

# *Lecture II*

## *Micro Data: A Helicopter Tour*

**Gianluca Violante**

New York University

Quantitative Macroeconomics

## Micro data and macro models

---

- Once upon a time there was the **representative agent...**
  - ▶ We have aggregation theorems (but require strong assumptions, typically complete markets)
  - ▶ Still used when question is not about redistribution and trade
  - ▶ Two-agent models a common framework these days (e.g., spender-forward looking or borrower-saver)
- Now, heterogeneous-agent (HA) incomplete-market models are a workhorse of macro
- Micro data essential for parameterization: want **to match heterogeneity in the model to that we observe in the data**
- Micro data + HA model: **more direct mapping between data and model parameters** compared to the RA

# Browning-Hansen-Heckman

---

- [Micro data and GE Models, Handbook of Macroeconomics, 1999](#)
- *The main goal of this essay is to foster the process of reattachment. Macro general equilibrium models provide a framework within which micro empirical research can be fruitfully conducted. At the same time, dynamic general equilibrium theory will be greatly enriched if it incorporates the insights of the micro empirical literature. **The micro foundations of macroeconomics are more fruitfully built on models restructured to incorporate microeconomic evidence.** This essay explores three challenges for closing the gap between empirical microeconomics and dynamic macroeconomic theory.*
- New Handbook of Macro in preparation...

<https://bfi.uchicago.edu/events/conference-handbook-macroeconomics-volume-2>

<http://www.hoover.org/events/conference-handbook-macroeconomics-vol-2>

# Browning-Hansen-Heckman

---

- Challenge I: Heterogeneity
- *Any careful reading of the empirical microeconomics literature on consumption, saving and labor supply reveals quantitatively important heterogeneity in agent preferences, in constraints, in dimensions of labor supply and skill, and in human capital accumulation processes. Modelling heterogeneity is required to calibrate dynamic models to microeconomic evidence.*

# Browning-Hansen-Heckman

---

- Challenge II: Uncertainty
- *Modern macroeconomics is based on models of uncertainty. Aggregating earnings across members in a household and across income types may create a disconnect between uncertainty as measured by innovations in time series processes of earnings and income equations and actual innovations in information. Government or interfamily transfers provide insurance that should be accounted for. Alternative risk components such as risks from changes in health, risks from unemployment and job termination, and risks from changes in family structure, have different degrees of predictability and are difficult to quantify. Measuring the true components of both micro and macro uncertainty and distinguishing them from measurement error and model misspecification remains an empirical challenge that is just beginning to be confronted.*

# Browning-Hansen-Heckman

---

- Challenge III: Synthesis
- *Synthesizing evidence across micro studies is not a straightforward task. Different microeconomic studies make different assumptions, often implicit, about the economic environments in which agents make their decisions. They condition on different variables and produce parameters with different economic interpretations. A parameter that is valid for a model in one economic environment cannot be uncritically applied to a model embedded in a different economic environment. Different general equilibrium models make different assumptions and require different parameters, many of which have never been estimated in the micro literature.*
- E.g., literature on micro vs macro labor supply elasticities

# Current Population Survey (CPS)

---

- Primary source of labor force statistics in the US. Conducted for more than 50 years.
- 100,000 households interviewed monthly. Cross-section with a short (2-year) rotating panel dimension.
- March Files (ASEC) the most comprehensive
- Detailed demographic, income, and employment info
- Special files, eg., Displaced Workers, Fertility, Computer Use
- **Since 2010: American Community Survey: 30 times larger. It allows for more geographical detail**

## Panel Study of Income Dynamics (PSID)

---

- Longitudinal study of a representative sample of U.S. individuals and their family units.
- Starting with a national sample of 5,000 U.S. households in 1968, the PSID reinterviewed individuals from those households every year until 1997, and every other year since that time.
- New households are added to the sample as the children of the panel families grow older and form family units of their own.
- It can be used to study intergenerational issues, and **estimate income processes**
- Contains also **wealth supplements** and, since 1999, good coverage of **consumption** expenditures (food, housing, transportation, health-care, education, utilities, child care)



# National Longitudinal Study of Youths (NLSY)

---

- NLSY79: longitudinal study that began in 1979 with 13,000 men and women ages 14-21, and has interviewed this cohort every year until 1994, and every other year since then.
- NLSY97: in this case, the study began in 1997 with 9,000 youths who were 12 to 16 years old in 1996.
- Unlike the PSID, there is only one cohort observed, thus **not representative of US economy**
- Survey designed to document transition from school to work and into adulthood, and collects extensive information about youths' labor market behavior and educational experiences over time.
- In the last five waves, enhanced consumption expenditures data

# Survey of Income and Program Participation (SIPP)

---

- Panel, starts in 1991, organized in waves: 1991-1993, 1996-2000, 2001-03, 2004-07, 2008-11, 2014-
- Every four months, each respondent answers a questionnaire about each of the past four months, hence, the data appears monthly but there is less stress on the respondent. **However, there is a recall bias.**
- Reflecting its origins to study government assistance programs, it contains many questions regarding the use of government services and government transfers.
- Used to study the effects of the Clinton's welfare reform (PRWORA) of 1996
- It contains data on assets and liabilities (especially good to study the bottom tail)

# Health and Retirement Survey (HRS)

---

- Longitudinal study that surveys a representative sample of more than 26,000 Americans **over the age of 50** every two years.
- First survey in 1992
- Explores the changes in labor force participation and the health transitions that individuals undergo toward the end of their work lives and in the years that follow.
- Collects information about income, work, assets, pension plans, health insurance, disability, physical health and functioning, cognitive functioning, and health care expenditures.
- Unique source of data to study bequests, retirement choices, and saving behavior of the elderly

# Consumer Expenditure Survey (CEX)

---

- Cross-sectional survey with information on the **consumption expenditures** of American households. Sample size: 10,000-15,000
- **Quarterly Interview panel survey** in which each consumer unit in the sample is interviewed every three months over a 15-month period. Good to capture large item purchases (e.g., durables).
- **Diary or record keeping survey** completed by the sample consumer units for two consecutive one-week periods. Good to capture frequent expenditures (e.g., groceries, gas)
- Asset data available but poor quality (last question of 2-hour survey: “fatigue bias”)
- Problem: averages not aligned with NIPA (see Aguiar-Bils, 2012)

# Survey of Consumer Finances (SCF)

---

- Conducted every three years to provide detailed information on the **finances of U.S. families**.
- It is a survey of the balance sheet (detailed information on portfolio of asset and liabilities), income, and other demographic characteristics of U.S. families.
- Oversamples the **rich households**: unique in the US
- Data are available for 1962-1963, 1983, 1986, 1989, 1992, 1995, 1998, 2001, 2004, **2007-09 (panel)**, 2010, 2013

## Other household data sets

---

- **American Time Use Survey**: since 2003, measures how people divide their time among life's activities. Earlier surveys: 1965, 1975, 1985, 1992
- **American Housing Survey**: wide range of housing subjects (size, features, mortgage and HELOCs, home value, etc)
- **Medical Expenditure Panel Survey**: most complete source of data on the cost and use of health care, health status, and health insurance coverage. Panel: can be used to estimate **stochastic process for health and medical expenditures**.
- **Statistics of Income**: income data from tax returns compiled and aggregated by the IRS
- **National Center for Education Statistics**: primary federal entity for collecting and analyzing data related to education.

## Restricted-access household-level data sets

---

- **Kilts-Nielsen data:** Two sources of data:
  - Retail scanner data** of weekly pricing and sales volume from more than 90 participating retail chains across all US markets since 2006.
  - Consumer panel data:** longitudinal annual data beginning in 2004 that track a panel of 50,000 US households and their purchases of retail goods.
- **Corelogic:** data on U.S. real estate, mortgage, consumer credit, bankruptcy by ZIP code. **Zillow:** similar data, but free!
- **Equifax:** household debt by ZIP code. **FRBNY Consumer credit panel:** free Equifax data until 2010

## Public firm-level data sets

---

- **Compustat**: detailed information on public companies going back to 1962. Stock prices, capital structure, executive compensation, credit rating, R&D.
- **Business Dynamics Statistics**: annual measures of business dynamics (e.g., job creation and destruction, establishment births and deaths) for the economy and aggregated by establishment and firm characteristics.
- **County Business Patterns**: measures of economic activity at high level of geographical disaggregation
- **Quarterly Workforce Indicators**: a set of economic indicators including employment, job creation, earnings, and employment flows. Reported by firm characteristics (geography, industry, age, size) and by worker demographics (sex, age, education, race).



## Restricted-access firm-level data sets

---

- **Census data:** Manuf., Construction, FIRE. Every 5 years
- **Annual Survey of Manufacturing:** sample estimates of statistics for all manufacturing establishments with one or more paid employee.
- **Longitudinal Business Database:** longitudinally linked data for establishments in all sectors
- **Longitudinal Employer-Household Dynamics:** combines different administrative sources (primarily state-level UI earnings data and firm-level surveys). Firm and worker information combined to create quarterly earnings history data for individual workers.

Access to these data occurs at one of the Census Research Data Centers (RDCs). There is one in NYC. **Factor in one year to get access (need a very well defined project)**