“Plants in Space”

Empirical Evidence replication instructions

The National Establishment Time Series (NETS) dataset used in this paper is proprietary and cannot be redistributed. Access can be purchased from Walls & Associates.

The instructions below describe how to query and process the data and create the figures and tables in the paper. A full description of what each code file does is included in the comments at the top of the file.

Each file sets the working directory as "C:\Plants_in_Space\" by default. This must be changed in every file in order to use a different directory.

Data/Raw subfolder

This folder contains the non-proprietary raw data used in this analysis and contains the following subfolders:

Gazetteer	This subfolder contains data on land area at the level of counties and ZIP Code Tabulation Areas (downloaded from https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.2014.html).
LODES	This subfolder contains the LEHD Origin-Destination Employment Statistics (LODES) Workplace Area Characteristics (WAC) data on employment at the census block level in 2011 (downloaded from https://lehd.ces.census.gov/data/lodes/LODES7/).
NHGIS	This subfolder contains population data from the 2010 decennial census (at the census block level) and the 2016 5-year ACS (at the level of counties and ZIP Code Tabulation Areas), downloaded from the National Historical Geographic Information System.
Zillow	This subfolder contains data on rents per square foot at the county and ZIP code level from Zillow, data series “Rental Values, Median ZRI Per Sq Ft: SFR, Condo/Co-Op” downloaded Nov. 2018.

Shapefiles/Raw subfolder

This folder contains the raw shapefiles downloaded from the Census used in this analysis:

US states	https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2018&layergroup=States+%28and+equivalent%29
2010 census blocks	https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2010&layergroup=Blocks
New York, New Jersey, and Virginia files	US counties and equivalent, Virginia area landmarks, New Jersey county subdivisions, New Jersey point landmarks, New York point landmarks, New Jersey-Hudson County water area, New Jersey-Bergen County water area, New York-New York County water area (downloaded from https://www.census.gov/cgi-bin/geo/shapefiles/index.php)

Code folder

The main folder contains two .do files, one .R file, and two subfolders:

Wrapper.do:	This file runs all the .do files required to reproduce the empirical results in the main text and the appendix in the proper order.
set_directories.do	This file creates all necessary subdirectories within the working directory.
clean_tables.R	This file imports the regression tables produced by Stata and reformats them the way they appear in the text. This file should run last (after all Stata files run).

Code/Setup subfolder

Continental_US_Grids.do	This file divides the contiguous US (i.e. the lower 48 states plus DC) into grids of M-by-M mile squares for M=3, 6, 12, 24, and 48.
Grid_Population_and_Employment_Data.do	This file collects data on census block population (from the 2010 decennial census) and employment (from 2011 LODES WAC), and

	maps 2010 census blocks into M-by-M mile squares based on the coordinates of the blocks' centroids. It totals population and employment in each M-by-M mile square, and computes population and employment densities.
Query_and_Clean_NETS_Data.do	This file queries and cleans the 2014 NETS data, removing government, education, nonprofits, and the USPS, establishments outside the contiguous US, and enterprises with less than 5 employees. It then maps establishments into M-by-M mile squares using latitude-longitude coordinates and computes total firm employment. Finally, it saves data for robustness exercises, 1) keeping only firms with at least X=10, 20, 50, and 100 plants, and 2) keeping only firms operating in industries where the firm with the most plants has at least X=10, 20, 50, and 100 plants.
Rent_and_Population_Data.do	This file collects data on rents per square foot from Zillow in 2018 and computes population density using data from the 2016 5-year ACS at the county and ZIP code levels.
NETS_Sorting_Data.do	This file processes NETS data in preparation for the regressions shown in Figure 13 and in Tables I, II, and V. For each firm, it calculates average employment density across all locations in which the firm is located and saves this information along with the firm's national employment.
NETS_Largest_Firm_in_Town_Data.do	This file processes NETS data in preparation for the regressions shown in Tables III and VI. In each industry-location pair, it finds the firm with the greatest number of plants and saves its national employment along with the location's population density.
NETS_Span_of_Control_Data.do	This file processes NETS data in preparation for the regressions shown in Tables IV, VII, and VIII. For each firm-location, it computes the firm's national employment excluding the firm's own contribution in the particular location, and the firm's average employment and number of plants in the location.

regressionMatrix_Ls.m	This file creates two datasets “Ls_sorting_regression_tbl1.csv” and “Ls_sorting_regression_tbl2.csv” that are both being used later to create the regression table X.
regressionMatrixDelta9_Ls.m	This file creates two datasets “Ls9_sorting_regression_tbl1.csv” and “Ls9_sorting_regression_tbl2.csv” that are both being used later to create the regression table X.
wrapper_square.m	This Matlab file creates a set of datasets that are later being used by “regressionMatrix_Ls.m”. It also produces the second row of table IX.
wrapper_square_delta9.m	This Matlab file creates a set of datasets that are later being used by “regressionMatrixDelta9_Ls.m”. It also produces the first row of table IX.

Code/Analysis subfolder

The subfolder "Code/Analysis" contains seven .do files that use inputs produced by files in the “Code/Setup” subfolder. Each file runs everything needed to produce the corresponding set of figures and tables listed in the title of the file.

Figure1.do:	This file identifies the 12-by-12 mile squares roughly containing Princeton, Richmond, and New York City, plots the Starbucks locations in each of these three squares, and produces the three panels presented in Figure 1.
Figure2.do:	This file creates the histograms showing the cumulative number of Walgreens and Rite Aid establishments by population density presented in Figure 2.
Figure3.pdf	This PDF file includes the LaTex code for figure3 that shows the catchment areas for one-

	dimensional case where 5 plants are located in a line. On the left figure b0 is constant and on the right one b0 varies.
Figure4.m	This Matlab file creates the graph of catchment areas for the case where 9 plants are located in the squared space. On the left figure b0 is constant and on the right one b0 varies.
Figure5.pdf	This PDF file includes the LaTex code for figure5 that shows the case where economic activities are uniform across space. On the left figure a firm places N plants in a one-dimensional space uniform space and on the right figure, a firm places N plants in a two-dimensional uniform space.
Figure6.pdf	This PDF file includes the LaTex code for figure 6 that shows a one dimensional representation of the efficiency of distribution.
Figure7.pdf	This PDF file includes the LaTex code for figure 7 which is a graphical representation of the location of plants for a high and low productivity firm.
Figure8_9_10.m	This Matlab file creates the graphs of Figure 8, 9 and 10. Figure 8 presents the distribution of plants, and sales for three representative firms: a firm with the lowest productivity, q = 0.1, a firm with intermediate productivity, q = 1, and a firm with the highest productivity, q = 10. Figure 9 reproduces figure 8 and compares it with findings for the case that the firms efficiency of control is three (sigma = 3). Figure 10 illustrates the impact of improvements in firms span of control technology on lambda (left figure) and x_s (right figure).
Figure11_12.m	This Matlab file produces figures 11 and 12. Figure 11 shows the effect of increase in φ from 0.04 to 0.4 on the mass of plants (left figure) and

	sales of the representative high, medium and low productivity firms. Figure 12 illustrates the effect of improvements in transportation efficiency on total firm sales.
Figure13.do:	This file runs a kernel regression of the relationship between log average firm employment density and log total firm employment after controlling for industry fixed effects, as presented in Figure 13.
TablesI_II_V.do:	This file runs the regressions presented in Tables I, II, and V.
TablesIII_VI.do:	This file runs the regressions presented in Tables III and VI.
TablesIV_VII_VIII.do:	This file runs the regressions presented in Tables IV, VII, and VIII.
FigureB1.do:	This file creates the binscatters of rents versus population density at the county and ZIP code levels presented in Figure B.1.
FigureG1.m	This Matlab file creates two graphs of FigureG1 in the appendix G of the draft. The first plot shows the normalized tightest bound withing the range of delta. The second one presents the log of the average of establishments implied by the upper and the lower bound for the case of k = 0.005.
TableIX_a.m	This file creates the first row of table X that shows the number of plants per firm when delta = 9.
TableIX_b.m	This file creates the second tow of table X that shows the number of plants per firm when delta = 1.
TableX.do	This file produces four columns of table X, showing the results of regressing firm's

employment on its weighted population for different values of delta and M.