

Regression Results for the Effect of Vacancy on NYC Real Estate Prices

Jason M. Barr

January 13, 2020

I. Data Sources.

- *Sale Prices* of individual properties along with sales date, building type, building and lot area is from the [NYC Department of Finance](#), for 2016 to 2017. Data included sales that were \$1000 or greater.
- *Real Sales Prices*: For the regression I used the natural log of real sales prices per square foot of building area, where sales price divided by the NYC-Metro Area CPI, all costs excluding shelter.
- *Vacancy Data*: I used the data from the American Community Survey for 2016 and 2017, respectively, at the 2010 Census Tract Level. which includes total number of units, the total vacant and occupied units, which is from file ACS_17_5YR_B25004_with_ann_ and ACS_16_5YR_B25004_with_ann. It also includes data on the nature of the vacancy, such as if the vacant unit has been sold or rented but not occupied, is currently on the market, is vacant now but is used for seasonal, recreation, or migrant work, and an other category. Note that the vacancy data is five-year averages, which is based on current year and four prior years.
- *Census Tract Income and Population counts*: I used median household income from the ACS for 2016 and 2017, file ACS_17_5YR_S1901_with_ann & ACS_16_5YR_S1901_with_ann. I used ACS_17_5YR_B25008_with_ann and ACS_16_5YR_B25008_with_ann for population counts.
- *Building level data* comes from the NYC Primary Land Use Tax Lot Output (PLUTO) file for 2015, which gives information about building characteristics, such as number of floors, number of residential units, building area, building area devoted to residential, number of structures on the property, and other location-based information, such as its latitude and longitude coordinates.

II. Data Processing

- The sales, vacancy data, and building data sets were merged. The regression results included sales that were above the 5th percentile and below the 95th percentile in terms of real prices per square foot. This was done to exclude outliers.
- When I merged the sales file with the PLUTO file, I only kept those data points where the building use and area were the same in the sales and PLUTO Files. In some cases, the building use changes, or the building is torn down after sale. These data points were excluded.
- For vacancy, in some specifications, I include the number of vacant units, while in others I include the percent of vacant units. Another variable I used is the weighted sum of number of vacant units outside the census tract, where the weights were calculated in the following manner. For each census tract, I determined the distance to the centroid of each other census tract. I then created weights where $w_{ij} = 1/d_{ij}$, where d_{ij} is distances, as

the crow files from census tract i to j . I then take the sum of the weights $W_i = \sum_j w_{ij}$ and normalize each weight by dividing it by the sum, so the sum of the normalized weights equals one. I then calculated the weighted distance-based # of vacant units as $v_{-i} = \sum_j v_j \left(\frac{w_{ij}}{W_i}\right)$. I also take the weighted sum of number of structures not in census tract i .

III. Descriptive Statistics

Table 1 provide descriptive statistics. Note that vacancy measures, income, and population are at the census tract level, but the statistics are given at the property level.

Variable	Mean	Std. Dev.	Min	Max	# Obs.
Real Price Per Square Foot	353.13	139.52	30.00444	693.3506	59,116
# Vacant Units, CT	112.27	81.28	0	1,303	37,475
% Vacancy, CT	7.84	4.32	0	36	37,475
# Units, CT	1,455	684	14	6,147	37,475
# Units for Rent, CT	24.07	31.98	0	364	37,475
# Unit Rented by Not Occ., CT	4.29	10.93	0	118	37,475
# Units for Sale, CT	13.24	20.87	0	158	37,475
# Units Sold but not Occ., CT	8.25	17.31	0	176	37,475
# Vacant Units for Seasonal or Rec., CT	12.94	23.62	0	942	37,475
# Vacant Units for Migrant Workers, CT	0.23	2.28	0	55	37,475
# Vacant Units for Other Reasons, CT	49.25	47.45	0	391	37,475
Population, CT	4,042	1,844	3	11,940	59,116
Med. HH Income (\$), CT	67,372	21,870	9053	165,753	59,106
Building Lot Area (ft ²)	3,143	4,110	200	679,000	59,116
Building Floor Area (ft ²)	2,584	11,473	324	1,741,458	59,116
# Residential Units per Building	2.32	10.52	0	1,327	59,116

Table 1: Desc. Stats. Note other variables included in regressions, but not given above. CT=2010 Census Tract. Data for 2016 and 2017. Vacancy data is five-year average.

IV. Regression Results

Table 2 provides the main specifications. Equation (1) has number of vacant units, number of units, weighted number of vacant units in all the other census tracts, weighted number of units in the other census tracts, and zip code fixed effects. Equation (2) adds building controls (lot size, building area, lot type, building types, age, height, etc.—note full results available upon request), and year-quarterly dummies. Equation (3) adds census tract income and population. Equation (4) adds census tract controls of total building area and number of residential units. Equation (5) is same as (4), but only for census tracts with median income in the bottom 25th percentile. Looking at Equation (4), the results suggest that an increase in within census tract vacant units by 10%, cet. par., reduces prices by 0.316%. An increase in the weighted average vacancy of the other census tracts by 10% reduces prices by 6.81%

	(1)	(2)	(3)	(4)	(5)
	Ln(rppsf)	Ln(rppsf)	Ln(rppsf)	Ln(rppsf)	Ln(rppsf)
ln(# Vacant Units, CT)	-0.0284***	-0.0256***	-0.0316***	-0.0324***	-0.0179
	(3.40)	(3.40)	(4.55)	(4.70)	(1.26)
ln(# Units, CT)	-0.00245	0.0247*	0.166***	-0.0646	0.0702
	-0.2	(2.44)	(4.56)	(0.68)	(0.46)
ln(# Weighted Vacant Units, -i CT)	-0.944**	-0.759**	-0.653**	-0.681**	-0.522
	(3.23)	(3.12)	(2.84)	(2.99)	(1.48)
ln(# Weighted Units -i ,CT)	1.642***	2.073***	1.914***	1.879***	1.320**
	(3.68)	(4.78)	(4.63)	(4.54)	(2.79)
ln(Med. HH Income, CT)			0.135***	0.134***	0.0604
			(5.88)	(5.97)	(1.95)
ln(Population, CT)			-0.131***	-0.127***	-0.0724
			(3.83)	(3.78)	(1.25)
ln(Total Building Area, CT)			0.0249*	-0.0442	
				(2.02)	(1.86)
ln(CT # Residential Units, CT)				0.207*	0.0546
				(2.42)	(0.42)
Constant	-1.607	-6.520*	-7.070**	-6.787**	-0.286
	(0.63)	(2.54)	(2.90)	(2.76)	(0.10)
N	40,585	37,014	37,014	37,014	8,170
R-sq	0.293	0.393	0.397	0.398	0.421
adj. R-sq	0.29	0.39	0.394	0.395	0.411
Year-Quarterly Dummies	No	Yes	Yes	Yes	Yes
Zip Code Fixed Effects	Yes	Yes	Yes	Yes	Yes
Building Controls	No	Yes	Yes	Yes	Yes

Table 2: Regression Results. Notes: Dependent variable is ln(Real Price Per Square Foot). Standard errors clustered by census tracts. T-stats below estimates. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. CT=2010 Census Tract

Table 3 provides additional specifications. Equation (1) also includes breakdowns for the reason for vacant units, which each variable is $\ln(I + count)$. This means that if there no observations for that variable it takes on the value of 0. I also include total number of vacant units, so the categories can be interpreted as additional impact above or below the average vacancy effect. In general, the vacancy effect remains similar to before, suggesting that empty units reduce prices in general. Equations (2) and (3) include the percent vacancy for each CT. Equation (3) is just for CTs in the lowest quartile.

	(1)	(2)	(3)
	Ln(rppsf)	Ln(rppsf)	Ln(rppsf)
CT Vacancy (%)		-0.00605***	-0.00195
		(5.27)	(1.08)
Weighted Vacancy (%)		-0.149***	-0.128*
		(3.89)	(2.19)
ln(# Vacant Units, CT)	-0.0298***		

	(4.02)		
ln(1+# For Sale or Rent, CT)	-0.00237		
	(0.91)		
ln(1+# Sold/Rented, not Occ., CT)	0.00301		
	(1.22)		
ln(1+# Seasonal, not Occ., CT)	-0.00307		
	(1.20)		
ln(1+# for Migrant Work, Not Occ., CT)	-0.0113		
	(1.00)		
ln(1+# Not Occ. For Other Reasons, CT)	-0.000219		
	(0.07)		
ln(# Units, CT)	-0.0628	0.150***	0.0778
	(0.67)	(4.31)	(1.44)
ln(# vacant weighted units, -i CT)	-0.668**		
	(2.94)		
ln(# units weighted, -i CT)	1.873***	1.411***	0.943**
	(4.56)	(4.57)	(2.63)
ln(Med. HH Income, CT)	0.136***	0.137***	0.0555
	(6.07)	(6.12)	(1.81)
ln(Population, CT)	-0.128***	-0.149***	-0.0746
	(3.80)	(4.42)	(1.28)
ln(PLUTO Total Building Area, CT)	0.0265*		
	(2.29)		
ln(PLUTO # Residential Units, CT)	0.205*		
	(2.39)		
Constant	-2.78	-4.976*	0.803
	(2.77)	(2.31)	(0.29)
N	37,014	37,462	8,178
R-sq	0.398	0.398	0.421
adj. R-sq	0.395	0.395	0.411
Year-Quarterly Dummies	Yes	Yes	Yes
Zip code Fixed Effects	Yes	Yes	Yes
Building Controls	Yes	Yes	Yes

Table 3: Regression Results. Notes: Dependent variable is $\ln(\text{Real Price Per Square Foot})$. Standard errors clustered by census tracts. T-stats below estimates. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. CT=2010 Census Tract.

V. Additional Comments

In all cases, I used ordinary least squares with many controls, including building-level variables, income and population variables, as well as zip-code (neighborhood) dummies. These OLS results provide strong evidence that the number of vacant units in and around census tracts significantly reduces prices. Adding additional controls because units are vacant does not appear change the results. Having said that I cannot rule possible endogeneity with respect to vacancy and prices in that vacancy might in some sense be determined by prices. I experimented with various instruments, such as population and building area as of

2010, but all the sets of instruments always rejected the null hypothesis with Overidentification Tests, suggesting invalid instruments.