The Census Bureau’s Business Microdata

Randy A. Becker
Center for Economic Studies
U.S. Census Bureau

Dallas-Fort Worth Research Data Center
Research Workshop
December 5, 2018
... And we can also hope to learn much more in future from the studies of the activities of firms which have recently been initiated by the Center for Economic Studies of the Bureau of the Census of the United States. ...” Ronald H. Coase, Nobel Prize Lecture, 1991
Business Microdata from the Census Bureau

• Business microdata (or “economic” microdata) include data at the:
  • Establishment-level (i.e., individual business location/address)
  • Firm-level (i.e., encompassing all establishments under the same ownership)
  • Transaction-level (exports, imports, domestic shipments)
  • Individual-level (non-employer businesses, business owners, workers)
  • Multiple levels
    ➢ Export & import shipments linked to the U.S. firms involved (LFTTD)
    ➢ Workers (earnings, characteristics) linked to the firms that employ them (LEHD)
Advantages of Census Bureau

Restricted-Use Business Microdata

- There is no public-use version of these
  - At best, published data is at the industry-level, at the county-level
  - It is difficult/impossible to produce public-use business microdata and maintain confidentiality
- Data for the entire universe of businesses, or for a fully representative scientific sample
- Mandatory reporting (very high response rates)
- Detailed industry (and even sub-industry/product-level detail)
- Detailed geography (sub-county, including address & latitude-longitudinal)
- Data from a business have been linked over time (longitudinal links)
- Data for a business from different Census surveys can be linked together (cross-sectional links)
- Data from non-Census sources (e.g., Compustat, patents, TRI) can be linked in via name-and-address matching
“Popular” Business Microdata

Business Register

- The Business Register (BR) is the backbone of nearly all Census Bureau business data (sampling frame)
  - Essentially the universe (7.5 million) of all business establishments in the United States with at least one paid employee
  - Data items are fairly limited:
    - Business name & address
    - EIN
    - Location (county, state)
    - Industry code
    - Employment
    - Payroll
    - Corporate structure (firm ID ➔ sibling establishments; legal form of organization)
    - Identifiers that allow cross-sectional matching across all business surveys & censuses
- We learn of new firms from the IRS
- We learn of new establishments within a firm from various Census Bureau surveys
“Popular” Business Microdata
Longitudinal Business Database

• The **Longitudinal Business Database (LBD)** is the “research version” of the Business Register
  • Created by the Center for Economic Studies
  • Essentially the universe (7.5 million) of all business establishments in the U.S. with at least one paid employee
  • Covers most industries (not farms, not government sector)
  • 1976 to present (currently 2015)
  • Similar data items as BR: location (county, state), industry code, employment, payroll, corporate structure (allows aggregation to the firm level)
  • No name & address information
  • Longitudinal identifier (LBDNUM) allows for the identification of establishment entry, exit, continuers (growth), and age
  • Can be linked to all other Census business surveys and censuses
“Popular” Business Microdata
Economic Censuses

- Richer data on establishments and firms come from the myriad of surveys conducted by the Census Bureau, including the **Economic Census**
  - Conducted every 5 years (for years ending in ‘2’ and ‘7’)
  - Covers nearly all sectors of the economy: retail, services, manufacturing, wholesale, construction, financial industries, mining, transportation, utilities
    - No agriculture (except agricultural services)
  - Approximately 7 million employer businesses
  - “All” establishments, except that there are non-mailed cases and non-respondents whose data are fully imputed using BR data
  - About 800 unique survey forms (now, electronic paths)
    - Collects data on value of output, payroll, employment, materials consumed, fuels & electricity consumed, depreciable assets, selected purchased services, inventories, and capital expenditures, output by product line, etc.
    - Also collects data specific to an industry or sector
  - The primary source of information about the structure and functioning of the nation’s economy
    - Important part of the framework for the national income and product accounts, input-output tables, etc.
    - Provides sampling frames and benchmarks for current business surveys, which track short-term economic trends, serve as economic indicators, and enter into GDP estimation
Annual surveys collect similar data to the Economic Census, except:

- Fewer sectors: manufacturing, services, retail, wholesale
- Only manufacturing is at the establishment-level (the others are at firm-level)
- Just a sample of the universe
  - Businesses above a certain size are “certainty cases” with weight=1 (at least at the outset)
  - Businesses below that size are “stratified”
    - Establishments are randomly chosen within a strata, with a probability proportionate to its size
    - Weight is constructed accordingly

- There are also quarterly and monthly surveys available
“Popular” Business Microdata
Manufacturing Surveys

• Census of Manufactures (1963, 1967, 1972, .... 2012)
• Annual Survey of Manufactures (1972-present) – includes constructed TFP
• R&D expenditures (BRDIS/SIRD)
• Capacity utilization (PCU/QPC)
• Energy consumption (MECS)
• Technology use (SMT, CNUS)
• Environmental expenditures (PACE)
• Exports & imports (LFTTD)
• Domestic shipment destinations (CFS)
• Patents
• Capital expenditures & ICT expenditures
• Management & organization practices (MOPS)
• Water use (SWUM)
“Popular” Business Microdata
Longitudinal Firm Trade Transactions Database

• The **Longitudinal Form Trade Transactions Database (LFTTD)** links individual trade transactions to the firms that undertake them.
  • Export transactions ↔ Exporter (37 million in 2015)
  • Import transactions ↔ Importer (77 million in 2015)

• Transaction information includes:
  • Commodity
  • Value
  • Quantity
  • Weight
  • Date
  • Origin
  • Destination
  • Method of transportation
  • Related-party indicator

• 1992 – 2015
“Popular” Business Microdata
Longitudinal Employer-Household Dynamics (LEHD) Data

• Constructed by the Center for Economic Studies

• Worker files
  • Employee earnings and employment (from state UI records)
  • Individual characteristics (sex, age, education, race, ethnicity)

• Employer files
  • Business characteristics (industry, location, size, age)
  • Workforce characteristics
## Business Register Data

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Frequency</th>
<th>Unit of Enumeration</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form 5500 Bridge File</td>
<td>Annually</td>
<td>Establishment</td>
<td>1992–2012</td>
</tr>
<tr>
<td>Ownership Change Database (OCD)</td>
<td>Every 5 Years (Discontinued)</td>
<td>Establishment</td>
<td>1963, 1967–2002</td>
</tr>
</tbody>
</table>

**The CSB is only available with a current Compustat license.**
<table>
<thead>
<tr>
<th>Data Set</th>
<th>Frequency</th>
<th>Unit of Enumeration</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census of Auxiliary Establishments (AUX)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1977–2012</td>
</tr>
<tr>
<td>Census of Construction Industries (CCN)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1972–2012</td>
</tr>
<tr>
<td>Census of Finance, Insurance, and Real Estate (CFI)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1992–2012</td>
</tr>
<tr>
<td>Census of Island Areas – Puerto Rico (CIAPR)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1992–2007</td>
</tr>
<tr>
<td>Census of Manufactures (CMF)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1963, 1967–2012</td>
</tr>
<tr>
<td>Census of Mining (CMI)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1987–2012</td>
</tr>
<tr>
<td>Census of Retail Trade (CRT)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1977–2012</td>
</tr>
<tr>
<td>Census of Services (CSR)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1977–2012</td>
</tr>
<tr>
<td>Census of Transportation, Communications, and Utilities (CUT)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1987–2012</td>
</tr>
<tr>
<td>Census of Wholesale Trade (CWH)</td>
<td>Every 5 Years</td>
<td>Establishment</td>
<td>1977–2012</td>
</tr>
</tbody>
</table>
### Other Establishment Surveys

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Frequency</th>
<th>Unit of Enumeration</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Industrial Reports (CIR)</td>
<td>Annually (Discontinued)</td>
<td>Establishment</td>
<td>1993–2011</td>
</tr>
<tr>
<td>Quarterly Survey of Plant Capacity Utilization (QPC)/Survey of Plant Capacity Utilization (PCU)</td>
<td>Quarterly</td>
<td>Establishment</td>
<td>1974–2017</td>
</tr>
</tbody>
</table>
## Firm Surveys

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Frequency</th>
<th>Unit of Enumeration</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Includes</em> Information and Communication Technology (ICT) Supplement starting in 2003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual Retail Trade Survey (ARTS)</td>
<td>Annually</td>
<td>Firm</td>
<td>1999–2015</td>
</tr>
<tr>
<td>Annual Survey of Entrepreneurs (ASE)</td>
<td>Annually</td>
<td>Firm</td>
<td>2014</td>
</tr>
<tr>
<td>Business Expenditures Survey (BES)</td>
<td>Every 5 Years (Discontinued)</td>
<td>Firm</td>
<td>1992–2007</td>
</tr>
<tr>
<td><em>Also known as</em> Business Expenses Supplement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>See also</em> SIRD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Must be used in conjunction with</em> Census Bureau restricted-use microdata</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manufacturers' Shipments, Inventories, and Orders (M3)</td>
<td>Monthly</td>
<td>Firm</td>
<td>1992–2017</td>
</tr>
<tr>
<td>Manufacturers' Unfilled Orders Survey (M3UFO)</td>
<td>Annual</td>
<td>Firm</td>
<td>2009–2013</td>
</tr>
<tr>
<td>Quarterly Services Survey (QSS)</td>
<td>Quarterly</td>
<td>Firm</td>
<td>2003–2014</td>
</tr>
</tbody>
</table>

---

The table above lists various surveys and their associated data sets, frequencies, units of enumeration, and availability periods.
Summary of Recovered Historical Economic Microdata
August 2012

This summary highlights some of the business microdata recovered from the Census ClearPath mainframe. Note it is very difficult to summarize this information succinctly, for a variety of reasons, but the last of the list below, there are thousands of files, many with only a crude description (if any at all) in the associated paper file. Also note, presence on the list below does not necessarily guarantee these are usable microdata files. And, in some cases, the lack of a record layout may make recovered data very difficult to use—unless related data and or record layouts exist. Further details are available upon request. Meanwhile, for the list of surveys and years that CBS currently has and makes available, see http://www.census.gov/acs/dataproducts/

For more information on the recovered data, email CBS.Data Recovery.List@census.gov or see http://www.census.gov/acs/dataproducts/recovered/

Manufacturing and Mining

Census of Manufactures
- Special inquiry data
- Class of customer data
  - 1967, 1977
- CAO (auxiliary establishments)
  - 1977
- MA-131 (Supplemental Survey of Materials Consumed)
  - 1972
- Location of plants
  - 1963
- Ancillary data (e.g., flags, name and address, edit parameters)
Census of Manufacturers (Puerto Rico)
- 1967, 1972, 1977
Census of Mining

Annual Survey of Manufactures
- Main files
- ITS (longitudinally-linked ASM) estabs
  - 1954-1961
- CAO (auxiliary establishments)
- Fixed Base Difference files
- MA-175 (Defense, Government)
  - 1981-84
- Defense Mobilization
  - 1987, 1989
- Ancillary data (flags, parameters, sampling universe, etc.)
  - Various years
- Name and address files
  - Various years
- Fuel and Electric Energy supplement
  - 1974, 1976
- MA-452 (Survey of the Consumption of Selected Hydrocarbons, Coal, and Coke Materials)
  - 1978-80
- Survey of Water Use in Manufacturing
- Capital Asset and Depreciation Survey
- Survey of Industrial R&D
  - 1963-70
- Survey of Plant Capacity
  - 1976-88 (additional data and flags)
- Pollution Abatement Costs and Expenditure Survey
- Zvi (Griffiths) project files: (includes R&D, CMS)
  - 1953-1990
- Lawrence Klein project files
  - 1954-59, 1972-81
- Annual Survey of Oil and Gas
  - 1974-82

Business (i.e., non-manufacturing)

Census of Business
- 1972
Census of Retail (various)
- 1963, 1967(Puerto Rico), 1972(Capital Expenditure)
Census of Wholesale
Census of Services

1 A small portion of the listed microdata were recovered from other sources.

2 This survey is at times also called the “Annual Trade Survey”.

Service and Retail Capital Expenditure Survey
- 1972
Annual Retail Trade Survey
Annual Wholesale Trade Survey
Services Annual Survey
- 1987, 1990-97
Housing Fuel Survey
- 1979-80
Survey of Gallium Nitride of Gasoline
- 1980-82
Canned Food Survey
- 1981
International Trade
Imports (transactions)
- 1980-90
Business Register
- 1974-87, 1989
County Business Patterns establishment files
- 1976-84
Other administrative records files
- 1977-82
Enterprise Statistics
- 1972, 1977
Non-employers
- Survey of Minority-owned Business Enterprises
Outlying Areas
Economic Census
- 1982
Puerto Rico
Census of Retail
- 1967
Census of Manufactures
- 1967, 1972, 1977
Census of Construction
- 1977
Transportation
Commodity Transport Survey
- 1967, 1972, 1977
Domestic and International Transportation of U.S. Foreign Trade
- 1976
DOT Truck Inventory and Use Survey (public use?)
- 1977
Railroad and Tracking Survey
- 1990-97
Construction
Building Permits
- 1975-86, 1989
Survey of Housing Starts, Sales, and Completion (HSS)
- 1979-83
Progress Reporting
- 1973-83
Agriculture
Census of Agriculture
- 1964, 1969, 1974, 1979 ("Farm Real Survey")
- 1998, 2002, 2007...
Census of Agriculture - Puerto Rico
Census of Agriculture - Guam, Virgin Islands
- 1982
Census of Agricultural Services
- 1969, 1978
Farm Finance Survey
Census of Commercial Fisheries
- 1967
Census of Irrigation
- 1969
Census of Horticulture
- 1975, 1988

1 This survey is at times also called the “Annual Trade Survey”.
# Transaction Data

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Frequency</th>
<th>Unit of Enumeration</th>
<th>Availability</th>
</tr>
</thead>
</table>
Longitudinal Employer-Household Dynamics (LEHD) Data

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Unit of Observation</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Register Bridge (BRB)</td>
<td>Establishment</td>
<td>1990–2008</td>
</tr>
<tr>
<td>Employer Characteristics Files (ECF)</td>
<td>Establishment – Quarter</td>
<td>1989–2011</td>
</tr>
<tr>
<td>Employment History Files (EHF)</td>
<td>Job (Person–Firm)</td>
<td>1985–2011</td>
</tr>
<tr>
<td>Geocoded Address List (GAL)</td>
<td>Establishment</td>
<td>1990–2011</td>
</tr>
<tr>
<td>Individual Characteristics Files (ICF)</td>
<td>Person</td>
<td>1985–2011</td>
</tr>
<tr>
<td>Quarterly Workforce Indicators (QWI)*</td>
<td>Establishment – Quarter</td>
<td>1990–2011</td>
</tr>
<tr>
<td>Unit–to–Worker (U2W)</td>
<td>Job (Person–Establishment)</td>
<td>1990–2011</td>
</tr>
</tbody>
</table>
The Productivity of RDC Researchers Using Census Bureau Microdata

• In 2017, RDC researchers using Census Bureau microdata published:
  • At least 91 working papers (CES, NBER, university mimeos, etc.)
  • At least 81 journal articles (including those known to be forthcoming)
    • 11 in the Top 5 journals in economics (AAA)
    • 11 in AA journals in economics (ranked 6-20)
    • 22 in A journals in economics (ranked 21-102)
    • 16 of the top 22 journals in economics published RDC research in 2017 (or forthcoming)

• From 2010-2017 (and forthcoming)
  • 10 articles in American Economic Review (does not include P&P)
  • 9 articles in Quarterly Journal of Economics
  • 8 articles in Journal of Political Economy
  • 3 articles in Review of Economic Studies
  • 2 articles in Econometrica

• For examples, see:
  • Appendix 2 of each CES & RDC Annual Report:  www.census.gov/ces/publications/publications.html
Research Tips
For Proposal Development & Beyond

• Find all relevant Census Bureau publications (or webpages) containing:
  • Sample details (size, coverage, etc.)
  • Survey methodology (editing, imputation, weighting, etc.)
  • Survey instrument & instructions (variable definitions)
  • Published tables

• Often PDF scans are available online (Census Bureau, archive.org, Google Books)
• RDC administrator (and RDC researchers) have access to survey forms archive
• Feel free to contact me for help in finding publication or survey form

• Recent dissemination has been all electronic (web pages, American FactFinder)

• Seek out working papers, publications, and researchers that have used the same data
Research Tips
For Early Stages of Research

• Try to replicate published statistics
  • Key variables: Number of firms, number of establishments, employees, etc.
  • Tabulate by industry, by geography, etc.
  • Look for “tabulation flag”
  • Use weights (if applicable)

  • Getting to within ±5% is a good way of knowing that you’ve got the right data, the right set of cases, and so forth

• Beware of imputation!
  • Survey data are generally considered the most reliable data, followed by administrative data, followed by imputed data
  • Look for impute or “source” flags
  • Most likely for smaller businesses
  • Imputes are designed to make reliable aggregates -- not necessarily reliable microdata
Graduate Students & Dissertations

• At the end of 2017, there were about 130 graduate students using the RDCs for their Ph.D. dissertation work
  • 48 different universities
  • 108 using Census Bureau microdata (vs. NCHS, AHRQ)
  • This does not include the many graduate students who use the RDCs as research assistants to others

• CES Dissertation Mentorship Program
  • 44 participants since 2008 from 25 different universities:
    • Columbia University, Cornell University, Duke University, Florida State University, George Washington University, Georgia State University, Georgia Tech, Harvard University, Massachusetts Institute of Technology, Michigan State University, Stanford University, Syracuse University, Texas A&M University, University of Houston, University of Illinois at Chicago, University of Iowa, University of Maryland, University of Michigan, University of Minnesota, University of Missouri, University of North Carolina, University of Pennsylvania, University of Texas, University of Virginia, University of Wisconsin
CES Dissertation Mentorship Program

• CES wants to assist doctoral candidates who are actively engaged in dissertation research in economics, sociology, demography, or a related field using Census Bureau microdata at a Research Data Center.

• Program participants receive two principal benefits:
  • One or more CES staff economists are assigned as mentors and advise the participant on the use of Census Bureau microdata.
  • A visit to CES where they meet with staff economists and present research in progress. Trip expenses (airfare, hotel, and meals) will be paid for by CES.

• Eligible candidates must be
  • Actively working on a Ph.D. thesis in economics or a related field using Census Bureau microdata at a Research Data Center.
  • A U.S. citizen

• To apply:
  • Submit the following via email to CES.Mentorship.List@census.gov
  • Contact information, name(s) of faculty advisor, and expected completion date, or a CV with that information.
  • Title and abstract of RDC–based dissertation research, including a description of microdata being used.

• Acceptance will be on a rolling basis, as funding and staffing resources allow.
My “Academic” Research

• The effects of environmental regulation on the behavior of manufacturing firms, including:
  • Plant location, investment, and survival (CM)
  • Production costs (CM, ASM, and PACE survey)
  • Environmental spending (ASM, PACE survey)
My Research (cont.)

• Environmental spending by business size (PACE, ASM, CM)

• Environmental spending & pollution havens (PACE, ASM)

• Spatial heterogeneity in environmental regulation (PACE)

• Productivity (CM, ASM)

• Employment (ASM, CM)
My Research (cont.)

• International trade (Imports, Exports)

• The “green industry” (CM, 1995 Survey of Environmental Products & Services)

• Industrial water use (ASM, Survey of Water Use in Manufacturing)
Some By-Products of This Research

• Involvement in the redevelopment of Pollution Abatement Costs and Expenditures (PACE) Survey, with EPA

• Advisory roles on the “green industry” with the Brookings Institute, Department of Commerce, and Bureau of Labor Statistics

• Water use in manufacturing & mining added to the 2017 Economic Census
Contact Information

Randy A. Becker
randy.a.becker@census.gov
301-763-1891