

A User's Guide to the PSU-Federal Statistical Research Data Center

Mark Roberts
Professor of Economics and Director PSU RDC
Chris Galvan
Administrator PSU RDC

March 2015

- Introduction to the Federal Statistical Research Data Centers (RDC)
- Background of the PSU RDC
- Data Resources in Economics
- Data Resources in Health and Demography
- Procedures for Using the PSU RDC.

Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau.

Federal Statistical Research Data Centers

- Network of secure data centers that provide researchers with access to restricted data
 - Census Bureau
 - Agency for Healthcare Research and Quality (AHRQ)
 - National Center for Health Statistics (NCHS)
 - Bureau of Labor Statistics (in progress)
- 19 RDCs in operation - most major Research-I Universities have access.
 - Washington, New York, Boston, Atlanta, Chicago areas.
 - Minnesota, Texas, Michigan, Cornell, Duke, Berkely, Stanford, UCLA, USC, Penn State, Yale
- 5 under construction with more planned in future years
- www.census.gov/fsrdc and www.census.gov/ces

- Public-use data sets are widely used in research: ACS, NHANES, PUMS, SIPP
- Limitations due to need to protect confidentiality
 - Limited geographic information. Limits ability to link contextual variables
 - Top-coding of responses (income)
 - Exclude sensitive personal characteristics (weight)
 - Difficult or impossible to link individuals or businesses across surveys
 - Difficult to reconcile differences in estimates across surveys
 - Data restrictions have tightened over time

What Does the RDC Provide?

- Access to the microdata on individuals or businesses in the surveys
 - Detailed geographic identifiers to allow merging of city, county, or state information.
 - More detail on characteristics - place of birth, date of birth, occupation, income, firm or plant size
 - Individual identifiers that allow merging with administrative data sources and across surveys
- Maintains confidentiality of individual responses
 - Formal review process for research proposals which examines need for confidential data and feasibility of the project
 - Security clearance for researchers
 - Secure computing environment (Linux) to access census data servers
 - Commonly used software on servers
 - Disclosure review of outputs

The PSU Research Data Center

- Opened in April 2014, Room 206 Paterno Library
- Funding provided by NSF, Office of the VP for Research, SSRI, Population Research Institute, University Libraries, Colleges of Liberal Arts, Health and Human Development, Agriculture and Eberly College of Science
- Available to all faculty and grad students (with faculty sponsor)
- Currently have 11 approved projects and 20 faculty and grad students have clearance
 - Sociology and Demography: Jennifer Van Hook, Melissa Hardy, Shannon Monnat
 - Economics: James Tybout, Steven Yeaple, Mark Roberts
 - Business: Yelena Larkin
 - Agricultural Economics: Spiro Stefanou and Pinar Geylani (Duquesne)
 - Geography: Chris Fowler

Primary source for micro data on firms and establishments. Covers all sectors of the economy. Only comprehensive longitudinal micro data source.

- Employer Dynamics - entry, growth, and exit of firms
- Employment Flows - job creation and destruction due to firm turnover
- Productivity Growth - role of turnover and investment in capital and R&D
- Vertical Integration of Firms - linkages across plants in production chain
- Firms in International Trade - importance of multinational firms in U.S. exports and imports, formation of trading networks

Longitudinal Business Database (LBD)

- annual 1976 - 2011
- covers 8.5 million establishments across whole economy (non-farm, paid employees)
- establishments are linked over time - permanent ID number
- establishments can be aggregated to the firm (enterprise) level - using firm ID
- limited information: geography, industry, payroll, employment
- major strength: plant ID number can be linked to other CB surveys and censuses
- <http://www.census.gov/ces/pdf/CES-WP-02-17.pdf>
- Synthetic version (SynLBD) available at Cornell RDC - no geographic information

Economic Censuses

Five year intervals, mix of establishment surveys and administrative data

- Manufactures 1963 - 2007
- Construction 1972 - 2007
- Retail Trade, Wholesale Trade, Services 1977 - 2007
- Mining 1987 - 2007
- Transport-Communications-Utilities 1987 - 2007
- Finance-Insurance-Real Estate, 1992 - 2007

Detailed establishment information (*later year, manufacturing example*)

- six-digit NAICS industry, geography (census place)
- employees, production workers, hours, payroll, employee benefits,
- value of shipments (primary and secondary products), inventories
- cost of materials, energy expenditures
- capital investment by categories (building, equip, transport)
- computer and communication expenditures, data processing

For sector detail see <http://www.census.gov/econ/census07/>

Annual Surveys.

Establishments or firms in non-census years

Samples that are stratified by size. Rotating sample every 5 years.

Goal: create industry (6-digit) and geographic totals for key variables

- Manufacturing - establishment based. 1973-2011, shipments, investment, payroll, employees
- Retail Trade - firm based. 1999-2010, revenues, inventories, operating expenses
- Wholesale Trade - firm based. 1995-2010
- Services - 1991-2010 except 1999-2001

Longitudinal Employer-Household Dynamics (LEHD)

- Combined worker-firm data
- Built up from state unemployment insurance records, merged with census surveys.
- Job-level quarterly earnings data, information on where workers live and work, and data on firm characteristics.
- All 50 states contribute data but access varies - Requires approval of every state used. Some states review every project, others give blanket permission. Approval for large, multi-state projects is difficult.
- 2000-2008, some states have older data.
- Used to construct **Quarterly Workforce Indicators (QWI)** - Public use versions. Employment levels and flows by firm characteristics (geography, industry, age, size) and worker characteristics (sex, age, education, race, ethnicity). Establishment level data available in RDC.

Other Specialized Economic Surveys

All have unique sampling frames. Can be linked to LBD.

- **Quarterly Financial Reports** - 10,000 firms in manufacturing, wholesale, retail, mining. Debt structure, liquidity, profits.
- **R&D and Innovation Surveys** - firm level, 1972-2009. R&D expenditure, employment, management, innovation, tech transfer, intellectual property. Large employment firms always included.
- **Survey of Business Owners** - firm level, 1972-, Ethnicity, gender, race, vet status. Includes non-employee firms.
- **Capital Expenditure Survey** - firm level. Spending on equipment and structures. Includes non-employee firms.

Other Specialized Economic Surveys (continued)

- **Pollution Abatement and Control Expenditures** - establishment, various years 1974 - 2005
- **Survey of Manufacturing Technology** - establishment, 1988, 91, 93
- **Manufacturing Energy Consumption** - establishment, various years 1985 - 2002
- **Manufacturing Plant Capacity Utilization** - quarterly, 1974 - 2011

Complete listing at

<http://www.census.gov/ces/dataproducts/economicdata.html>

Longitudinal Firm Trade Transaction Database

- Export Transactions - monthly 1994-2011, value, date, country of origin or destination, quantity, product code. Covers whole economy (34 million records in 2011).
- Import Transactions - similar to exports (59 million in 2011).
- Includes firm id so they can be linked to the LBD and other census surveys