Register-based data on firms – International perspectives on research potential and availability

Tor Eriksson
Dept of Economics & Center for Corporate Performance,
Aarhus School of Business

Last 20-25 years

- Huge changes
- Substantial shift from macro level (country, industry), typically time series data to micro data (individuals, households, firms), first as cross-sections, later as panels
- Tremendous development of econometric techniques and software to analyse the new data

Background

- U S Negative Income Tax experiments
- PSID and others. Later NLSY,HRS,…
- In Europe: Level of Living Surveys
- Study of firms and their employees relied data set up for other purposes. Business statistics, Labour Force Surveys.
- Employers and employees lived separate lives
- Tax reforms => focus on labour supply
- A decade of studies of supply behaviour

The eighties

- Stafford (1986), Leontief (1985)
- Less attention paid to the demand side, to firms
- IO turned into game theory, information economics
- Distance between theoretical concepts and the variables observed in readily available data sets grew

Return of firms

- Gradually firms and firm data returned
- Firm panels
- Linked employer-employee data
- IO: firm demography, analysis of performance differences, R&D
- Growth of the subfield of corporate finance
- Labour: what goes on within firms and organisations (personnel economics)?

Register based data

- North America: mainly survey data; in US relatively few research data bases. Canada many survey-based data sets
- (Especially Northern) Europe: register based data and hybrid data; more data sources
- Having good data is not enough!

Advantages, disadvantages

- Register data can provides sources of truly independent variation in explanatory variables (see Raaum & Roed, 2005 for examples)
  - Due to public regulation (e.g. calendar time)
  - Exogenous shifts in policies, practices
  - Calendar itself exogenous to the subject
- Much richer analysis: heterogeneity, dynamics

Drawbacks

- General equilibrium effects – Achilles heal
- Collected for administrative purposes
- Sensitive nature => hampers free exchange of data

Dearth of firm data

- Less data on firms than individuals, because:
- A ”firm” is not as clearly defined as an individual (especially in longitudinal settings)
- Who should answer a questionnaire on the behalf of a firm?
- Firms much more reluctant to answer requests for data

Moreover

- Small and young firms particularly problematic. Less reliable and low quality information.
- Data produced for other than research purposes reflect differences in corporate law, accounting systems => hence, hard to do cross-country comparisons.
- Growing problem: internationalisation of firms, whereas statistical agencies and their products remain national. =>”Misleading statistics” and not helpful in answering one of the key questions of our time.

Motivating high costs

- Construction of panels and linked e-e data in particular (giant jigsaw puzzles) are costly
- However, data collection costs are sunk
- Additional costs: tools for handling the data and adapting for research purposes
- Not for research only
- Policy analysis. Monitoring the economy
- Best: statistical products

Firm data sets

- Some main fields of application:
- Firm demography (exit, entry, survival); productivity, R&D and innovation, outsourcing, government regulation, entrepreneurship, multinationals, corporate finance, corporate governance
- All highly relevant wrt business and industrial policy
- Less interesting from an employment policy perspective
- Employment dynamics; but requires much more detailed data than what normally are available

Cross-country analysis

- Not much using firm data
- And yet, a lot could be learnt from cross-country-comparisons
- Attempts to build have not been very successful; EU in mid-90s; OECD some progress in connection with its productivity project
- Private initiative: Amadeus produced by Bureau van Dijk in Brussels: employment, balance sheet info, industry, location. Strongly biased against small firms.

Pan-European firm data sets

- Some first steps
- PIEP (Pay Inequality and Economic Performance) project
- Made use of 1995 European Structure of Earnings Survey for Belgium, Denmark, Ireland, Italy, Spain and UK. Can be supplemented with firm performance data from EUROSTAT's Structure of Business Statistics
- Pioneer in creating a system of secure remote access to pan-European microdata
- 2002 ESES is available and many more countries have agreed to allow remote access
- New EU project tries to take some further steps

Linked e-e data

- Why?
- Because we are (should be) interested in activities on **both** sides of labour markets
- Both for more academic research: ”understanding functioning of labour markets” and for policy analysis.
- Partial analysis may paint a flawed picture (as when neglecting the supply side)

Mainly known for:

- Job Creation - Job Destruction studies (Davis and Haltiwanger, 1999)
- Role of firm and individual effects in wage determination (Abowd, Kramarz & Margolis, 1999)
- But can/should be used for studying many other topics as well (save, labour demand)

Examples

- Effects of displacement on workers' subsequent earnings and employment history
- Event studies
- Regional labour markets. Effects of opening/closure of major plant/firm on wages/employment in firms in the local area (like Wal-Mart studies)
- Impact of tax incentives, subsidies, infrastructure investments (transportation economics)

Internal workings of firms

- Most previous work based on single firm studies a la Lazear (1992) and Baker, Gibbs and Holmström (1994)
- Recent NBER project exploits LEED from several countries (the Nordic, Germany, France, Italy, Portugal, US) to examine wage structures and mobility

NBER project

- Examples of simple questions never addressed before:
- Differences in wage structures across firms; what accounts for them?
- Is wage variation across individuals due to workers being different (high or low wage) firms, or to being in high and low wage jobs found in all firms?
- Are firms with compressed wage structures also low wage (productivity) firms?
- Are raises uniform within firms? Are differences in wage growth in the economy primarily due to firm differences?
- Does pay compression result in loss of the most able or the retention of the least able?

Surprisingly little

- Programme evaluation research does not exploit linked data
- Silent on firms
- What characterises:
  - Employers in successful programmes
  - Co-workers when policy works
- Ageing
  - As if retaining employees did not depend on where they are employed, how they are treated, who their co-workers are
- Great potential to enhance our understanding is still unexploited

Availability

- LEED for whole economy and long time periods: France, Denmark, Finland, Norway and Sweden
- Have unique identifiers of employees, establishments and enterprises
- Possible to link employees to workplaces and firms
- But also to other sources of information: other registers, surveys

Pretty good

- Portugal, Germany, US, Canada, NZ
- Not so bad: Austria. Belgium, some CEE countries
- Italy? Strange sample selection
- UK? In theory, yes. In practice?
- More? LEEDs are not built over night

And in the U.S.?

- LEHD project at the Census Bureau
- Main use is not research but statistical products, monitoring and guidance of policy (leading indicators)
- Built state by state
- For specific research projects
- Rather little content (few variables)

For what there is…

- Existing LEED have not been designed in the same way as to
  - sampling frame,
  - info content,
  - length of observation period and
  - treatment of attrition problems
- Therefore, comparative research has to work on a project basis

Accessibility

- Key to ensure cofidentiality and security of the information
- Rules change all the time
- Technology for providing high security improves continously
- Demand for LEED information increases, as well

Why in some countries, not in others?

- Some countries have tradition for keeping several registers
- Big vs small government
- Homogeneity? Thrust?
- Not much of a pattern – what the France, Portugal and the Nordics in common?

Case: Denmark

- Early eighties: Cooperation between researchers in Aarhus and Statistics Denmark to build a longitudinal labour market research (LLMR) register (individuals; focus on unemployment, policies, later wages)
- Workplaces had unique identification numbers. Also in early eighties, but within SD and mainly involving sociologists, major project to construct a linked employer-employee data set => IDA
- In mid-nineties, Aarhus researchers and SD cooperated on merging IDA with firm info and aggregating workplaces to firms

- The LLMR (1 or 5% samples) was available outside SD already in the 90s
- IDA and firms+IDA not
- Gradually access improved:
- SD branch office in Aarhus
- Next, National Center for Register Research established in Aarhus. Today, mainly psychiatric research
- In last 5 years access to register data from secured offices at universities and several large research institutes

- Use of register data has spread quickly and are now widely used
- Recent trend: merge surveys with registers (need permission from respondents and the ombudsman). Increase value added significantly. Problem: register info arrives with a considerable time lag
- Also used in ministries, the Economic Council; important for understanding their value among decision makers

- For five years the Ministry of Science, Technology & Innovation has supported register-based research in state budget with annual payments to SD and National Board of Health
- Use administered by board of leading scholars
- Prices for access and use of register have fallen significantly => demand goes up

Recent initiative

- More efficient and cost-effective use of registers (reduce overlap in purchases)
- Proposal: build a common longitudinal research database for social sciences, containing 300 most used variables in SD for the whole population
- Validate and document this, and draw general samples based on (consensus) recommendations from researchers
- In the budget proposal for next year

Concluding remarks

- Register data on firms: much richer analyses
- Still a great potential to be tapped
- Cross-country comparisons of firm panels associated with large payoffs. Common standards are needed and remote access pan-European data sets
- LEED data also for analyses of policy and ”event studies”
- Improved availability and accessibility is a long process
- More policy relevant research using them is helpful for spreading use of register data
- Can’t answer all questions. Need many tools.