"Diverging Trends in National and Local Concentration"

Replication instructions

The National Establishment Time Series (NETS) dataset used in this paper is proprietary and cannot be redistributed. It may be used by employees of the Federal Reserve System and others who have purchased access from Walls & Associates.

The online appendix provides a detailed description of the dataset. The instructions below describe how to query and process the data and create all the figures in the main text and the online appendix. A full description of what each file does is included in the comments at the top of the file. The files should be run in the order listed below as many files use data created in another file as inputs.

Each file changes the working directory to "C:\DTNLC\" by default at the top of the file. This must be changed in every file in order to use a different directory. Extract all code files into a subfolder called "Code" in the working directory, and extract data files into a subfolder "Data".

## Data folder

This folder contains the non-proprietary data used in this analysis.

The subfolder "Raw\CBP\" contains the NAICS to SIC crosswalks which are publicly available from the Census Bureau. It also contains the empty subfolders "_nat", "_county", and "_zip". The County Business Patterns data to fill these folders can be accessed publicly at: https://www.census.gov/programs-surveys/cbp/data/datasets.html. The programs require national and county CBP data from 1990-2014 and zipcode CBP data from 1994-2014, and these data files should be extracted into the subfolders "_nat", "_county", and "_zip", respectively.

The subfolder "Miscellaneous" contains:

County_CBSA_Crosswalk.dta:          a mapping of counties to CBSAs.

County_Lat_Lon_Centroid2012.dta:   a list of all counties in the U.S. states and D.C.

Divisions,MajorGroup,IndustryGroup,SIC4.dta:
                                    descriptions and details about industries at the SIC4 level

sic2_descriptions.dta:              descriptions of industries at the SIC2 level

sic8_descriptions.dta:              descriptions of industries at the SIC8 level

## Code folder

The main folder contains three Stata .do files:

| | |
|---|---|
| wrapper.do: | This creates all necessary subdirectories within the working directory and runs all the .do files required to reproduce the results in the main text and the appendix in the proper order. |
| query_data.do: | This runs a loop to query 25 separate raw NETS data files. Each file corresponds to one year between 1990 and 2014. In each year we get data on all establishments that have positive sales and employment in that year. |
| clean_data.do: | This runs a loop to clean each of the 25 raw data files collected in "query_data.do" and removes the set of industries described in the main text. |

## "Main_Figures" subfolder

The subfolder "Main_Figures" contains 8 Stata .do files. Each file runs everything needed to produce the corresponding set of figures listed in the title of the .do file.

| | |
|---|---|
| figures_1_2_5_6.do: | This produces figures 1, 2, 5, and 6 in the main text. In all figures, we calculate the HHI of sales in each industry-geography-year pair, and for each industry-geography pair, calculate the change in the HHI from the first year that industry-geography pair has an establishment with positive sales and employment. In figures 1 and 2, we take an average of this change across all industry-geography pairs in each year, weighted by employment in each pair, at different definitions of geography. In figures 5 and 6, we take a weighted average only across pairs that have establishments present in all 25 years in our sample. |
| figures_3_4.do: | This produces figures 3 and 4 in the main text. In these figures, we calculate the HHI of either sales or employment for each industry-geography-year pair, and in each year take a weighted average of changes in the HHI across all industry-geography pairs, while varying the definitions of industries and geographies. Figure 3 measures the HHI of sales; figure 4 measures the HHI of employment. |
| figure_7.do: | This produces figure 7 in the main text. For each industry-year pair, we take a weighted average of changes in each industry-geography-year associated with that pair. We then run OLS of weighted average changes in concentration on the year and collect the coefficients at both the national and ZIP code measures of geography to classify industries as having positive or negative national and ZIP code trends. Within each SIC2 code, we |

| | |
|---|---|
| | calculate the percentage of employment in industries with various combinations of national and ZIP code trends. |
| figures_8_9_13_14.do: | This produces figures 8, 9, 13, and 14 in the main text. For each industry, we find the top industry as measured by sales in 2014, and then calculate changes in concentration at the national and local level both including and excluding each industry's top enterprise. Figures 8 and 9 plot these weighted average changes including and excluding the top enterprise for each year from 1990 through 2014. Figures 13 and 14 are event studies that look at changes in concentration before and after the opening of a top enterprise in a local market. |
| figures_10_11.do: | This produces figures 10 and 11 in the main text. Figures 10 and 11 reproduce figures 8 and 9, respectively, but looking at the second and third ranked enterprises in each industry instead of the top enterprise. |
| figure_12.do: | This produces figure 12 in the main text. For each ZIP code-industry-year pair, we determine if that industry's top enterprise is present, and in each year calculate the percent of total employment located in pairs with that top enterprise present. Among this set of pairs with the top enterprise present, we also calculate the percent of aggregate employment in pairs that have either the second or third ranked enterprise (or both) present in addition to the largest enterprise. |
| figures_15_16_17_18.do: | This produces figures 15, 16, 17, and 18 in the main text. Figure 15 repeats the event study in figures 13 and 14, but looking at only the Discount Department Stores industry, Walmart's primary industry where it is the largest enterprise. Figure 16 plots the number of establishments present in each ZIP code-industry-year pair before and after a Walmart opening in that market. Figures 17 and 18 repeat figures 15 and 16, but for the Ready-Mixed Concrete industry, a manufacturing industry whose largest enterprise is Cemex. |
| figures_19_20.do: | This produces figures 19 and 20 in the main text. Figure 10 looks at standardized employment growth in our NETS database versus employment growth in CBP data, from 1990-2014. Figure 20 looks at correlations of county-level and ZIP code-level employment between NETS and CBP in each year between 1990 and 2014. |

**"Appendix" subfolder**

The subfolder "Appendix" contains 9 subfolders. Each subfolder contains a subset of the .do files in the "Main Figures" folder, with names slightly changed, which reproduce the results in the main text for cases where we use alternative specifications of the NETS data, use a different measure of concentration, or use an alternative level of industrial or geographical aggregation. Each file name specifies the figures in the main text reproduced by the file and the modification used.

1. "Removing_HQs_fewer_than_X_employees": This subfolder contains code to reproduce the results in the main text for the case where, in each year, we exclude from the NETS data all enterprises with fewer than X employees in that year, for X=2, 5, and 10.

> figures_1_2_5_6_removing_HQs_fewer_than_X_employees.do
> > (figures 1, 2, 5, 6, 19, 20, 23, 24, 37, 38, 41, and 42 in the appendix)

> figures_3_4_removing_HQs_fewer_than_X_employees.do
> > (figures 3, 4, 21, 22, 39, and 40 in the appendix)

> figure_7_removing_HQs_fewer_than_X_employees.do
> > (figures 7, 25, and 43 in the appendix)

> figures_8_9_13_14_removing_HQs_fewer_than_X_employees.do
> > (figures 8, 9, 13, 14, 26, 27, 31, 32, 44, 45, 49, and 50 in the appendix)

> figures_10_11_removing_HQs_fewer_than_X_employees.do
> > (figures 10, 11, 28, 29, 46, and 47 in the appendix)

> figure_12_removing_HQs_fewer_than_X_employees.do
> > (figures 12, 30, and 48 in the appendix)

> figures_15_16_17_18_removing_HQs_fewer_than_X_employees.do
> > (figures 15, 16, 17, 18, 33, 34, 35, 36, 51, 52, 53, and 54 in the appendix)

2. "Adjusted_HHI": This subfolder contains code to reproduce the results in the main text for the case where we use as our measure of concentration the adjusted HHI of sales, rather than the (unadjusted) HHI of sales.

> figures_1_2_5_6_adjusted_HHI.do
> > (figures 55, 56, 59, and 60 in the appendix)

> figures_3_4_adjusted_HHI.do
> > (figures 57 and 58 in the appendix)

> figure_7_adjusted_HHI.do
> > (figure 61 in the appendix)

figures_8_9_13_14_adjusted_HHI.do
   (figures 62, 63, 66, and 67 in the appendix)

figures_10_11_adjusted_HHI.do
   (figures 64 and 65 in the appendix)

figures_15_17_adjusted_HHI.do
   (figures 68 and 69 in the appendix)

3. "Top_1_share": This subfolder contains code to reproduce the results in the main text for the case where we use as our measure of concentration the share of the top enterprise in total sales, rather than the HHI of sales.

figures_1_2_5_6_top_1_share.do
   (figures 70, 71, 74, and 75 in the appendix)

figures_3_4_top_1_share.do
   (figures 72 and 73 in the appendix)

figure_7_top_1_share.do
   (figure 76 in the appendix)

figures_8_9_13_14_top_1_share.do
   (figures 77, 78, 81, and 82 in the appendix)

figures_10_11_top_1_share.do
   (figures 79 and 80 in the appendix)

figures_15_17_top_1_share.do
   (figures 83 and 84 in the appendix)

4. "HHI_emp": This subfolder contains code to reproduce the results in the main text for the case where we use as our measure of concentration the HHI of employment, rather than the HHI of sales.

figures_1_2_5_6_HHI_emp.do
   (figures 85, 86, 87, and 88 in the appendix)

figure_7_HHI_emp.do
   (figure 89 in the appendix)

figures_8_9_13_14_HHI_emp.do
   (figures 90, 91, 94, and 95 in the appendix)

figures_10_11_HHI_emp.do
   (figures 92 and 93 in the appendix)

figures_15_17_HHI_emp.do
    (figures 96 and 97 in the appendix)

5.  "Balanced_panel": This subfolder contains code to reproduce the results in the main text for the case where we only consider industry-geography pairs which have at least one establishment present in every year in our panel, i.e. a balanced panel.

    figures_3_4_balanced_panel.do
        (figures 98 and 99 in the appendix)

    figure_7_balanced_panel.do
        (figure 100 in the appendix)

    figures_8_9_13_14_balanced_panel.do
        (figures 101, 102, 105, and 106 in the appendix)

    figures_10_11_balanced_panel.do
        (figures 103 and 104 in the appendix)

    figures_15_16_17_18_balanced_panel.do
        (figures 107, 108, 109, and 110 in the appendix)

6.  "HQs_at_least_10_employees_non_imputed_estabs_HHI_emp": This subfolder contains code to reproduce the results in the main text for the case where, in each year, we exclude from the NETS data all enterprises with fewer than 10 employees, then keep only remaining establishments whose employment values are non-imputed, and calculate the HHI of employment.

    figures_1_2_5_6_HQs_at_least_10_employees_non_imputed_estabs_HHI_emp.do
        (figures 111, 112, 115, and 116 in the appendix)

    figures_3_4_HQs_at_least_10_employees_non_imputed_estabs_HHI_emp.do
        (figures 113 and 114 in the appendix)

    figure_7_HQs_at_least_10_employees_non_imputed_estabs_HHI_emp.do
        (figure 117 in the appendix)

    figures_8_9_13_14_HQs_at_least_10_employees_non_imputed_estabs_HHI_emp.do
        (figures 118, 119, 123, and 124 in the appendix)

    figures_10_11_HQs_at_least_10_employees_non_imputed_estabs_HHI_emp.do
        (figures 120 and 121 in the appendix)

    figure_12_HQs_at_least_10_employees_non_imputed_estabs_HHI_emp.do
        (figure 122 in the appendix)

figures_15_16_17_18_HQs_at_least_10_employees_non_imputed_estabs_HHI_emp.do
(figures 125, 126, 127, and 128 in the appendix)

7. "Misc": This subfolder contains two.do files. The file "effect_top_enterprises_number_establishments.do" explores the effect of a top enterprise's opening on the number of establishments in an industry-geography pair, essentially repeating figures 16 and 18 in the main text but for all industries instead of just the industries containing Walmart and Cemex. The file "figures_8_9_top_3_enterprises.do" repeats figures 8 and 9 in the main text but looking at the effects of the top three enterprises in each industry, as opposed to just the top enterprise.

effect_top_enterprises_number_establishments.do
(figure 129 in the appendix)

figures_8_9_top_3_enterprises.do
(figures 130 and 131 in the appendix)

8. "County_CBSA": This subfolder contains code to reproduce figures 2 and 3 in the main text at the county and CBSA levels.

figure_2_county_CBSA.do
(figures 132 and 134 in the appendix)

figure_3_county_CBSA.do
(figures 133 and 135 in the appendix)

9. "SIC4": This subfolder contains code to reproduce figures 1 and 2 in the main text at the SIC4 level rather than at the SIC8 level.

figures_1_2_SIC4.do
(figures 136 and 137 in the appendix)