The Atlanta Census Research Data Center

Julie Hotchkiss
Executive Director, ACRDC
Research Economist and Policy Advisor
Federal Reserve Bank of Atlanta

August 24, 2011

What is a Census Research Data Center?

- A secure physical location where researchers conduct approved research using restricted data collected by Census and other agencies.
- Established through an agreement between Census and a local research community.
- Provided for through:
  - Title 13, sec. 9 & 214 (Census)
  - use of the data only by employees and sworn officers of Census
  - Title 26, subtitle f, Ch. 61, Sub Ch. B, § 6103(j) (IRS)
  - allows sharing of information with Commerce for statistical use
  - CIPSEA (Title 44, Ch. 35, Sub Ch. I, § 3501)
  - use of confidential data must have a statistical purpose
  - provides for severe penalties for disclosing confidential data

Census Research Data Center Locations

The Atlanta Census Research Data Center

- Located in the Federal Reserve Bank of Atlanta
  - corner of 10th & Peachtree
- Consortium Members
  - Emory University
  - University of Georgia
  - Georgia State University
  - Georgia Institute of Technology
  - Federal Reserve Bank of Atlanta
  - University of Alabama-Birmingham
  - Centers for Disease Control and Prevention
- Expected Opening: September 2011

Advantages of Access to Restricted Data

- Data not available elsewhere
  - Establishment level business data
  - Linked worker-firm data (LEHD)
- Detailed geo-spatial variables
  - Census block group
  - Longitude and latitude
- Virtually no top or bottom coding of variables
  - earnings, income

Advantages, cont.

- Can link many Census datasets with each other
  - Current Population Survey with LEHD
  - Economic Census with Survey of Business Owners
- Can link to other non-Census data
  - NCHS data with ACS
  - Kauffman Firm Survey of new business formation
Types of Data Available

- Economic Data
  - Economic Census
  - Business Register data
  - Establishment data vs. Firm data
  - Transactions data

- Demographic Data
  - Census data (Decennial Census, ACS, SIPP, etc.)
  - American Housing Survey

- LEHD Data

- Health Data
  - National Center for Health Statistics
  - Agency for Healthcare Research & Quality

Types of Data Available: Economic Data

- Economic Census
  - Establishment level surveys conducted every 5 years
  - Public use versions limited to data in aggregate form

- Manufactures, Mining, Construction, FIRE, Services, TCU, Retail & Wholesale Trade

- Sample Research
  - "Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s" Foster et al.

  Using the Census of Retail Trade, the authors quantify and explore the relationship between the restructuring in the retail sector that occurred in the 1990s and reallocation and labor productivity dynamics.

Types of Data Available: Economic Data

- Business Register Data
  - Standard Statistical Establishment List (SSEL), COMPSTAT-SSEL Bridge
  - Form 5500 Bridge File
  - Longitudinal Business Database, Integrated Longitudinal Business Database
  - Ownership Change Database

- Sample Research
  - Quarterly Review of Economics and Finance 2011

  The authors combine COMPSTAT and the Annual Capital Expenditure Survey to examine the relationship between firm (investment) diversity and firm value. Firms with greater diversity inefficiently allocate capital for equipment, but they efficiently allocate capital for structures.

Types of Data Available: Economic Data

- Firm Surveys and Data
  - Business Research & Development and Innovation Survey
  - Survey of Industrial Research & Development (NSF)

- Sample Research
  - "Why are Black-owned Businesses Less Successful than White-owned Business?..." Fairlie and Robb

  Using the Survey of Business Owners, the authors examine why African American-owned businesses lag substantially behind white-owned businesses in sales, profits, employment, and survival. The lack of prior work experience in a family business among black business owners, limiting their acquisition of general and specific business human capital, negatively affects black business outcomes.

Types of Data Available: Economic Data

- Newest Firm Survey Data: Kauffman Firm Survey
  - On-going panel study of 4,928 businesses founded in 2004
    - Firms were selected randomly from the Dun and Bradstreet database
    - Businesses with employment focused in research and development were oversampled
  - Non-restricted data
    - Firm characteristics: industry, # employees, types of customers
    - Owner characteristics: educ., age, race, gender, hours, experience
    - Detailed financial info: equity/debt financing, balance sheet, income
    - Details on employees: benefits offered, task/work structure
    - Innovation, R&D strategies, product/service, intellectual property
  - Restricted data
    - 5-digit NAICS, zip code geography, continuous variables
    - Ability to link with other Census datasets
Types of Data Available: Economic Data

• Transactions Data
  • Foreign Exports and Imports Trade Data (Customs Bureau)
  • Commodity Flow Survey (Dept. of Transportation)

• Sample Research
  • "Multi-product Firms and Trade Liberalization"
    Bernard et al., QJE (forthcoming)
  Paper uses the Exports and Imports Trade Data and the Census of Manufactures to
  confirm predictions from a theoretical model. Specifically, trade liberalization
  fosters productivity growth within and across firms and in aggregate by inducing firms to
  shed marginally productive products and forcing the least-productive firms to exit.
  Though exporters produce a smaller range of products after liberalization, they increase
  the share of products sold abroad as well as exports per product.

Types of Data Available: Demographic Data

• Census Data
  • Decennial Census
  • American Community Survey
  • Current Population Survey (with BLS)
  • National Longitudinal Survey (with BLS)
  • Survey of Income and Program Participation

• Key Features
  • No top-coding of income and wealth
  • Fine geographic detail (block group)
  • Consistent Personal Identification Keys (PIKs)
    • => able to link to other Census data

Types of Data Available: Demographic Data

• American Housing Survey
  • Agency Sponsor: Housing and Urban Development
  • Started in 1973, same sample of housing units since 1985
  • Housing units in national survey surveyed every 2 years
  • Housing units in metro survey surveyed every 6 years

• National Crime Victimization Survey
  • Agency Sponsor: Bureau of Justice
  • Survey of housing units about personal victimization & property crime
    • unit in sample for 3 years, interviewed 7 times

Types of Data Available: Demographic Data

• Sample Research
  • This paper sheds new light on the forces that drive observed segregation patterns. Making
    use of restricted micro-Census data from the San Francisco Bay Area and a new
    measurement framework, it assesses the extent to which the correlation of race with other
    household characteristics, such as income, education and immigration status, can explain
    a significant portion of observed racial segregation.

  • "Location, Location, Location: The 3L approach to House Price Determination" Zabel and Kiel, Journal of Housing
    Economics (2008)
  • This study creates a unique data set based on data from the American Housing Survey
    (AHS) consisting of small clusters of housing units with information on their housing
    characteristics and resident characteristics that is merged with census tract-level attributes.
  • We find that all three levels of location (MSA, town, and street) are significant when
    estimating the house price hedonic equation.

Types of Data Available: Demographic Data

• Details on the SIPP
  • Nationally-representative longitudinal household survey
    designed to obtain detailed information on individuals’
    income and participation in public programs
    • households are interviewed every four months within a panel
    • length of panel range from between 2 1/2 to 4 years
    • panels began in 1984; most recent one is 2008-2012
  • Core set of questions plus topical modules
    • health status and disability, wealth, public program eligibility,
      migration history, childcare, etc.
  • Advantage of using SIPP in an RDC
    • geography is measured down at the census tract level (sub-county)
    • less top-coding or censoring of variables, such as earnings
    • availability of more confidential variables, such as date
      of birth.

• Sample Research
  • Notable advantage of using restricted version of SIPP is
    the ability to link individuals with other Census data sets
    • can examine subsequent employment outcomes of those receiving
      welfare or job training even after the SIPP panel has ended
    • can compare characteristics of the reported survey job with details
      of that job found in other Census administrative data

  • "Comparing Measures of Earnings Instability Based on
  • The authors compare earnings levels and jobs in the SIPP and administrative data
    (LEHD) to identify differences in measured earnings instability. Preliminary findings
    suggest that much of the difference in earnings instability across the two data sets
    comes from the lower tail of the earnings distribution. There is systematic difference in
    how workers' in the lower tail of the earnings distribution report their earnings
    compared with reporting behavior of workers higher up in the earnings
    distribution.
Types of Data Available: LEHD

- **Longitudinal Employer Household Dynamics**
  - State level Unemployment Insurance data that are used to
    consult the publicly available BDD & QCRXW
  - Consists of several separate data files that provide
    characteristics of every employer/worker combination in
    some states since 1989, most states since 2000

- **Sample Research**
  - "Increased Accountability, Teachers' Effort, and
    Moonlighting" Champion (Stanford), Nov. 2010
    This paper uses the LEHD to investigate moonlighting behavior of teachers. This analysis
    reveals a substantial reduction in public school teacher moonlighting for teachers in
    schools exposed to performance pressure from a state accountability system. A reasonable
    interpretation of this evidence is that moonlighting is a good indicator of slack off effort
    and that accountability systems do actually create incentives for teachers to make effort in their schools.

(http://lehdmap.did.census.gov/)

Types of Data Available: Health Data

- **National Center for Health Statistics (NCHS)**
  - National health status surveys
  - National healthcare (provission) surveys
  - Vital Statistics data
  - Linked data (across NCHS surveys, SSA data, EPA data)

- **Agency for Healthcare Research and Quality (AHRQ)**
  - Medical Expenditure Panel Survey (MEPS)
    - panel study of households (health status, healthcare utilization,
      expenditures, insurance coverage)
    - follow up surveys of providers & pharmacists (confirm household data)
    - follow up with employers (data on insurance options)
    - links with NHIS (from NCHS)
  - Only household & provider data available publicly

Research Published in Non-economic Journals

- "The Effects of Social Capital on White, Korean, Mexican, and Black Business Owners' Earnings in the U.S.” Valdez
  Traditional sociological approaches rooted in social capital arguments emphasize the
  primacy of group membership in facilitating entrepreneurial participation, which is
  associated with economic success. Using non-public characteristics in the Survey of
  Business Owners, this paper suggests that social capital facilitates entrepreneurial
  participation; however, it is human and capital that are essential for economic
  success.

- "Assessing Crime as a Problem...” Hipp
  Crime and Delinquency (Forthcoming)
  Using the American Housing Survey, this study examines the relationship between
  official crime rates in census tracts and resident perceptions of crime, and finds that
  tract violent crime is the strongest predictor of residents’ perception of crime.
The Nuts & Bolts of Doing Research in a RDC

- Contact the RDC Administrator
  - Census employee responsible for submitting proposals
  - Atlanta’s Administrator: Melissa Barzahf
  - Cookies (preferably chocolate chip)
- Access and fees (project based)
  - No cost for Consortium Member researchers (& co-authors)
  - Non members: $20,000/project disclosure fee
- Develop proposal
  - Scientific merit (similar to any other grant proposal)
  - Demonstrate need for restricted data, feasibility
  - Benefit to Census (legal requirement of CIPSEA)
    - statistical purpose for access to restricted data
    - administrator provides lots of guidance here

The RDC Advantage

- Exclusivity and expanded research questions
  - Only 13 RDC labs across the country
  - however, can easily work with colleagues close to other RDCs
  - Detailed micro data expands research agenda
- The Consortium funds 1 and 1/2 RAs for RDC use
- Taylor-made funding opportunities through NSF
  - NSF-Census Research Network (NCRN), Fall 2010
    - first of several expected joint calls for proposals from Census and NSF
  - Science of Science and Innovation Policy (SciSIP), Sept 2011
    - will fund research that “develops methodologies to analyze science and technology data, and to convey the information to a variety of audiences.”

Important Web Sites

- Atlanta Census Research Data Center
  - http://aysps.gsu.edu/acrdc/
- Center for Economic Studies, U.S. Census
  - http://www.census.gov/ces/
- NCHS Research Data Center
  - http://www.cdc.gov/rdc/
- MEPS: Restricted Data Files Available at Data Centers
  - http://www.meps.ahrq.gov/mepsweb/data_stats/onsite_datacenter.jsp