

This is the replication file corresponding to *Vertical Integration and Input Flows*. For any questions regarding the replication of this article, please send an e-mail to any of the following addresses: atalay@uchicago.edu, hortacsu@uchicago.edu, or Chad.Syverson@chicagobooth.edu.

Enghin Atalay, Ali Hortaçsu, and Chad Syverson

November 6, 2013.

Access to the Data

The three primary datasets used for this study are the Census of Manufacturers, the Commodity Flow Survey, and the Longitudinal Business Database. Access to these data is administered by the Census Bureau, via the Center for Economic Studies (CES) program. For approved projects only, the data can be accessed via the Census Research Data Centers. See <https://www.census.gov/ces/rdcresearch/howtoapply.html> for information on how to apply for access to the Census data.

Programs

There are nine program files, three written in SAS (to pull and clean the data), four written in STATA (to construct the definitions of vertically linked industries), and an additional two written in STATA (to produce the tables and figures of the paper). Programs 4-7 should be executed first, then programs 1-3, and finally programs 8-9.

1. `getdata12.sas`: This .sas file pulls the necessary info for Figure 1 and Tables 1, 2, A2, A3, and A4 of the paper.
2. `getdata345.sas`: This .sas file pulls the necessary info for Figure A1 and Tables 3, 4, 5, A5, and A6 of the paper.
3. `getIntangible.sas`: This file contains the calculations that will be used in the (STATA) analysis corresponding to Tables 6 and A7-A10 of the paper.
4. `io.do`: This do-file produces the definitions of vertically linked industries. It calls the files described in points 5-7, below.
5. `doioindsic.do`: This code is called by `io.do`. It is used to build a correspondence between the BEA IOIND codes and SIC codes. The correspondence draws on information contained in the BEA's industry accounts manual; see in particular pages A33-A36 of the following document <http://www.bea.gov/scb/pdf/national/inputout/meth/io1992.pdf>.
6. `doioindCon.do`: This file, which is called by `io.do`, creates a correspondence between stcc codes and (downstream) construction SIC industries.
7. `doioindRet.do`: This file, which is called by `io.do`, creates a correspondence between ioind codes and (downstream) retail SIC industries. The correspondence is supposed to indicate which particular ioind code commodities can potentially go to any particular retail industry

8. analysisVI_revised.do: This file produces most of the tables and figures of the paper. The only figures/tables not produced by this file are the ones relating to the sharing of intangible capital (see analysisIntangible.do for the code to make these tables/figures). Before running this .do file, the user should run getdata12.sas and getdata345.sas.
9. analysisIntangible.do: This file produces Table 6 and Appendix Tables A7-A10. Before running this piece of code, the user should run getIntangible.sas.

Non-CES Data

Besides the micro data provided to us by the Census Bureau, there are ten data files:

1. bea_use_87_02.csv: This dataset was taken from the 1987, 1992, 1997, and 2002 BEA Use Tables. In the end, we only used data from the 1992 version of this dataset. These data are used to construct the definitions of pairs of industries that are vertically linked to one another.
2. downstream1987.csv: This dataset is taken by combining the 1987 BEA Make and Use Tables. This dataset lists the fraction of IOIND industry J's purchases that come from different upstream industries. This dataset only plays a role in the construction of rows 9 and 10 of Table A2.
3. external.csv: Data from the NBER-CES Productivity database, as well as capital rental rates from the Bureau of Labor Statistics. This dataset is used by getdata345.sas, for the purpose of computing plants' TFPs (as in Tables 3-5, A4, and A5 of the paper).
4. prodMatching8792.csv: This dataset was downloaded from the Census's website (see <http://www2.census.gov/econ1992/MN92/>). Specifically, we downloaded the data from the MN92NCOM.DBF link, and then re-formatted the variables so that they could be read in by SAS. The dataset provides a correspondence of 7-digit product codes for the 1987 and 1992 Economic Censuses. This dataset, and the following dataset, are used in getIntangible.sas, for the purpose of figuring out whether a plant is producing the same product in two consecutive Census years (or whether it is producing a product that its acquiring firm or the acquired firm used to produce).
5. prodMatching9297.csv: This dataset was downloaded from the Census's website (see <http://www.census.gov/prod/ec97/97numlist/97numlist.html>). We then re-formatted the variables so that they could be read in by SAS. The dataset provides a correspondence of 7-digit product codes for the 1992 and 1997 Economic Censuses.
6. retail_purchases.csv: This dataset was downloaded from the Census's website, <http://www2.census.gov/retail/releases/current/arts/> (Specifically, we downloaded the data by clicking purcha.xls on the given website). The dataset contains information on the total purchases made by different retail industries for the years 1992-2009. We then re-arranged this dataset to only include information for four years of interest: 1992, 1993, 1997, and 2002.
7. stcccodes.csv : This concordance between STCC codes, SCTG codes, and SIC codes was provided to us by John Fowler, an economist at the Census Bureau. When providing us

with the dataset, he stressed to us that this classification is not an official Census concordance. It was provided to us with the understanding that this dataset was not intended for general use.

8. `upstream1987.csv`: This dataset is taken by combining the 1987 BEA Make and Use Tables. This dataset lists the fraction of IOIND industry I's sales that go to different downstream industries. This dataset only plays a role in the construction of rows 9 and 10 of Table A2.
9. `wholesale_purchase.csv`: This dataset was downloaded from the Census's website, <http://www2.census.gov/wholesale/xls/awts/> (Specifically, we downloaded the data by clicking `2009_awts_purchmarg_nosmbo_table4.xls` on the given website.) The dataset contains information on the total purchases made by different wholesale industries for the years 1992-2009. From this table, we took info regarding the purchases and gross margins from 1993.
10. `zipcodes.csv`: The source for this database is <http://www.unitedstateszipcodes.org/zip-code-database/>