Data and Results for Skyline vs. the Spawl-line

Jason M. Barr
September 15, 2020

The Data Sets

Household Carbon Emissions

The dependent variable used here is average estimated household greenhouse gas emissions by zip code. The data set is from Jones and Kammen (2013). I summarize their process here, but details can be found in the paper and their supplementary paper (at https://pubs.acs.org/doi/10.1021/es4034364). Their process is to create a carbon footprint estimate using the following formula.

$$HCF_z = \sum_{i=1}^{n} C_{iz} E_i,$$

where for each zip code (z) they create a predicted value of household consumption ($\hat{C}_{iz}$) for each of *n* sectors/categories from *n* separate regressions. Then then multiply the estimated average household consumption by E_i, which is an emissions conversion factor that converts consumption into a carbon dioxide emissions value.

Their sectors/categories include:

1. *Household Electricity Use*. Here they regress electricity consumption in kWh/year on several controls, such as housing characteristics, demographic characteristics, price of electricity, local temperatures, and income to get predicted electricity use.
2. *Natural Gas Consumption*: Dep. Var.: natural gas consumption in BTU/year. Control variables: housing characteristics, price of natural gas, temperature data, region and demographic data.
3. *Fuel Oil*. Regress fuel consumption on housing and demographic characteristics.
4. *Vehicle miles travelled (VMT):* Regress MVT on commuting and household characteristics.
5. *Food consumption; other goods and service consumption; and consumption of water, waste and building construction, respectively*: regressions of various household consumption items on household size and income.

In short, their HCF value for each zip code is a non-linear sum of many household, housing, locational, and climate-related variables, since each sector is a predicted value from two or more righthand side variables. Also note that these data are from about 2007.

Buildings Data

The New York City Department of City Planning provides a data set on every tax lot in New York City, which contains information about building areas, lot areas, number of floors, age, and so on (NYC PLUTO). Variables were created here at the zip code level for residential buildings (i.e., had at least one residential unit with non-zero floor area used for residential purposes.) The key independent variables are: average age of residential structures in the zip code, average number of floors per residential building, total zip code floor area for residential use and total lot area for these buildings. The residential Floor Area Ratio (FAR) was calculated for each zip code by dividing the total residential floor area by the total residential lot area.

Descriptive Statistics for the data are given in Table 1. Note here that I used building data as of middle of 2012. This means that the righthand side variables are a few years after the GHG variable. However, the average and totals for housing in a zip code does change very much in a five year span and additional there is not much reason to believe that in NYC, GHG emission in 2012 were much different from 2007, at least relatively across zip codes (in the future, I plan to check this with using 2007 building data). The Jones and Kammen (2013) and the NYC PLUTO (2012) data were merged by zip code.

Variable	Mean	Std. Dev.	Min	Max	Obs
Avg. HH CO2 Emissions (metric tons)	36.9	6.0	26.2	64.0	178
Floor Area Ratio	2.5	2.6	0.42	17.7	180
Avg Number of Housing Units	39	174	2	1844	180
Avg. # Floors	4.3	4.7	1	34	181
Total Residential Floor Area (sq. ft)	16,200,000	10,300,000	850	47,700,000	181
Total Res. Lot Area (sq. ft.)	9,699,320	6,402,292	152,374	35,700,000	180
Avg. Income per H.H.	43,681	19,264	2,499	112,947	178
Population	44,963	26,845	7	106,154	178
Avg. Residential Year Built	1,937	21	1,899	2,004	181
Avg Vehicle Miles Travelled per Year	10,388	4,314	3,271	28,270	178
Avg Electricity Usage (kWh)	5,949	1,370	3,628	10,319	178
Avg. Fuel Oil (Gallons)	220	91	3	679	178
Avg. Natural Gas (Cu. ft.)	41,800	12,262	8,043	73,118	178
Bronx	0.140				178
Brooklyn	0.208				178
Queens	0.337				178
Staten Island	0.067				178

Table 1: Descriptive Statistics Data. The data is at the zip code level. Note CO_2 emissions, vehicle miles, electricity, fuel, natural gas, and income are from Jones and Kammen (2013). The rest are from NYC PLUTO (2012).

Regression Results

Below I provide several regression tables with different specifications. Every regression has the same dependent variable: Avg. Household CO2 production by zip code. I have put in bold those variables are constancy statistically significant within the same regression (and have the same sign across specifications). In generally I used log-log specifications to estimate elasticities. In each table equation (1) has only the key independent variables (such as floor area or floor area per person, and average floor height). Then each one adds or substitutes additional variables to see how robust the building form variables are.

Table 2

	(1)	(2)	(3)	(4)	(5)
	lnCO2	lnCO2	lnCO2	lnCO2	lnCO2
lnFloor area per cap.	-0.0556*	0.0134			
	(-2.20)	(0.70)			
lnHHIncome	**0.168*** **	**0.147*** **	**0.152*** **	**0.152*** **	**0.137*** **
	(11.89)	**(10.50)**	**(9.50)**	**(9.02)**	**(11.47)**
lnVehicleMiles	**0.235*** **	**0.0779**	**0.0674**	**0.0750***	**0.114***
	(9.99)	**(1.60)**	**(1.69)**	**(2.25)**	**(2.16)**
Bronx	0.00515	0.0117	0.0181**	0.0252**	0.0227*
	(0.34)	(1.24)	(4.23)	(3.12)	(2.50)
Brooklyn	-0.00226	-0.00438	0.00558	0.0120	0.00413
	(-0.16)	(-0.35)	(0.71)	(0.93)	(0.34)
Queens	-0.0110	-0.00492	-0.00144	0.00544	0.00314
	(-0.44)	(-0.23)	(-0.08)	(0.30)	(0.16)
Staten Island	-0.00225	-0.00959	-0.00227	0.00606	0.000397
	(-0.07)	(-0.36)	(-0.11)	(0.28)	(0.02)
lnAvgNumUnits		-0.00233			-0.0226
		(-0.57)			(-1.52)
lnElectrcity		**0.606*** **	**0.582*** **	**0.566*** **	**0.521*** **
		(5.82)	**(5.10)**	**(5.56)**	**(5.23)**
lnFuel		**0.0557** **	**0.0514** **	**0.0548** **	**0.0550** **
		(3.87)	**(3.60)**	**(3.26)**	**(3.77)**
lnNatGas		**-0.108***	**-0.106**	**-0.0874**	**-0.0951** **
		(-2.52)	**(-1.98)**	**(-1.84)**	**(-2.91)**
lnTotalFloor area			-0.00702*	-0.0112*	-0.00123
			(-2.57)	(-2.72)	(-0.19)
lnPop			0.0000694	0.00663	0.00215
			(0.01)	(1.17)	(0.24)
lnAvgFloors				0.0177	0.0614
				(1.09)	(1.77)
_cons	-0.00772	-3.142***	-2.716***	-2.881***	-2.735***
	(-0.02)	(-5.14)	(-7.81)	(-6.99)	(-9.70)
N	174	174	175	175	174
R-sq	0.918	0.967	0.966	0.967	0.970
adj. R-sq	0.915	0.965	0.964	0.964	0.968
AIC	-598.2	-752.8	-754.1	-757.1	-770.6
BIC	-588.7	-740.2	-741.5	-744.5	-758.0

t statistics in parentheses * p<0.10, ** p<0.05, *** p<0.01. Standard errors clustered by borough.

Equations (1) and (2) above look at ln(Floor area per capita). The rest split up floor area and population as two independent variables. Across specifications, floor area is not statistically significant and does not seem a strong determannt of emissions, after controlling for other factors.

Table 3

	(1) lnCO2	(2) lnCO2	(3) lnCO2	(4) lnCO2	(5) lnCO2
lnFloorArea	-0.156***	-0.107	-0.0112	-0.00570	0.0409**
	(-6.66)	(-1.35)	(-0.25)	(-0.18)	(2.89)
lnTotalLotarea	0.0880**	0.0505	-0.0607	-0.0585**	-0.0617**
	(3.02)	(0.59)	(-1.98)	(-2.84)	(-3.67)
Bronx		-0.0127	0.0147	0.0206	0.0327***
		(-0.14)	(0.88)	(2.04)	(9.20)
Brooklyn		0.0128	0.0104	-0.00365	0.0202***
		(0.13)	(0.61)	(-0.41)	(6.73)
Queens		0.0811	0.0107	0.00554	0.0270***
		(0.57)	(0.34)	(0.26)	(6.54)
Staten Island		0.167	0.0306	0.00555	0.0307***
		(0.90)	(0.78)	(0.21)	(5.17)
lnAvgFloors			-0.0208	0.0926***	0.0389***
			(-0.95)	(5.34)	(6.72)
lnHHIncom			**0.147*** **	**0.125*** **	**0.123*** **
			(10.36)	**(10.42)**	**(14.08)**
lnPop			**0.0626** **	**0.0628*** **	**0.0188** **
			(4.28)	**(5.14)**	**(4.47)**
lnAvgYearBuilt			-1.038**	-0.0995	-0.124
			(-4.17)	(-0.42)	(-0.64)
lnVehicleMiles			**0.317*** **	**0.320*** **	**0.176*** **
			(11.74)	**(16.45)**	**(7.10)**
lnAvgNumUnits				**-0.0553** **	**-0.0160** **
				(-4.27)	**(-3.77)**
lnElectrcity					**0.528*** **
					(9.45)
lnFuel					**0.0536*** **
					(7.02)
lnNatGas					**-0.0760** **
					(-3.48)
_cons	4.757***	4.497***	7.491**	0.461	-2.357
	(37.76)	(17.43)	(3.81)	(0.27)	(-1.60)
N	174	174	174	174	174
R-sq	0.512	0.571	0.934	0.957	0.981
adj. R-sq	0.507	0.555	0.930	0.953	0.980
AIC	-287.1	-311.4	-633.2	-706.6	-853.6
BIC	-277.7	-305.1	-620.6	-693.9	-841.0

t statistics in parentheses, * p<0.10, ** p<0.05, *** p<0.01. Standard Errors clustered by borough.

The above table provides additional specifications, an again, floor area is not stat. sig. across equations.

Table 4

	(1) lnCO2	(2) lnCO2	(3) lnCO2	(4) lnCO2	(5) lnCO2
lnAvgFloors	-0.557***	0.104*		0.0542	
	(-8.20)	(2.54)		(1.47)	
$lnAvgFloors^2$	0.131***				
	(10.57)				
Bronx	-0.0822	0.0169	-0.000795	0.0266*	0.0137
	(-2.00)	(0.93)	(-0.06)	(2.29)	(1.69)
Brooklyn	-0.0977	-0.0190	-0.0322	0.00549	-0.00400
	(-2.13)	(-1.28)	(-2.10)	(0.45)	(-0.37)
Queens	0.0129	-0.0135	-0.0294	0.00609	-0.00238
	(0.25)	(-0.48)	(-1.13)	(0.30)	(-0.15)
Staten Island	0.104	-0.0182	-0.0304	0.00526	-0.00176
	(1.83)	(-0.48)	(-0.82)	(0.22)	(-0.10)
lnFloorArea		-0.0402	-0.0341	0.00211	0.0133
		(-1.66)	(-1.21)	(0.46)	(0.84)
lnAvgNumUnits		-0.0564**	-0.0433***	-0.0206	-0.00770
		(-3.17)	(-5.20)	(-1.37)	(-1.16)
lnHHIncome		**0.142*** **	**0.151*** **	**0.136*** **	**0.140*** **
		(9.87)	**(8.43)**	**(10.88)**	**(9.44)**
lnPop		0.0458*	0.0376	0.00234	-0.00535
		(2.52)	(1.56)	(0.29)	(-0.30)
lnzipArea		-0.00715	-0.0104	-0.00681*	-0.0108***
		(-1.24)	(-1.52)	(-2.34)	(-4.72)
lnAvgYearBuilt		-0.245	-0.297	-0.292	-0.422
		(-0.55)	(-0.60)	(-0.76)	(-1.02)
lnVehicleMiles		**0.264*** **	**0.254*** **	**0.120***	**0.0902** **
		(9.06)	**(10.22)**	**(2.54)**	**(2.97)**
Avg 3-4 floors dummy			0.0319		0.0111
			(1.56)		(0.90)
Avg 5-6 floors dummy			0.0427		0.00230
			(2.11)		(0.15)
Avg 7-9 floors dummy			0.0635**		-0.00145
			(4.29)		(-0.15)
Avg 10-15 floors dummy			0.105**		0.0879***
			(4.07)		(7.02)
Avg 16-20 floors dummy			0.0752**		0.0275
			(3.71)		(0.79)
Avg >20 floors dummy			0.164***		0.0368*
			(5.43)		(2.19)
lnElectrcity				**0.541*** **	**0.651*** **
				(5.42)	**(6.94)**
lnFuel				**0.0519** **	**0.0557***
				(3.28)	**(2.47)**
lnNatGas				**-0.104** **	**-0.138** **
				(-2.94)	**(-3.43)**

_cons	4.049***	1.717	2.173	-0.682	-0.138
	(40.23)	(0.57)	(0.60)	(-0.23)	(-0.04)
N	175	174	174	174	174
R-sq	0.594	0.947	0.944	0.970	0.973
adj. R-sq	0.579	0.943	0.938	0.968	0.969
AIC	-322.6	-670.2	-663.2	-773.1	-786.2
BIC	-316.3	-657.5	-650.6	-760.5	-773.6

t statistics in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Standard errors are clustered by borough.

Table 3 investigates average number of floors across equations. Equations (3) and (5) break down the average floor heights into different categories. The result again show that while floor height could be positive across specifications, the results are not robust across specifications.

Table 5

	(1) lnCO2	(2) lnCO2	(3) lnCO2	(4) lnCO2	(5) lnCO2
lnAvgFloors	-0.557***	0.0896		0.0556	
	(-8.20)	(1.98)		(1.41)	
lnAvgFloorsq	0.131***				
	(10.57)				
Bronx	-0.0822	0.00635	-0.00422	0.0267*	0.0130
	(-2.00)	(0.29)	(-0.30)	(2.28)	(1.45)
Brooklyn	-0.0977	-0.0361	-0.0408*	0.00595	-0.00283
	(-2.13)	(-1.97)	(-2.42)	(0.48)	(-0.23)
Queens	0.0129	-0.0175	-0.0268	0.00591	-0.00515
	(0.25)	(-0.52)	(-0.99)	(0.29)	(-0.29)
Staten_Isl~d	0.104	-0.0271	-0.0335	0.00541	-0.00233
	(1.83)	(-0.63)	(-0.86)	(0.23)	(-0.12)
lnAvgNumUn~s		-0.0587**	-0.0478***	-0.0211	-0.00925
		(-4.27)	(-10.19)	(-1.32)	(-1.12)
lnHHIncom		**0.131*** **	**0.139*** **	**0.137*** **	**0.143*** **
		(7.75)	**(5.79)**	**(10.90)**	**(8.61)**
lnPop		0.0159*	0.0128*	0.00401	0.00468
		(2.47)	(2.42)	(0.88)	(0.68)
lnzipArea		-0.0121*	-0.0146**	-0.00658	-0.00935***
		(-2.45)	(-3.80)	(-2.07)	(-5.48)
lnAvgYearB~t		-0.428**	-0.375	-0.274	-0.372
		(-2.87)	(-0.92)	(-0.72)	(-0.98)
lnVehicleM~s		**0.300*** **	**0.289*** **	**0.120***	**0.0895*** **
		(23.43)	**(13.63)**	**(2.53)**	**(2.82)**
floorcat3_4			0.0326		0.0124
			(1.66)		(0.90)
floorcat5_6			0.0482**		0.00381
			(2.93)		(0.22)
floorcat7_~9			0.0656***		0.00260
			(4.75)		(0.20)
floorcat1~15			0.107***		0.0904***
			(4.96)		(6.97)
floorcat1~20			0.0724**		0.0341
			(3.94)		(0.81)
floorcatgt20			0.149**		0.0515
			(4.46)		(1.71)
lnElectrcity				**0.535*** **	**0.616*** **
				(4.90)	**(5.81)**
lnFuel				**0.0516*** **	**0.0532***
				(3.35)	**(2.61)**
lnNatGas				**-0.103*** **	**-0.135*** **
				(-2.81)	**(-3.17)**
_cons	4.049***	2.574*	2.278	-0.765	-0.139
	(40.23)	(2.30)	(0.71)	(-0.26)	(-0.05)

N	175	174	174	174	174
R-sq	0.594	0.941	0.941	0.970	0.972
adj. R-sq	0.579	0.937	0.935	0.968	0.969
AIC	-322.6	-653.6	-651.6	-773.1	-783.4
BIC	-316.3	-641.0	-639.0	-760.4	-770.8

t statistics in parentheses. * p<0.10, ** p<0.05, *** p<0.01. Standard errors clustered by borough.

Table 6: Dep. Var. Avg HH CO_2 in levels.

	(1)	(2)	(3)	(4)
lnFloorArea	-4.439	-0.383	-0.806	1.159*
	(-1.57)	(-0.24)	(-0.65)	(2.75)
lnTotalLot~a	2.193	-2.297*	-2.025**	-2.112***
	(0.74)	(-2.39)	(-3.22)	(-6.48)
lnAvgNumUn~s		-1.164*	-2.000**	-0.364
		(-2.51)	(-4.34)	(-2.04)
lnHHIncom		4.937***	4.394***	4.533***
		(7.88)	(7.21)	(6.42)
lnPop		2.037**	2.543**	0.635*
		(3.23)	(3.41)	(2.21)
lnAvgYearB~t		12.95	9.373	-7.354
		(0.66)	(0.60)	(-0.63)
lnVehicleM~s		11.00***	11.69***	4.340**
		(14.12)	(16.86)	(3.39)
lnAvgFloors			3.235***	0.773
			(5.30)	(1.68)
lnDistESB				0.520
				(1.61)
lnElectrcity				23.74***
				(7.34)
lnFuel				1.693**
				(4.15)
lnNatGas				-4.404*
				(-2.67)
_cons	74.67***	-190.3	-168.6	-154.4
	(7.14)	(-1.22)	(-1.38)	(-1.79)
N	174	174	174	174
R-sq	0.595	0.945	0.952	0.982
adj. R-sq	0.581	0.941	0.948	0.980
AIC	941.1	598.9	575.8	406.2
BIC	947.4	611.6	588.4	418.8

t statistics in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Takeaways from these regressions:

- Floor height seems to be positively related to CO_2 but is not robust across specifications.
- Avg # of units per building is negatively related to CO_2—not stat. sig. across all specifications, but is negative in all of them and about the same size.
- Floor area is not robust indicator of $C0_2$ emissions.
- Strong predictors: Income, energy usage and vehicle Miles (which are used indirectly to construct the $C0_2$ measure from Jones and Kammen).
- Conclusion: regressions do not support the idea that residential building shape matters all that much after controlling for income, energy, height and cooling, and driving frequency.

- Caveats: The Jones and Kammen data are somewhat indirect in that they use consumer behavior and correlates with behavior to get predicted values. It may be that their method of generating average household CO_2 emissions is the reason for no effect from building form, rather than reflecting some true lack of relationship. However, without further data sets on CO_2 emissions at the local level, we cannot say for sure.