Use of administrative and Business Register (BR) data in the compilation of Finnish Structural Business Statistics (SBS)

Enterprise and local KAU databases

Voorburg Group Meeting, Statistics Finland 28 September 2005

Ville Tolkki
The structure of the presentation

- Use of administrative data at Statistics Finland - general overview
- Framework of Finnish SBS
- Statistical process of SBS at Statistics Finland
  - Source data
  - Methods
  - SBS output (Enterprise and local KAU databases)
  - Process timetable
Use of administrative data at Statistics Finland, evolution (1)

- First the use of administrative sources expanded in social and demographic statistics and this was followed by the expansion in business statistics
- Modern statistical uses of administrative data dates back to 1970 population and housing census
- The 1990 population census was collected exclusively from registers
- The use of Tax Authority data in SBS 1995 and in monthly turnover and wage bills 1998
Use of administrative data at Statistics Finland, sources (2)

- 96% of data reserves in Statistics Finland comes from registers. That is, 4% is covered by direct inquiry.
- Major register sources used in Statistics Finland:
  - The population information system (population, buildings, dwellings etc.)
  - The real estate information system
  - The business information system (Trade Register, Finnish Tax Administration data and business register)
  - Other major registers: taxation, employment, pension and register of Job Applicants
Use of administrative data in Statistics Finland, population census (3)
Framework of SBS in Finland

- Historical background
  - 1974 Industrial statistics adopted UN 1968 recommendations
  - Financial Statements statistics 1974 onwards
  - The new system of SBS 1995 onwards
- SBS Regulation (58/97)
  - Legal framework (annexes 1 - 4)
  - Coverage, variables, units
- National Accounts ESA 1995
  - Practical framework
  - Goal: quality data for NA to compile National totals
- Statistical Act 2003, recommends the usage of the administrative sources in statistics production
Source data, direct inquiry (1)

- Is collected by census and covers roughly all enterprises with more than 20 employees
- About 7000 enterprises and 400 variables
- About 5000 local KAU (industries C-E) and 150 variables
  - Of which 2000 are part of multiregional or multi-industrial enterprise
  - That is, 3000 are uniregional and uni-industrial
- Accounting data: balance sheet and profit and loss account data
- Information on investments
- Details on income, expenses and personnel
Source data, Tax Authority (2)

- Is collected by census and covers all enterprises paying income taxes in Finland
- About 295 000 enterprises and 200 variables
  - Accounting data
  - Investments
- Covers also activities outside the SBS-regulation
- Small enterprises employing less than 20 persons are considered as uniregional and uni-industrial
Source data, benefits in Tax data (3)

- Improved coverage of the statistics
- Reduction of response burden (especially for small enterprises)
- Cost advantage
- More detailed presentation is possible
- Better precision
- Needs of main users (e.g. National and Regional Accounts) satisfied
Source data, defects in Tax data (4)

- Includes errors of various type
  - Need to develop and maintain an automated editing system
- The scope of information is too small for SBS needs
  - Need to develop and maintain an automated imputation system
- Limited contact to actual statistical unit
  - Collecting largest units by direct inquiry
- Limited independency of SBS
  - More intensive co-operation with register authorities needed
Source data Business Register (BR) (5)

- Defines the frame
- Gives principal activity and number of personnel for Tax data
- Covers also small number of enterprises not included in the Tax data
  - Tax data includes all enterprises that have ended their accounting period during the year
  - Enterprises that have operated but have not ended their accounting period are received from Business Register but not from Tax data (unit non-response)
  - Also some small enterprises are missing from Tax data
Furthermore, BR is the major source of background and classification information

- Activity
- Location
- Owner type
- Institutional sector etc.
Source data, coverage by number of enterprises (7)

Coverage by number of enterprises, 2003

- Total (C-O less J)
- Industry (CDE)
- Construction (F)
- Distributive trade (G)
- Services (HIK)
- Social services (MNO)

Legend:
- Direct inquiry
- Tax authority
- Unit non-response (BR)
Source data, coverage by turnover (8)

Coverage by turnover, 2003

- Total (C-O less J)
- Industry (CDE)
- Construction (F)
- Distributive trade (G)
- Services (HIK)
- Social services (MNO)

Legend:
- Direct inquiry
- Tax authority
- Unit non-response (BR)
Methods, general overview (1)

- Direct inquiry
- Tax data
- Item non-response
- Unit non-response

Variables: 400

Enterprises: 7,000, 204,000, 20,000
Methods, editing (2)

- Direct inquiry: manual editing
- Tax data: automated editing
  - Logical editing → Using linear equality based accounting identities the defects in company accounts are highlighted
  - Defects are edited using
    - Outlier detection
    - Re-scaling if error/turnover < 5%
    - Donor based imputation
      - Previous year
      - Nearest neighbour
Methods, imputation (3)

- Unit non-response: nearest neighbour imputation
  - Distance measure compiled using turnover and personnel data
- Item non-response: regression imputation
  - Typical variables are sub-items of turnover or expenses
  - (Weighted) least squares regression with automated outlier detection
  - Only one explanatory variable, the corresponding total
  - A separate model for each sub-item and each principal activity using information from the direct data collection.
  - Predicted values are re-scaled to sum up to the total.
SBS output

- Enterprise database
  - 230,000 units
  - 400 variables
- Local KAU database
  - 60,000 units
  - 150 variables
- From above databases we produce following services
  - Statistical publication
  - Internet StatFi
  - Ad hoc reports
SBS timetable

- Production base updated by the Business Register in November
  - Frame for the survey
- Questionnaires sent to the reporting units (t+2 months)
- First results concerning large enterprises (t+6 months)
- Coverage of Tax data and direct inquiry sufficient for preliminary SBS (t+6 months)
  - 71% number of enterprises and 90% turnover, 2004
  - Missing data estimated using turnover change (from BR), structural change (from SBS) and previous year data
- Closure of the survey data treatment (t+10 months)
- Business register completed, final frame (t+11 months)
- Survey, taxation and BR data combined (t+11 months)
- The SBS databases completed (t+13 months)