Using Households Microdata to Understand Households Decisions

Luigi Guiso (EIEF)

Prepared for the Third IMF Statistical Forum

Deutsche Bundesbank

Frankfurt, November 19-20 2015

Lot of progress in understanding households behavior

- What type of conceptual framework are most promising for understanding their behavior
  - Do they behave according to classical life cycle models?
  - Do they conform to Merton's principle in their investment decisions?
  - Do they select mortgage type as risk-management models predict?
- What are the macroeconomic implications of they micro choices (and inform policy).
  - Savings dynamics of a country
  - Predicting bankruptcies and aggregate financial fragility

Successful outcomes

- The study of the aggregate implications of households microeconomic behavior is one of the most successful area of research over past 25 years
- Not surprising, this year Nobel prize was awarded to Angus Deaton - the founding father of this area
- Key ingredients: 1) Heterogeneity, 2) Aggregation
- No room for representative agents. As Deaton puts it:
  - "Representative agents have two great failings: they know too much, and they live too long." Aggregate of individuals is not likely to behave like the single individual of the textbook"
- Future progress is most likely to come when aggregation is taken seriously, and when macroeconomic questions are addressed in a way that uses the increasingly plentiful and informative microeconomic data

A two way interaction

1. Study of microeconomic household behavior made progress thanks to impressive greater availability of micro data
2. Past 20 years massive gathering of households surveys
    1. Lower collection costs
    2. Access to internet surveys
    3. Increased commercial value of microeconomic information
    4. Increased policy value of redistributive effects of macro policies
3. Greater availability has revived microeconomic research in households behavior and made this research available for macro purposes

The administrative records revolution

1. We are probably now at a new turning point
2. Post war period: the age of time series
3. Past 25 years: the age of Household surveys
4. We are now starting a new age: the age of administrative records
5. The new future is the availability of large administrative records that:
    1. Combine different datasets each gathered for a specific administrative purpose
    2. Can inform on many dimensions of the household and its components – economic and non-economic

Drivers of the revolution

1. Upward trend in the use of administrative records for research in all countries
    1. Matched employer-employees data from SS
        1. Countries: Italy, Germany, France, Portugal among others
        2. Examples: within firm insurance, human capital, unemployment transitions etc.
    2. Matched households loans/portfolios and banks datasets
        1. Countries: Italy, Germany, Argentina etc.
        2. Examples: lending patterns, portfolio allocations, quality of financial advice
2. Expansion of types of administrative records
    1. Tax records, Social security, Military, Education, banks and financial institutions, Health records, Genetic records
3. Increase in the "mergibility" of administrative records
    1. Single identifier in many countries: SSN, tax code etc.
4. Can get comprehensive picture of the households

The Nordic countries: a benchmark: 1

- They are precursors, they are ahead of others
  - Records cover the population
  - Can merge essentially any sort of data
  - Members of the family (parents, kids, brothers can be linked/connected)
  - Coverage is very broad: from income to assets, to firms and their balance sheets, to mortality, health, casualties, genetics, lotteries and of course demographics

The Nordic countries: a benchmark: 2

- Something missing? Yes
  - what is in people heads (beliefs, beliefs) opinions and preferences
  - What is not of interest for the administration that gathered the data (often demographic info is scarce)
  - Not a problem if you can merge *all* datasets; but this is cumbersome (even in Nordic countries)
  - Surveys give comprehensive picture

Why are administrative records important? I

- **They can help solve at least four problems that plague HS**
  - Measurement error
  - Attrition
  - Sample size and representativeness (non response)
  - Tails (the very poor, the very rich)
  - Under-reporting

Why are administrative records important? II

- These problems:
- Present almost always (though to various degrees) in all surveys
- For some issues they can be critical and solution in surveys impossible
- Some of them have grown in importance and surveys quality has deteriorated (Meyer et al 2015, *Household Surveys in Crisis JEP*)
  - Upward trend in both *unit* and *item* non-response – perhaps households overburdened)

Examples

- Give five/six examples:

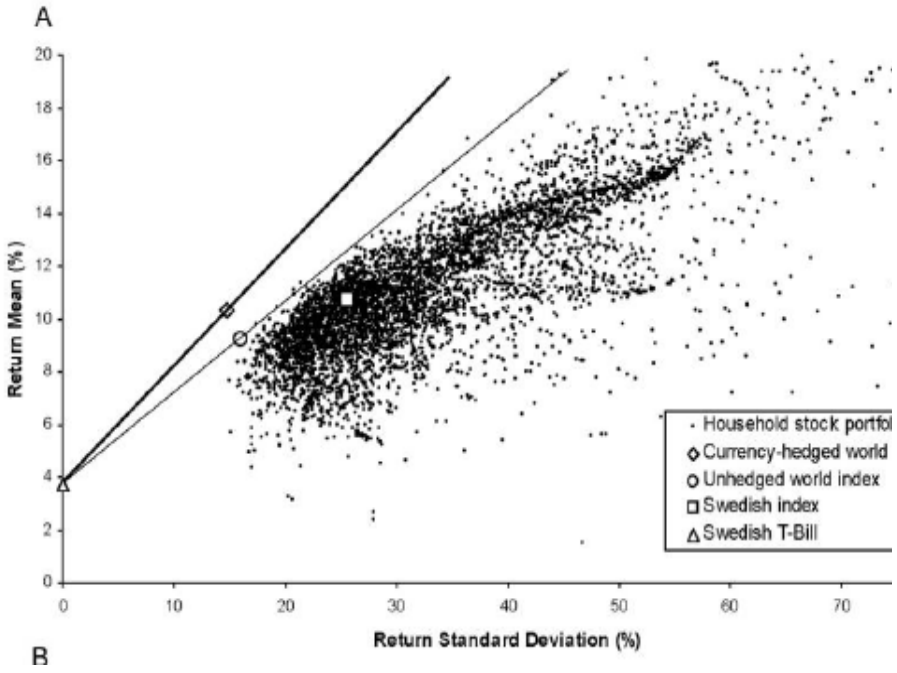
1. Portfolio diversification (Calvet et al)
2. Wealth inequality: measurement
3. Wealth inequality: heterogeneity in income or heterogeneity in returns?
4. The causal effect of wealth effects (Cesarini)
5. Genetics: the size problem (Cesarini et al)
6. Life cycle accumulation and portfolio allocation

Ex. 1: Portfolio Diversification (Calvet et al)

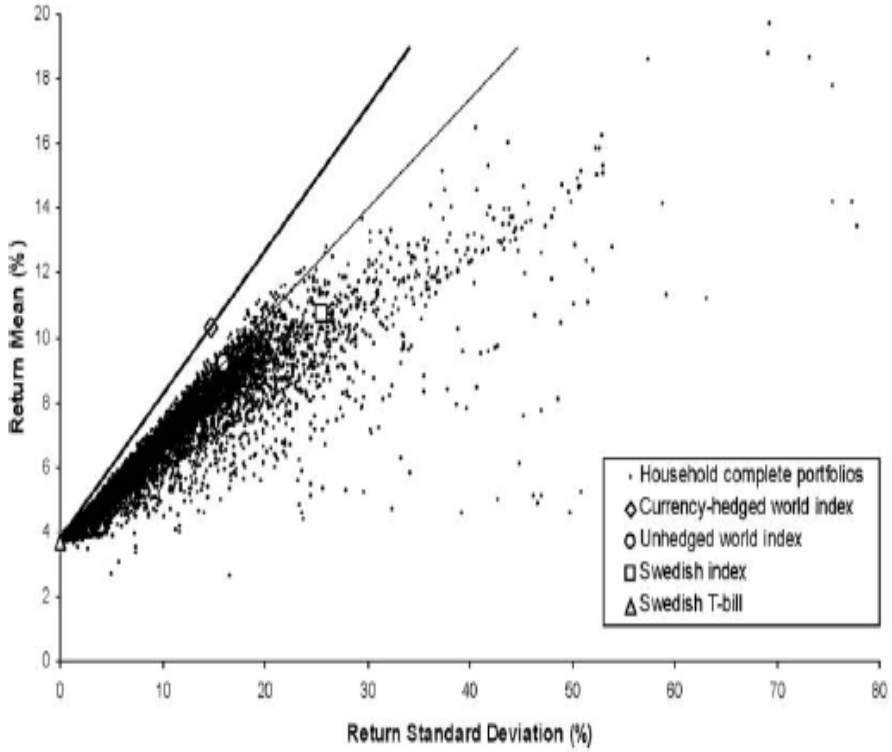
- Hard to measure, requires knowledge of
  1. All single risky (and safe) assets in one household portfolio
  2. Time series of returns on each risky asset to construct the portfolio return and check how it differs from the market portfolio
- These data are not available in surveys (ISIN code impossible to collect
- But available in Sweden for the whole population

Ex. 1: Findings

Stock Portfolio



Complete portfolio



Ex. 2 - Wealth inequality: measurement I

- What happens at the top of the wealth distribution? Is there a trend?
- Topical issue (after Piketty)
- But wealth at the top hard to measure:
  - Surveys miss the top and wealth is very concentrated
  - Wealth not generally collected (except by Forbes 500)

Ex. 2 - Wealth inequality: measurement II

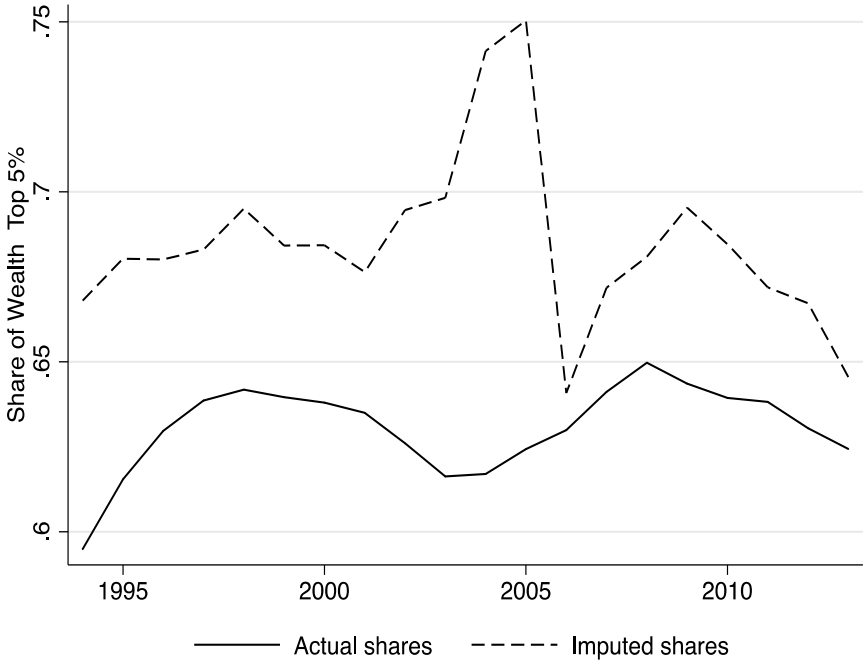
- Alternative (Saez&Zucman, 2015, *QJE*): infer wealth from income from capital from tax records
  - Available for the whole population , for many years: compute

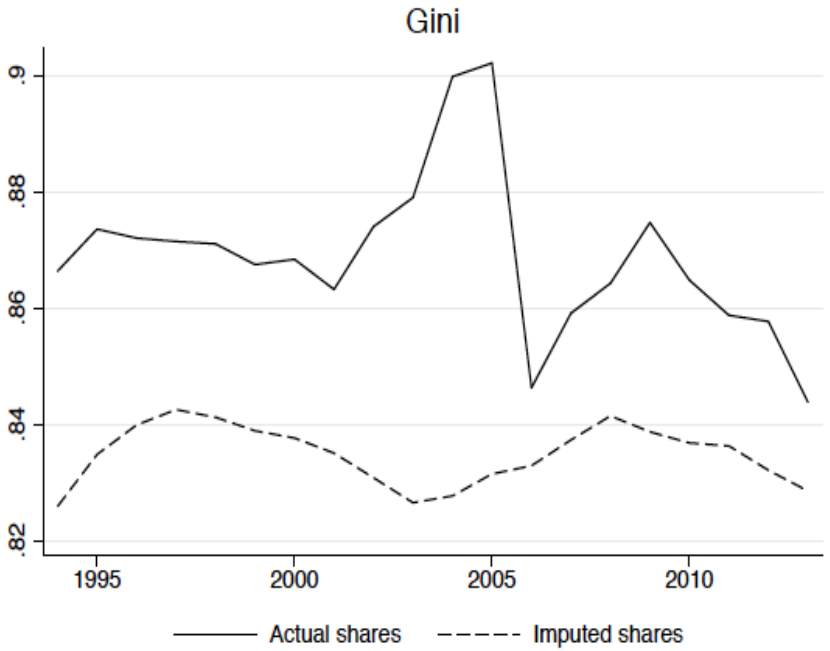
$$w_i^I = \frac{r_i w_i}{\bar{r}}$$

Nice but problems:

- What happens to estimated inequality if returns on wealth are heterogeneous across individuals
- To validate method need tax records and wealth data for many years: Norway has both

Ex. 2 – Some findings (Fagereng, Guiso, Pistaferri AER, 2016)

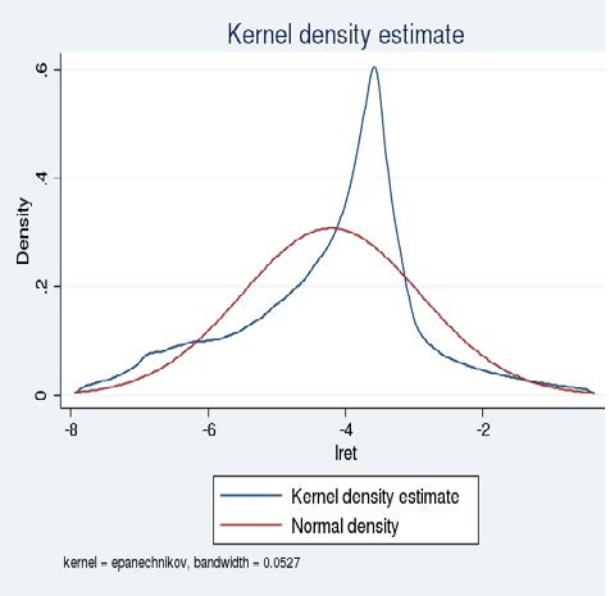




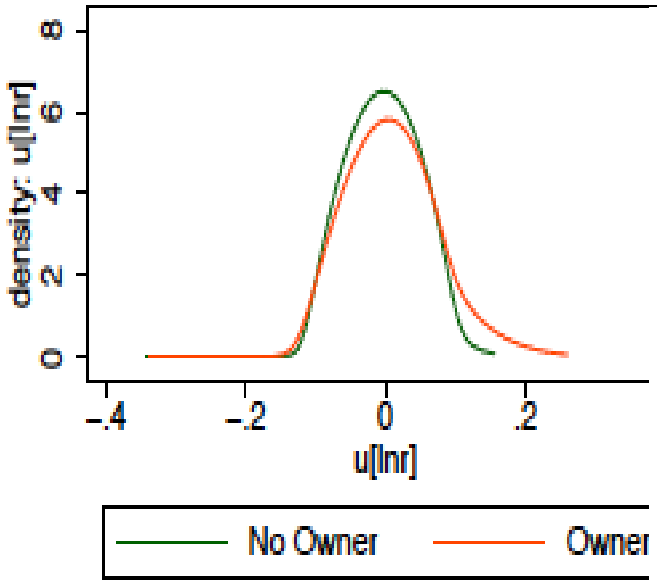
Ex. 3: Wealth inequality: heterogeneity in income or returns to wealth?

- Previous data useful for a bigger debate: what explains wealth inequality, particularly the tails.
- Heterogeneity in income realizations and time discount can help, but :
  - Hard to explain the tail with realistic differences in patience and in human capital
- Persistent heterogeneity in returns to wealth can explain the tail (Benhabib and Bisin, 2011)
- But how much heterogeneity and how much persistence is there?

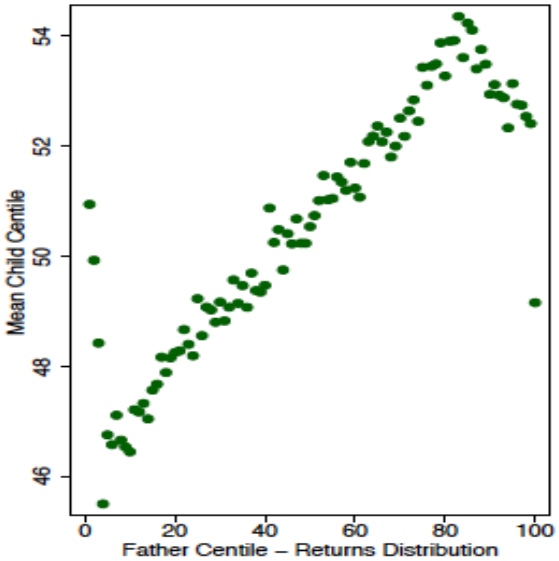
Ex. 3: Some findings (FGP, 2016)



Returns very heterogeneous



Heterogeneity is persistent *within* generations



Persistent also *across* generations but mean reversion

Ex. 4: The causal effect of wealth effects I

- Variation in wealth in principle key for several critical variables
  - Consumption, Labor supply, Participation in assets markets

1. But very hard to identify: ideally would like to observe random allocations of substantial sums of wealth to individuals
2. But that is unfeasible
3. Exception are large lotteries, but need population data

Ex. 4: The causal effect of wealth effects II

- Denmark has both
  - Several years of lottery with data on winners *and participants*: ID, cost of ticket, money won
  - Can be linked to administrative records on portfolio holdings, labor supply, health, mortality and fertility
  - Wealth shocks are random among participants
  - => Can identify causal effect of wealth shocks
- Findings (Cesarini et al) : wealth effects are much smaller than typically believed both on labor supply, participation in (risky) assets markets, no effect on health and mortality

Ex. 5: Genetics: the size problem (Cesarini, 2014)

- Genetics and economics is the topic of the new millennium
  1. Massive progress in Genetics thanks to the sequencing of DNA
  2. We can thus address important questions in health economics and behavioral genetics/economics
  3. But cannot be done using surveys with information on behavioral outcomes and genetic material ail behavior

Ex. 5: Genetics: the size problem (Cesarini, 2014)

- Size problem: association between an economic trait and a genetic marker is very plausibly small=> genetic studies can confidently be conducted only using very large samples.
- Large samples are also needed to identify the weights used to aggregate different Single-Nucleotide-Polimorphisms and conduct what are called genome-wide association studies
- Denmark allows this
  - Genetic data available for whole populations
  - Can be merged with behavioural variables

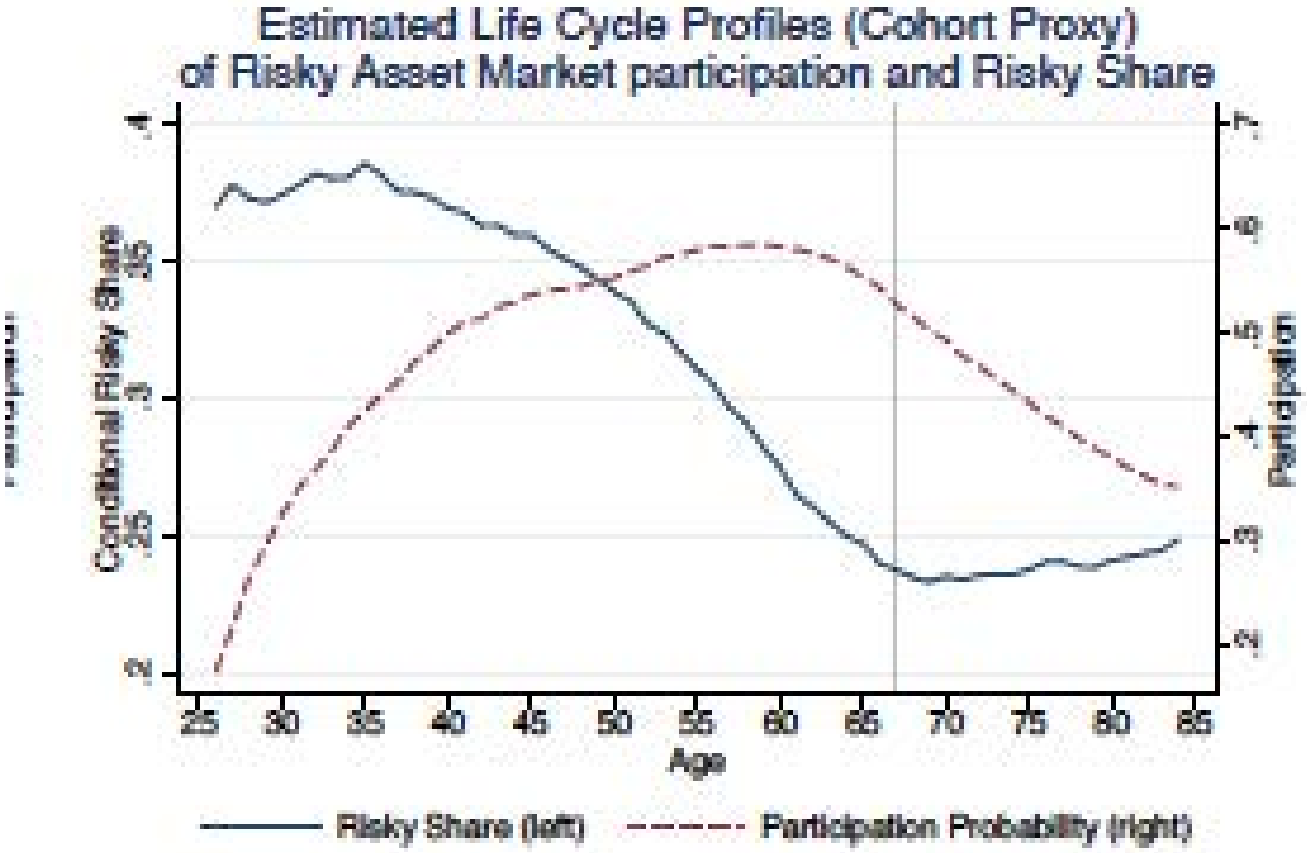
Ex. 6: Life cycle portfolio choice

1. How do people allocate their portfolio over the life cycle? Do they rebalance as they age?
2. Theory and practitioners: sell stocks as you age
3. Hard to find evidence in surveys: risky share profile essentially flat (participation hump shaped)
4. But survey data are problematic : reporting error and under-reporting possibly correlate with age

Ex. 6: Life cycle portfolio choice

1. Administrative records from a single bank/intermediary/pension fund not a solution
2. Need comprehensive data covering all sources of wealth for many years
3. Norway wealth data fit the requirement
   - Long panel
   - All assets, details about the portfolio
   - No measurement error

Ex. 6: Some findings (Fagereng, Gottlieb, Guiso 2016 JF)



Conclusions

- A new data age has begun: the age of administrative records
- Many statistical advantages
  - No/little measurement error
  - Very large coverage of tails
  - No selection and attrition
  - No collection costs
- Their exploitation requires some organizational capital: e. g. subjects need an ID at birth, linked to their families
- Is this the end of surveys? Not necessarily: at lot is collected administratively but not all
- Not every where: not in developing countries
- The two can be usefully complemented