**Core Datasets**

This paper uses data from several datasets.

County Business Patterns, US Department of the Census

We obtain county-industry-level data on the number of establishments, the number of employees, and the number of establishments in several employment size categories from County Business Patterns. We extracted these data for three industries - hotels (SIC 7011/NAICS 7211), eating and drinking places (SIC 58/NAICS 722) and gasoline service stations (SIC 554/NAICS 4471) – each year between 1964-1992.

The data from 1964-70 are posted on ICPSR. These data include corrections that we made to the 1964-70 data based on cross-validation and comparing to printed sources.

<https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/25984/datadocumentation>

All other years are on ICPSR, and were posted by the U.S. Census.

<https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/22>

Interstate Highway Construction Data

We obtain data on the opening date of each mile of the US Interstate Highway System from the US Department of Transportation’s “PR-511” file. These data are a print-out of the last known version (from 1992) of the file from the Department’s mainframe computers. We hand-entered in these data. Using the PR-511’s record of highway mileposts, we then hand-merged these data with geographic mapping data from the National Highway Planning Network to identify the county in which each segment that was completed in the year was built. (<http://www.fhwa.dot.gov/planning/nhpn>.) We then calculated the number, share, and cumulative share of miles in each highway-county that opened in each calendar year.

We then defined highway corridors, and calculated the number, share, and cumulative share of miles in each highway-corridor that opened in each calendar year. See the text, and the Appendix of Campbell and Hubbard (2016), for a full description of the method.

Temperature Data

We obtain data on temperatures, by county, from the North America Land Data Assimilation System, which reports average high and low temperatures for each county. Our temperature measure is the average between the average high and average low.

<https://wonder.cdc.gov/nasa‐nldas.html>

**Sample Inclusion**

As the text discusses, we restrict the sample to smaller counties where through traffic patterns are simple, and where one might expect highway construction to have a meaningful effect on demand for hotels in the county. We therefore use only counties with only one one- or two-digit interstates and no three digit interstates. This eliminates nearly all counties that are part of large cities (because interstates commonly intersect in such cities). However, a few counties that are part of large cities satisfy this criteria – most notably, Manhattan (New York County, New York). A simple way of eliminating these counties was to simply drop any county whose total employment exceeded 200,000 in 1992.

We also dropped counties in which an interstate passes but there is no exit – one would not expect highway construction to increase demand for hotels in places where travelers cannot exit. We identified county-highways with no exits using various websites that report the milepost of each highway exit, and linking this information to our merged highway construction data.

Finally, we restricted the sample to a balanced panel of counties, where the panel is “balanced” if we observe the number of hotels in each of our sample years. Although there are no disclosure-related issues from the US Census’ perspective in reporting the number of hotels, even in counties with few hotels (the fact that a hotel exists is considered public information), the Census tended not to publish data for industry-counties with fewer than 100 employees during the 1960s and early 1970s. We conjecture that this reflects that reports were printed, and doing so economized on costs.

In this step, we started with the counties used in Campbell and Hubbard’s (2016) related analysis of industry structure in gas retailing – which used only the 676 counties where they observed the number of gas stations in each year between 1964-92. We then restricted the sample to the 227 counties where we observed the number of hotels in each sample year.

In using the data, note that it is common for the data to have missing values for employment (and therefore employees/establishment) for particular county-years. These missing values reflect that the Census did not report hotel employment for that county-year for disclosure-related reasons – it deemed that doing so could reveal an individual hotel’s employment. One example of this is Greenbrier County, WV. This county has a substantial number of individuals employed at hotels, but most of them are employed at the Greenbrier Resort, which is far larger than any other hotel in the county. As a consequence, despite the fact that the number of hotels in this county is reported in each of our sample years, the number of employees/hotel is reported in few years.[[1]](#footnote-1)

Our sample is thus balanced across years when we are analyzing the number of hotels, is unbalanced when we are analyzing hotel employment or employees/hotel.

**Using the Data and Program**

The main dataset is “masterData.dta.” This is the 676 county\*29 year dataset used in Campbell and Hubbard (2016). A secondary dataset is “countytemp.dta,” which contains the county-level temperature data. Both are Stata datasets. They can be read in using Stata 15, and possibly earlier versions of Stata.

The program is “hotels final.do.” This is a Stata do file. It does some straightforward processing with the CBP variables, and merges the countytemp.dta file in with the main dataset masterData.dta. The panel regressions at the end generate all of the paper’s regression results.

The regressions are either ARs (when hotel employment is the dependent variable) or VARs (when the number or employment size of hotels is the dependent variable). As discussed in the text, the definition of “number of hotels” changed between 1973-74. Before 1974, it was the number of firms operating a hotel in the county. (So that if one firm operated two hotels in the county, this was counted once.) Starting in 1974, it was the number of hotels (so that if one firm operated two hotels in a county, this counted as two.) Because of this, we allow the coefficient on the VAR terms to differ for the year 1974 (because the lagged dependent variable is defined differently than the current dependent variable for that year) – including the variables “lestpe\_hotm” and “lempest\_hotm”, which are interactions between our dependent variable and a dummy for the year 1973 accomplishes this. (We do something analogous for our “eating and drinking places” and “gas stations” regressions.)

1. This is the most extreme example in our data. In some specifications, this county effectively plays no part of the analysis because of the number of missing values – Stata will report N=226 rather than 227 in such cases. [↑](#footnote-ref-1)