Data Sources and Calculations for Blog Post on Transit-Oriented Design

Jason Barr
Rutgers University-Newark
January 17, 2023

Data Sources

PLUTO File and Map Pluto: The NYC Dept of City Planning produces these files (CSV and shapefile format) which contain data on every lot in New York. Data includes details about height, size, use, and zoning regulations. Source: https://www.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page. I used 2022 version 1.

Subway stops shapefile: https://data.cityofnewyork.us/Transportation/Subway-Stations/arq3-7z49

Census Tract Data: Census Tract Population Data and were obtained from the American Community Survey for 2020 at https://data.census.gov. The Shapefile of the 2020 Census Tracts are here: https://www.nyc.gov/site/planning/data-maps/open-data/census-download-metadata.page.

Subway Ridership by Stop: Ridership by subway stop is from https://new.mta.info/document/91476 at https://new.mta.info/agency/new-york-city-transit/subway-bus-ridership-2021

The 1KM Subway Zone

In QGIS, I created a 1-kilometer (1KM) "buffer" around each subway stop in NYC, including the stops on Staten Island. I then obtained that part of the PLUTO file (in Map PLUTO) that was in the buffer.

Using this new version of the PLUTO file, I was able to obtain data about every building and the current zoning rules for each lot within a half mile of a subway station. For each lot, data included the total number of residential units, the floor heights, maximum allowable floor area ratios, and total floor area for all the buildings. Note I also eliminated properties in this file that are classified by the city as open spaces or for transportation or utilities (see Pluto Data diction for land use categories: https://s-media.nyc.gov/agencies/dcp/assets/files/pdf/data-tools/bytes/PLUTODD.pdf?v=22v3)

For example, there are 452,934 lots in the 1KM zone (outside of Manhattan), and the average building has 4.476 residential units, making a total of just a little over 2 million housing units.

```
. sum bldgarea builtfar resarea unitsres residfar if  borough!="MN"

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
    bldgarea |    452,934    5907.306    37003.29          0   1.35e+07
    builtfar |    452,934    1.171041    1.224521          0        437
     resarea |    452,934    4186.743    30153.07          0   1.32e+07
    unitsres |    452,934     4.47635    44.12291          0      19201
    residfar |    452,934    1.507363    1.074547          0         10
     lotarea |    452,934    4635.907    53432.24          0   2.80e+07
-------------+---------------------------------------------------------
```

The average built FAR is only 1.17. Note that the average lot size is 4,636 ft^2, and there are 452,934 lots. This gives a total developable land area of 2.1 billion ft^2 or 75 square miles of land. And if there was an average FAR of 4 on these properties that would 8.4 billion square feet of residential space (for 8.4 million units, at 1000 ft^2 (93 m^2) per unit; see below).

Here is a table of the number of floors that are 3 or less in the 1km zone outside of MN (Note that are some building with non-integer floor levels, not shown, but they are a small fraction of the total; rest of table above 3 stories now shown).

NumFloors	Freq.	Percent	Cum.
0	19,290	4.26	4.26
1	31,085	6.86	11.13
2	226,763	50.07	62.56
3	89,297	19.72	91.56

Here is a table of the maximum allowable residential FAR in the 1KM zone (outside of MN). Note that 63% have a max FAR of 1.35 or less.

```
tab residfar ifborough!="MN”.
```

ResidFAR	Freq.	Percent	Cum.
0	16,419	3.63	3.63
.5	74,694	16.49	20.12
.75	84,978	18.76	38.88
.85	827	0.18	39.06
.9	5,135	1.13	40.19
1.1	4,105	0.91	41.10
1.25	73,643	16.26	57.36
1.35	25,105	5.54	62.90
2	53,677	11.85	74.75
2.43	66,091	14.59	89.35
3	19,093	4.22	93.56
3.44	14,198	3.13	96.70
4	9,007	1.99	98.68
4.2	783	0.17	98.86
5	938	0.21	99.06
6.02	3,518	0.78	99.84
7.52	242	0.05	99.89
9	20	0.00	99.90
10	461	0.10	100.00
Total	452,934	100.00	

I also calculated the values of current built FAR minus allowable FAR in percentage terms = 100*(builtFAR-allowable Resid FAR)/allowable Resid FAR (for buildings that are currently residential). Here is the fraction that are either over the FAR or under by 25% or less. 56% are either over or within 25%.

FARDiffPost _wi25	Freq.	Percent	Cum.
0	220,794	44.00	44.00
1	281,000	56.00	100.00
Total	501,794	100.00	

Note that 30% of current buildings in the 1KM zone (outside MN) are at or above the current FAR limit for residential buildings.

. tab FARDiff_pos

FARDiff_pos	Freq.	Percent	Cum.
0	344,232	69.66	69.66
1	149,911	30.34	100.00
Total	494,143	100.00	

To make calculations on the number of new units and the population in the 1KM zone (outside of Manhattan) if built out to the maximum new FAR scenarios, we need to know estimates of average unit sizes and average population per unit.

To get the average unit size, I first calculated the total building area for all buildings currently in the 1KM zone (outside of Manhattan) that are 100% residential. I then calculated the total number of units in these buildings. To get the average unit size I divided total building area by the total number of units. This gives an average size of 973 square feet per unit.

```
(sum bldgarea unitsres if bldgarea==resarea & unitsres>0 & resarea>100 & numbldgs==1 &
borough!="MN")
```

Variable	Obs	Mean	Std. dev.	Min	Max	Total
bldgarea	206,727	4590.902	12446.89	200	623806	949,063,398
unitsres	206,727	4.719538	13.20739	1	1024	975,656

I repeated this exercise for all of New York City and all of New York City excluding Manhattan, and the results give 998 ft^2 and 1017 ft^2 (i.e., not just in the 1KM zone), respectively. To keep things simple, I assumed that the average of all new units in the various TOD scenarios is 1000 ft^2.

The population of NYC from 2020 is 8,804,190 and there are 3,658,920 units in total (from PLUTO), giving 2.4 persons per unit. The population of 4 boroughs excluding Manhattan is 7,109,939 and units are 2,704,762 is 2.63 people per household. (https://www.census.gov/quickfacts/NY). To make things simple I assume that in the new scenarios, the average unit will have 2.5 people per unit.

Note if there are 2 million units in the 1KM zone and we assume a 2.6 people per unit, that gives an estimated population of about 5.2 million people.

Case I: Assume an FAR of 4 for all residential properties unless already over and set their max FAR to their current built FAR. To do this exercise, I took 4*lot area or current FAR*lot area to get the total possible residential floor area. When I looked at the distribution of floor areas from this, I noticed some possible extreme outliers. Thus to keep things "reasonable" I allowed a maximum floor area of 200,000 ft^2 for any given property (this will also diminish maximum number of possible units).

This gives the following statistics that allows for a calculation of total possible floor area:

```
. sum FA4orSame if FA4orSame<200000 & borough!="MN"

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   FA4orSame |    449,715    14106.94     16480.6          0     199916
```

Taking 449,715*14106.94 = 6,344,102,522 ft^2. Assuming 1000 square feet per unit gives 6,344,103 units. Assuming 2.5 people per unit gives a total population of 15.86 million people. Since the average building is 14,107 square feet, that gives an average of 14 units per building.

Case II: Same as above but limit maximum FAR to 3 or current built FAR and limit now building to be more than 200,000 square feet.

```
. sum FA3orSame if FA4orSame<200000 & borough!="MN"

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
   FA3orSame |    449,715    10859.24    13480.58          0     199706
```

This gives an estimated total floor area of 4,883,563,117 ft2, with 4.883 million units, holding a population of 12.21 million.

Case III: Using a gradient approach for maximum allowable FAR is to allow Max of 6 within 200 meters of a stop, 4 from 200 to 500 meters, and 3 from 500 to 1000 meters. Again if the current built far is above that we keep it and we limit maximum building area to 200,000 ft^2.

```
. sum FAR_Gradient if FA4orSame<200000 & borough!="MN"

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
FAR_Gradient |    449,715    13641.03    17036.95          0     299400
```

This gives a total possible floor area of 6,134,575,806 ft^2 with 6.13 million units and a population of 15.34 million.

Subway Ridership

To estimate the impact of residential population on subway ridership, I took the following steps. First, I merged the subway ridership by stop data with the stops shapefile, to get a shapefile that had ridership at each stop. Next using 2020 census tracts and 2020 ACS population counts, I added up the total population of the census tracts that fell within (or crossed) the 1KM buffer for each stop.

I then ran two regressions. One is the log of total annual ridership at each stop as of 2020 on the log of the population estimate, and log ridership in 2019 on the population estimate.

Here are the scatter plots of the relationships (outside of Manhattan). It seems that during the Pandemic that was an shifted downward of ridership, but that the population-ridership elasticity (i.e., the slope) did not change.

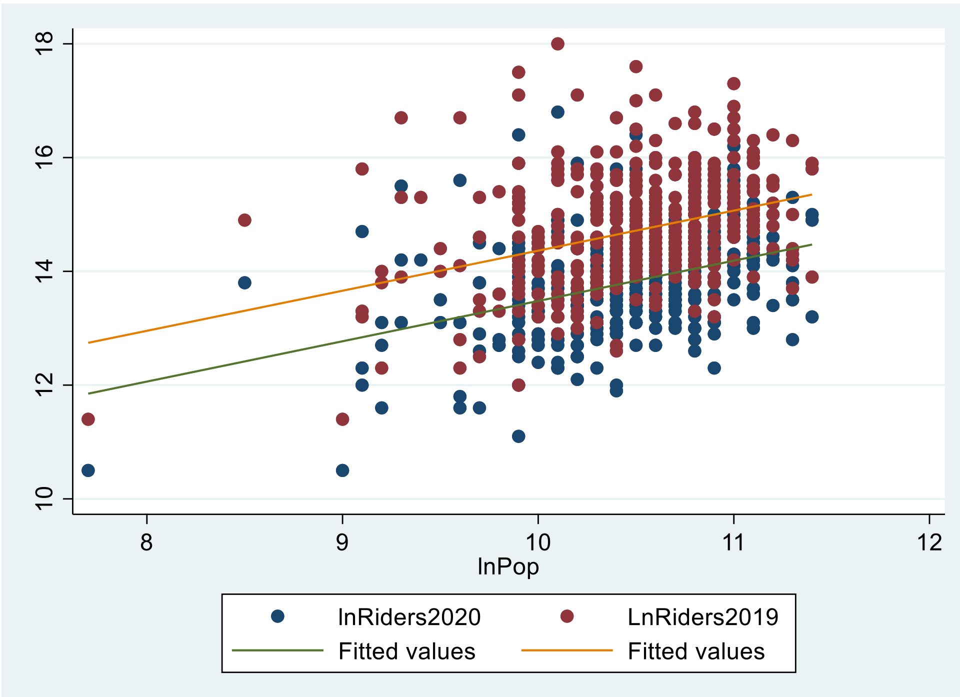

Figure 1: Ln(Stop Ridership) vs Ln(1KM Buffer Zone Pop) outside of Manhattan

Here are the regressions results. They suggest that impact of a 10% increase in residential population is associated with a ~7% increase in ridership. This suggests that for each doubling of the population within the zone, ridership would increase by about close to 70%.

```
. reg lnriders2020 lnpop

      Source |       SS           df       MS      Number of obs   =       423
-------------+----------------------------------   F(1, 421)       =     69.84
       Model |  46.3287688         1  46.3287688   Prob > F        =    0.0000
    Residual |   279.26951       421  .663348006   R-squared       =    0.1423
-------------+----------------------------------   Adj R-squared   =    0.1403
       Total |  325.598279       422  .771559903   Root MSE        =    .81446

------------------------------------------------------------------------------
lnriders2020 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       lnpop |   .7084946   .0847777     8.36   0.000     .5418542    .8751349
       _cons |   6.394261   .8898454     7.19   0.000     4.645168    8.143354
------------------------------------------------------------------------------
```

```
. reg lnriders2019 lnpop

      Source |       SS           df       MS      Number of obs   =       423
-------------+----------------------------------   F(1, 421)       =     56.32
       Model |  45.8009601         1  45.8009601   Prob > F        =    0.0000
    Residual |  342.343066       421  .813166428   R-squared       =    0.1180
-------------+----------------------------------   Adj R-squared   =    0.1159
       Total |  388.144026       422  .919772574   Root MSE        =    .90176

------------------------------------------------------------------------------
lnriders2019 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       lnpop |   .7044472   .0938644     7.50   0.000     .5199459    .8889484
       _cons |   7.319443   .9852208     7.43   0.000     5.382879    9.256008
------------------------------------------------------------------------------
```