

Data Description Sheet

Public Firm Presence, Financial Reporting and the Decline of US Manufacturing

Stephen Glaeser, Jim Omartian

This version: June 23, 2022

Data Description

1. A description of which author(s) handled the data and conducted the analyses.

Jim Omartian handled the data and conducted empirical analyses with the exception of creating the ICScore variables, which was performed by Stephen Glaeser. Stephen Glaeser monitored the data collection process and empirical analyses since the start of the project.

2. A detailed description of how the raw data were obtained or generated, including data sources, the specific date(s) on which data were downloaded or obtained, and the instrument used to generate the data (e.g., for surveys or experiments). We recommend that more than one author is able to vouch for the stated source of the raw data.

We use data from a variety of sources as listed and outlined our manuscript.

- US import data from the US Census Bureau Data was downloaded from Peter Schott's website in June 2018. These data were used to create our ImportComp variable. The data can be accessed at the following URL:
<https://www.dropbox.com/sh/7r0gj1dhl6qkww9/AABEbzavXNXXIYoOh7a3LJ4ga?dl=0>
- US Annual Survey of Manufactures and Census of Manufactures information were downloaded from the Census Bureau's website in June 2018.
 - The following datasets were downloaded from the Census Bureau's FTP server (www2.census.gov) for the following years:
 - 2002: ec0231sg102.dat
 - 2003: am0331gs102.dat, am0331gs105.dat, am0331gs106.dat
 - 2004: am0431gs102.dat, am0431gs105.dat, am0431gs106.dat
 - 2005: am0531gs102.dat, am0531gs105.dat, am0531gs106.dat
 - 2006: am0631gs101.dat
 - 2007: ec0731sg2.dat
 - 2008: am0831gs101.dat
 - 2009: am0931gs101.dat
 - 2010: am1031gs101.dat

- 2011: am1131gs101.dat
 - 2012: ec1231sg1.dat
 - 2013: am1331gs101.dat
 - 2014: am1431gs101.dat
 - 2015: am1531gs101.dat
 - 2016: am1631gs101.dat
- For 2001 and prior, data were obtained by scraping PDFs of the Census publication “Statistics of Industry Groups and Industries: 2001.” We downloaded the document in June 2018 and it can be accessed at the following URL: <https://www.census.gov/prod/2003pubs/m01as-1.pdf>
- For our public presence variables we use Compustat annual fundamentals data (last downloaded in August 2018) and segment files (last downloaded in October 2019). These data were obtained from the WRDS web interface.
- Industry concentration data is obtained from the US Census Bureau’s “American FactFinder” website in the following datasets. Each was downloaded in December 2018.
 - EC0231SR12
 - EC0731SR12
 - EC1231SR2
- Related-party trade data were downloaded from the Census Bureau in August 2018:
 - 2005-2016 data were obtained from this website: relatedparty.ftd.census.gov
 - 2000-2004 data were downloaded from the website https://www.census.gov/foreign-trade/Press-Release/related_party/index.html. Follow the links for years 2000-2004 and download exhibit 4 from each.
- SOXBHAR data was constructed from the WRDS event study application in September 2018. We use an event period of 12 trading days starting with 7/8/2002. We estimate abnormal returns as the residual from a model of expected returns based on the Fama-French and momentum factors. We estimate firms’ factor exposures using firm returns over the 100-day window (requiring at least 70 return observations per firm) prior to a 50-day gap before the event
- For matching industries across datasets we use a variety of concordances:
 - NAICS descriptions are obtained from <https://www.census.gov/foreign-trade/reference/codes/naics/naicsmst.txt> in August 2019 to develop a cross-reference for a few Census-specific rollup industries to 4-Digit NAICS codes
 - HS-NAICS concordance cones from Peter Schott’s website and were downloaded in October 2019 (accessible at https://www.dropbox.com/s/0yk03fmitdphhf1/hssicnaics_20181015.zip?dl=0)
 - NACE-NAICS concordance downloaded from https://ec.europa.eu/eurostat/ramon/reactions/index.cfm?TargetUrl=LST_REL&StrLanguageCode=EN&IntCurrentPage=11 in November 2019.
-
- NTR Gap information is from the replication files of “The Surprisingly Swift Decline of US Manufacturing Employment” Pierce and Schott (AER 2016), downloaded in June 2018. Accessible from https://www.aeaweb.org/aer/data/10607/20131578_data.zip
- Import data for non-US countries was obtained from the BACI database downloaded in October 2019 (accessible at http://www.cepii.fr/cepii/en/bdd_modele/presentation.asp?id=1)

- Data for the construction of ICscores is from CRSP (returns and trade volume), Compustat (earnings, industry, earnings release dates, and shares outstanding), and EDGAR (filing dates).
- EDGAR download data was obtained from the EDGAR logfiles on the SEC's website (<https://www.sec.gov/dera/data/edgar-log-file-data-set.html>) during April 2019.
- IP location data was obtained from lite.ip2location.com/database/ip-country during April 2019.
- Bureau van Dijk's Orbis data are used for our German and UK analyses. We obtain these data from University of Michigan's June 2019 historical data snapshot. Specifically, we use the following files:
 - Key_financials-USD.txt
 - Industry_classifications.txt
 - Legal_info.txt

Stephen Glaser and Jim Omartian can vouch for all sources.

3. If the data are obtained from an organization on a proprietary basis, the authors should privately provide the editors with contact information for a representative of the organization who can confirm data were obtained by the authors. The editors would not make this information publicly available. The authors should also provide information to the editors about the data sharing agreement with the organization (e.g., non-disclosure agreements, any restrictions imposed by the organization on the authors, such as restrictions to publish certain results).

No confidential data is used in this study. All datasets are publicly available for download from the internet with the exception of Compustat, BvD Orbis, and BACI. Each of these three datasets is commercially available from the data providers and was used under the University of Michigan's licenses with the data providers. Kresge Library (contact Todd Elkins: tle@umich.edu) manages subscriptions. BACI is available to the University of Michigan as a result of the University's subscription to the United Nations COMTRADE database.

To be provided in the paper or the online appendix:

4. A complete description of the steps necessary to collect and process the data used in the final analyses reported in the paper. For experimental and survey papers, we require information about the instructions and instruments used to generate the data, subject eligibility and/or selection, as well as any exclusion criteria. The full set of instructions and instruments can be provided in the online appendix.

Raw data were obtained from the sources listed in Item #2 above and imported into R will be included in the final dataset. The respective code will be made available according to the journal guidelines. The paper verbally describes the complete data collection and processing steps. No survey instrument was used, nor was any experiment conducted.

To be provided upon acceptance of the paper and prior to publication:

5. The computer programs or code used to convert the raw data into the final dataset used in the analysis plus a brief description that enables other researchers to use this program. The purpose of this requirement is to facilitate replication and to help other researchers understand in detail how the raw data were processed, the final sample was formed, variables were defined, outliers were treated, etc. This code or programming is in most circumstances not proprietary. However, we recognize that some parts of the code or data generation process may be proprietary, including from the authors' perspective. Therefore, instead of the code or program, researchers can provide a detailed step-by-step description of the code or the relevant parts of the code such that it enables other researchers to arrive at the same final dataset used in the analysis. In such cases, the authors should inform the editors upon initial submission, so that the editors can consider an exemption from the code sharing requirement. Whenever feasible, authors should also provide the identifiers (e.g., CIK, CUSIP) for their Authors should consult our FAQ Sheet on the JAR website for further details.

The R code and data identifiers to recreate our results are available on the JAR website with our submission.

6. An assurance that the data and programs will be maintained by at least one author (usually the corresponding author) for at least six years, consistent with National Science Foundation guidelines.

Jim Omartian will maintain the data and programs for at least six years.