	Yes
P-value for Fixed Effects				0.00		0.00

Table 2: Dependent Variable: ln(Total Greenhouse Gas Emissions) for 2016 only. Note: All regressions produced robust standard errors that were clustered by the census tract. P-values are below coefficient estimates. Note that p-value less than 0.1 indicates statistical significance.

3. Expanded Regression Results

Table 3 expands upon Table 2 by including additional controls, which include the *ln(Lot Area)*, *ln(1+Garage Area)*, the decade in which the structure was built, and 2014 and 2015 year dummies. One specification (Equation (4)) includes the slenderness ratio. Garage area was included since presumably the more area devoted for garage space would reduce emissions. the year dummies control for time varying effects. Note that though the data set is an unbalanced panel I did not employ an panel methods, because building fixed effects or differencing would have removed the flour count variable from the regressions.

Lastly, all equations include two sets of additional dummy variables. First is census tract fixed effects to help control for income and demographic characteristics that might determine GHG independent of a building's form. Second is additional set of building type controls that make more fine distinction about the types of structure in the data set, such as whether the building is a condominium, coop, walk up or elevator rental building. Across all specifications, both sets of dummy variables are jointly significant at greater than 99%.

Also note Equations (2)-(4) include the floor count squared. The negative coefficient estimate also suggests a marginal reduction of GHGs with height. Equations (3) and (4) include building area in levels as an additional specification. Equation (5) includes *ln(Floor Count)*, *ln(Building Area)*, and the interaction of the two, showing similar results as in Table 1, Equation (6).

Variable	(1)	(2)	(3)	(4)	(5)
Floors	0.019	0.03	0.036	0.037	
	0.00	0.00	0.00	0.00	
$Floors^2$		-0.0002	-0.0004	-0.0004	
		0.06	0.01	0.01	
ln(Floors)					-0.317
					0.33
ln(Floor) x ln(Building Area)					0.050
					0.07
ln(Building area)	0.73	0.71			0.598
	0.00	0.00			0.00
Building Area (per 10,000 sq. ft.)			0.071	0.071	
			0.00	0.00	
Building $Area^2$ (per 10,000 sq. ft.)			-0.001	-0.001	
			0.00	0.00	
Slenderness Ratio				-0.004	
				0.67	
ln(Lot Area)	0.16	0.17	0.24	0.24	0.17
	0.00	0.00	0.00	0.00	0.00
ln(1+Garage Area)	-0.006	-0.006	-0.008	-0.008	-0.006
	0.06	0.09	0.03	0.03	0.07
Decade Built	-0.01	-0.01	-0.01	-0.01	-0.01
	0.03	0.02	0.02	0.02	0.02
2015 Dummy	-0.026	-0.026	-0.027	-0.027	-0.026
	0.02	0.02	0.02	0.02	0.02
2016 Dummy	-0.050	-0.050	-0.051	-0.051	-0.051
	0.00	0.00	0.00	0.00	0.00
Constant	-1.72	-1.46	5.25	5.26	-0.54
	0.08	0.14	0.00	0.00	0.66
Nobs.	11,213	11,213	11,213	11,213	11,213
R^2	0.61	0.61	0.60	0.60	0.61
Adj. R^2	0.57	0.57	0.56	0.56	0.57
AIC	10665.7	10658.1	10816.3	10818	**10653.9**
BIC	10782.9	**10782.6**	10948.1	10957.2	10785.8
P-value for Building Type Fixed Effects	0.00	0.00	0.00	0.00	0.00
P-value for Census Tract Fixed Effects	0.00	0.00	0.00	0.00	0.00

Table 3: Dependent Variable: ln(Total Greenhouse Gas Emissions). Note: All regressions include robust standard errors that were clustered by the census tract. P-values are below coefficient estimates. Note that p-value less than 0.1 indicates statistical significance.

Looking at Equation (1), for example, we see that the inclusion of the off the additional controls increases the floor count coefficient from 0.009, in Table 1 Equation (3), to 0.019. The additional controls reduce the *ln(Building Area)* coefficient from 0.88 to 0.71.

A few other thing to note. The results suggest about 2.5% annual reduction in GHGs from these buildings. Further, on average, new buildings have less GHGs than older buildings. Slenderness, per se, does not seem to matter. Finally, garage area reduces GHGs, while increased lot size increase GHGs, all else equal (why this may be is left for future investigations).

4. Predicted Effects of Height and Area

Using coefficient estimates from Table 3, Equation (5), the aim was to give estimates to see how increasing height and floor area impacted the growth in greenhouse gas emissions. To this end, the following exercise was performed. First, I assumed a fixed building size, then I created the "floor effects" from the following equation (given the fixe building area):

$$\ln(Floor\ Effect) = -0.317 \ln(Floor\ Count) + 0.05 \ln(Floor\ Count) \times \ln(Bldg.Area).$$

One caveat is in order. In Table 3, Equation (6), the AIC is minimized suggesting this the best functional form. However, the coefficients for the floor count is not statistically significant. Nonetheless the coefficient estimate appear plausible and so are used for this analysis.

Figure 1 shows the Floor Effect (i.e., *exp(ln(Floor Effect)*) for building areas of different sizes (100,000 sq. ft., 200,000 sq. ft., 300,000 sq. ft., 400,000 sq. ft. 500,000 sq. ft). The graph shows that for each building area, the height has a diminishing marginal effect on GHG emissions. It also shows that GHG emissions rise with building area.

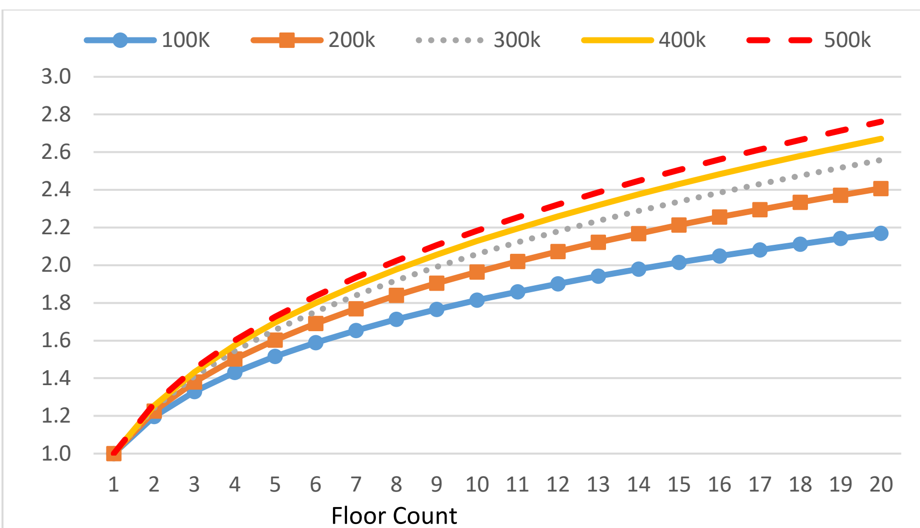

Figure 1: Estimated Relative Floor Effects for Buildings of Different Floor Areas (sq. ft.) Using Coefficient Estimates from Table 3, Equation (5). The graph shows two things. First, for a given building height, GHG emissions increase with building height but at a declining rate. Second, larger buildings emit more greenhouse gases, but, again, the increases occur at a decreasing rate (the space between lines diminishes going from the smaller to the large building).

Figure 2 shows the Floor Effects divided by the estimated building population. The assumption is that each occupant uses 400 square feet of space (about the average for New York City). So occupancy is determined by **# Occupants = Building Area/400**.

Then I take the relative floor effects and divide them by building population to get a measure of the relative per capita effect. The shape of the floor effect is preserved. However, we see that the 100,000 square foot building has the greatest per capita GHG emissions; while the largest building has the smallest effect.

The graph shows that as buildings get larger, it mitigates the floor effect, so that larger buildings have a lower net impact.

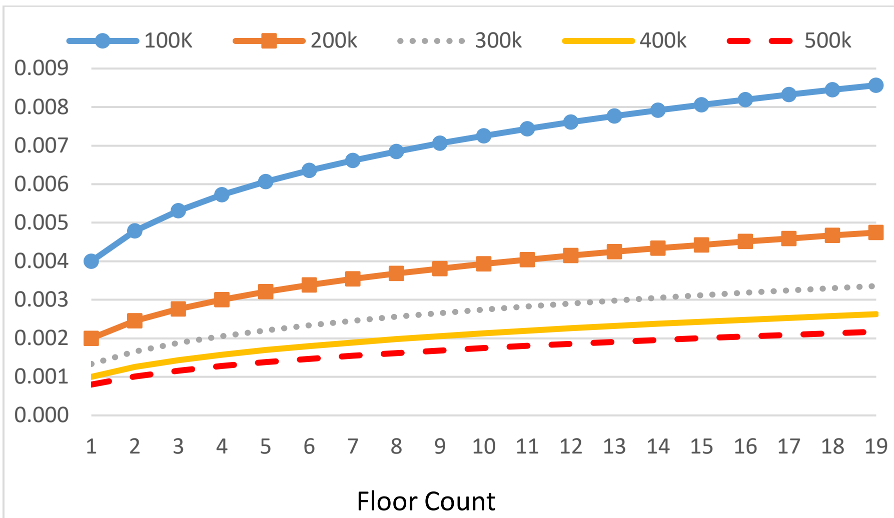

Figure 2: Estimated per capita floor effects. The relative floor effects take from Figure 3 divided by number of building occupants, where each occupant is assumed to consume 400 feet of building space.