

It is always good to have just one number, but it is better to have more than one way to get it – Results from the project „Combined Firm Data for Germany (KombiFiD)“

European Establishment Statistics
Workshop, EESW11
12.-14.9.2011, Neuchâtel

Stefan Bender,
Anja Gruhl

Starting point (Situation in Germany)

- Linked firm data do not exist between different data producers.
- Adjustments of the data do not exist, too.

In KombiFiD: linkage of firm surveys and administrative data of different data producers in Germany for the first time.

Comparision of variables from different data producers with (nearly) the same content.

Project partners

- Federal Statistical Office (*FSO*)
- German Federal Employment Agency (*BA*) / IAB
- Deutsche Bundesbank
- Leuphana University of Lüneburg
- University of Applied Science Mainz

Support by the German Federal Ministry for
Education and Research (BMBF)

Central project idea

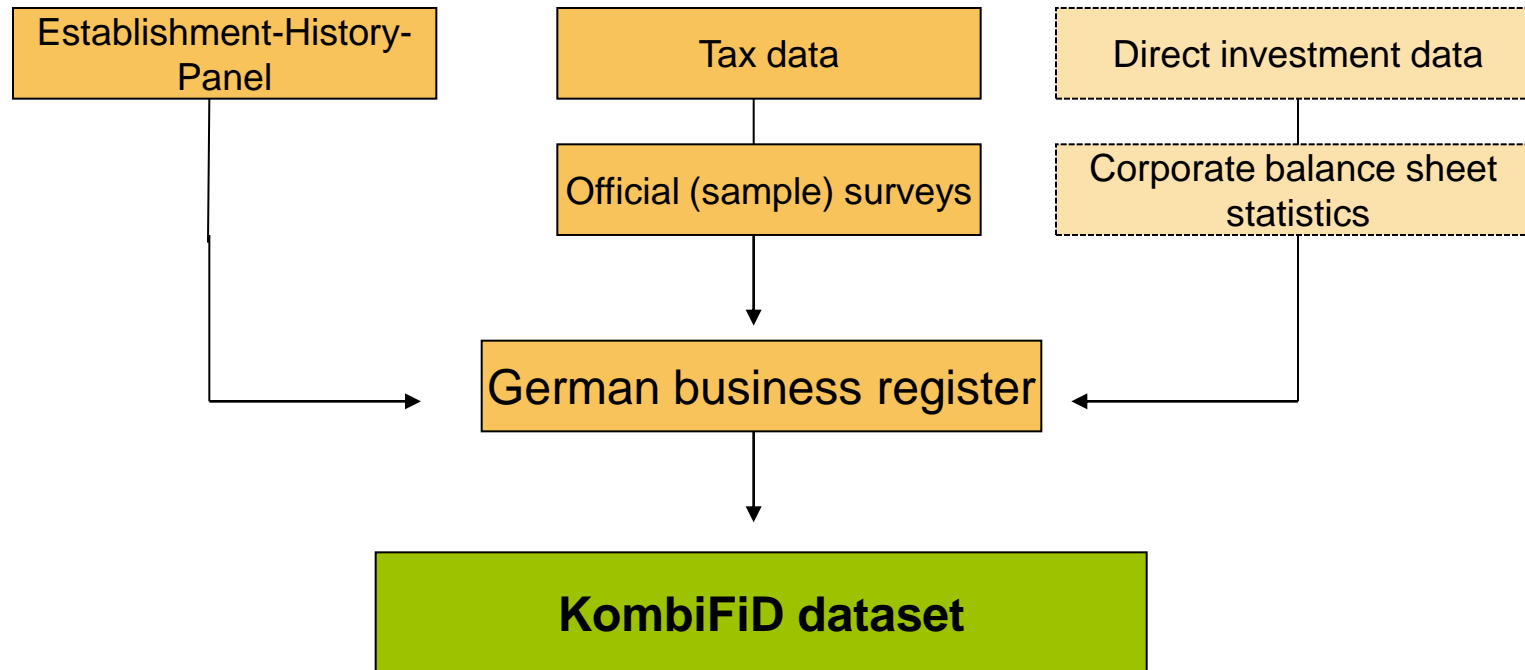
- The *German Business Register System (URS)* contains business register IDs, establishment numbers and tax numbers for all firms → *URS as masterfile*.
- A written agreement by the firms is mandatory to link the firm information of different data producers.

Sample design

- sample of 54,960 firms were asked for their permission
- 1,033 undeliverable addresses
- 30,944 firms answered (response rate: 57.4 %)
- 16,571 agreed (about 30 %)

Survey entity: firm as a legally independent unit (see European; Council Regulation No 696/93, Annex IIIa).

Data by producers within the KombiFiD project



Linkage procedure BHP/URS - URS as masterfile

- URS - recorded units:
 - economically active / inactive firms
 - firms with employees subj. to social security contributions
 - firms with taxable turnover
- Features:
 - tends to contain more establishments than BHP
 - tends to contain more firms than Turnover tax statistics

Linkage procedure BHP/URS – data cleansing I

- missing establishment numbers
- incompletely recorded firms

| Year | Number of all firms (URS) | Number of incompletely recorded firms | Number of all establishment numbers (URS) | Number of missing establishment numbers |
|------|---------------------------|---------------------------------------|---|---|
| 2005 | 16,082 | 1,739 | 27,728 | 6,619 |
| 2006 | 16,048 | 1,671 | 27,824 | 6,421 |
| 2007 | 16,014 | 1,627 | 28,109 | 6,280 |
| 2008 | 15,963 | 1,555 | 28,451 | 5,994 |

Note: Year refers to URS year. This results in time delay of 2 years with respect to BHP.

Linkage procedure BHP/URS – data cleansing II

- **Reference date:**
 - BHP: 30 June of each year
 - URS: 31 December of each year
- **different sampling frames:**
 - BHP: only establishments with employees subject to social security contributions
 - URS: also establishments without employees subject to social security contributions
- time delay of 2 years URS \Leftrightarrow BHP

Therefore not all establishment numbers can be linked.

Linkage procedure BHP/URS – Number of cases in the URS after data cleansing

| Year | Number of firms (URS) | Number of establishments (URS) | Number of OEF (URS) |
|------|--------------------------|--------------------------------------|------------------------|
| 2005 | 14,343 | 21,109 | 12,326 |
| 2006 | 14,377 | 21,403 | 12,263 |
| 2007 | 14,387 | 21,829 | 12,173 |
| 2008 | 14,408 | 22,457 | 12,112 |

Note: Year refers to URS year. This results in time delay of 2 years with respect to BHP.

Linkage procedure BHP/URS – Number of cases after linkage I

| Year | Complete URS-firms identified in the BHP | | Number of establishments in these firms | |
|------|--|----------------------------------|---|---|
| | Total | Percentage of complete URS-firms | Total | Percentage of URS-establishment numbers |
| 2003 | 13,722 | 95.7% | 17,103 | 81.0% |
| 2004 | 13,653 | 95.0% | 16,968 | 79.3% |
| 2005 | 13,580 | 94.4% | 16,978 | 77.8% |
| 2006 | 13,549 | 94.0% | 17,022 | 75.8% |

Linkage procedure BHP/URS – Number of cases after linkage II

Number of OEF identified in the BHP

| Year | Total | Percentage of all URS-OEF |
|------|--------|---------------------------|
| 2003 | 12,256 | 99.4% |
| 2004 | 12,173 | 99.3% |
| 2005 | 12,077 | 99.2% |
| 2006 | 12,019 | 99.2% |

Special features depending on the data producer

- Different data collection methods and sampling frames:
 - total population vs. samples
 - primary statistics vs. secondary statistics
 - FSO-data: data collected by the Statistical Offices of the Länder and reported to the FSO vs. data collected by the FSO-departments itself
- Different definition of industry:
 - *BA*: number of employees
 - *FSO*: main business activities

Selected data for the analysis

- One-Establishment-Firms (OEF) → “simple” structure
- pooled data set (2003-2006)
- 49,613 observations
- two combinations of (original) datasets used for the analysis:
 - service sample
 - production sample } most frequently observed within KombiFiD
- Classification of Economic Activities: 2-digit code

Combination of dataset with highest number of OEF:

■ Service sample:

- 12,254 included OEF
- included datasets:
 - BHP
 - Structural survey in services sector
 - Turnover tax statistics

■ Production sample:

- 11,468 included OEF
- included datasets:
 - BHP
 - Cost structure survey in manufacturing, mining, quarrying
 - Turnover tax statistics

Number of OEFs by dataset:

- BHP: 48,515 (97.8%)
- Turnover tax statistics: 39,000 (78.6%)
- Cost structure survey in manufacturing, mining, quarrying: 17.558 (35.4%)
- Structural survey in services sector: 15.590 (31.4%)

Number of all OEF in the KombiFiD dataset: 49.613

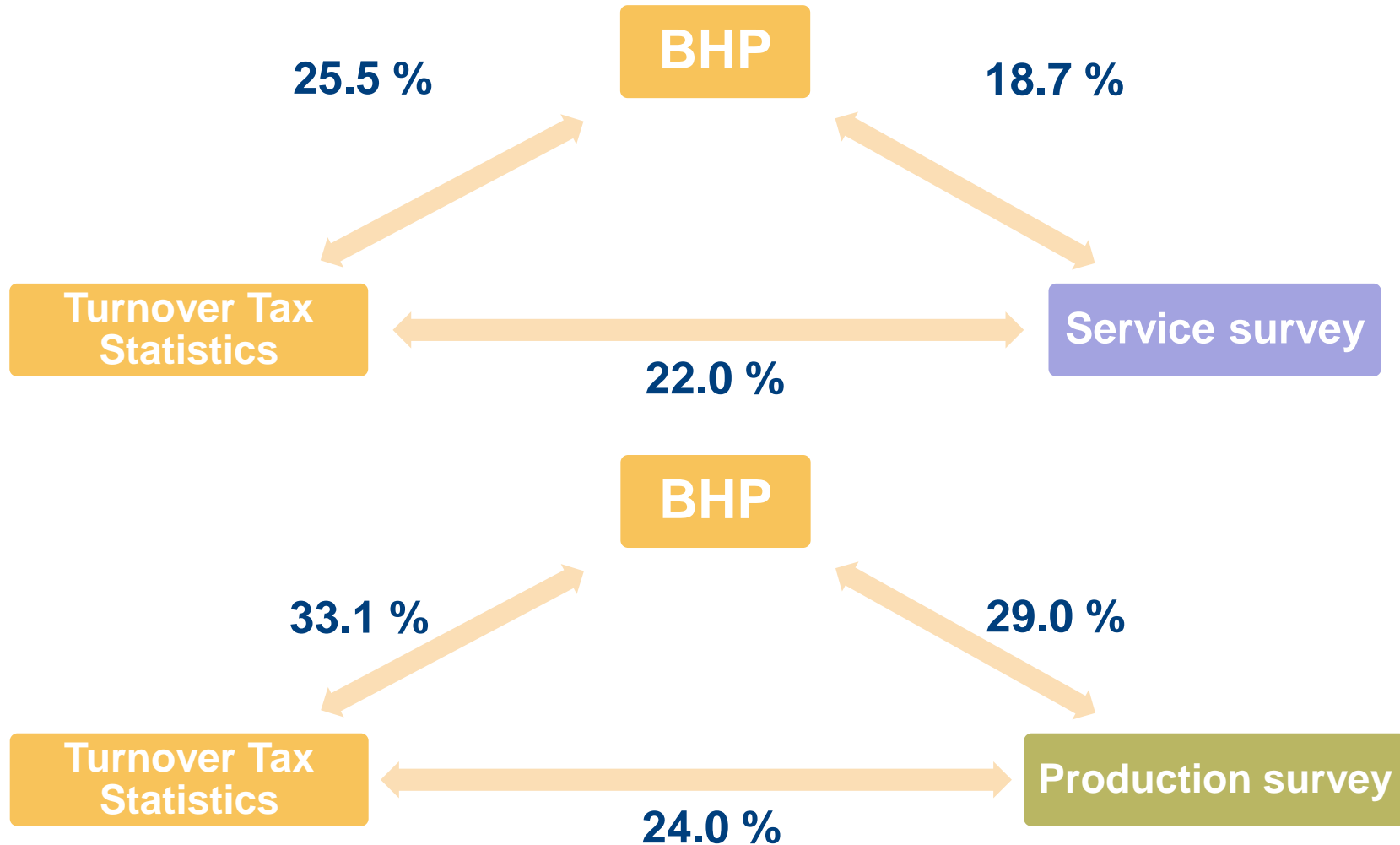
Results

- Location (German Federal State):
 - 98% to 99% consistency of the variable location
 - Linkage of identical OEFs
- Industry (2 digit):
 - Nearly no deviations between surveys of FSO (0.00 – 0.17% deviation)
 - Remarkable deviations between data of different data producers
 - Differences in production sample higher as in the service sample

Deviation of industry (service sample, production sample)

Service sample

Production sample



Summary I

- It is possible to link German firm data of different data producers with the URS
- Comparison of location & industry:
 - Location: no deviation
 - Industry: no deviations between FSO surveys, but remarkable deviations between surveys of different data producers.

Summary II

- Reasons for deviations:
 - different data collection methods
 - different definitions of variables (e.g. industry)



- Best consistency: one data producer with identical methods of data collection and equal definitions of variables is collecting the data.

But

- Need to know more about data generating processes (data quality->Blue-ETS)

Long-term perspectives

- To improve data quality:
 - Standardization of definitions (variables) for several datasets.
 - Exchange of knowledge.
 - Close coordination of data producers, researchers, ...
- Vision:
 - A central data collection, where a few relevant variables are surveyed and stored.
 - Variables are available for every survey conducted by researchers, FSO, BA...



As the result every firm survey will include the same information concerning relevant variables.

Thank you for your attention

Stefan Bender
stefan.bender@iab.de

<http://www.kombifid.de>

<http://fdz.iab.de>

Back-up

Datasets provided by the FSO and included in the KombiFiD dataset

| Dataset | Full sample/sample | Reporting unit | Reporting path |
|---|--------------------|----------------|------------------------------------|
| German Business register system (URS95) | full | firm | locally |
| Cost structure surveys: | | | |
| Cost structure survey in manufacturing, mining and quarrying | sample | firm | locally |
| Cost structure survey in the building industry | sample | firm | centrally |
| Annual surveys/reports: | | | |
| Annual survey in wholesale and retail trade | sample | firm | partly locally partly centrally |
| Annual survey incl. survey of investments in the building industry proper and in the finishing trade | sample | firm | locally |
| Annual report on enterprises in manufacturing, mining and quarrying | sample | firm | locally |
| Other official surveys: | | | |
| Monthly report incl. survey of orders received for local units in manufacturing, mining and quarrying | full | establishment | locally |
| Survey of investments in manufacturing, mining and quarrying | full | firm | locally |
| Structure of earnings survey | sample | establishment | locally |
| Structural survey in the services sector | sample | firm | locally |

Number of firms included in the KombiFiD dataset, ordered by the original data sources

| Dataset | Number of firms | Percentage |
|---|-----------------|--------------|
| BHP | 54,510 | 83.6% |
| Turnover tax statistics | 50.020 | 76.7% |
| Monthly report incl. survey of orders received for local units in manufacturing, mining and quarrying | 23,019 | 35.3% |
| Cost structure survey in manufacturing, mining and quarrying | 22,796 | 35.0% |
| Annual report in manufacturing, mining and quarrying | 22,680 | 34.8% |
| Survey of investment in manufacturing, mining and quarrying | 22,543 | 34.6% |
| Structural survey in services sector | 19,255 | 29.5% |
| Annual survey in wholesale and retail trade | 15,917 | 24.4% |
| Annual survey incl. survey of investments in the building industry and in the finishing trade | 4,780 | 7.3% |
| Cost structure survey in building industry | 3,436 | 5.3% |
| Structure of earnings survey | 2,816 | 4.3% |
| Number of firms in the KombiFiD dataset | 65,231 | 100 % |

Most frequently observed dataset combinations within the KombiFiD project

| | Dataset | Number of OEF |
|---|---|---------------|
| 1 | BHP/ Structural survey in services sector/ Turnover Tax Statistics | 12,254 |
| 2 | BHP/ Monthly report incl. surveys of orders received for local units in manufacturing, mining, quarrying/ Turnover Tax Statistics/ Cost structure survey in manufacturing, mining, quarrying/ Annual report on enterprises in manufacturing, mining, quarrying/ Survey of investments in manufacturing, mining, quarrying | 11,468 |

Number of OEFs included in the KombiFiD dataset, ordered by the original data sources

| Dataset | Number of OEF | Percentage |
|---|---------------|--------------|
| BHP | 48,515 | 97.8% |
| Turnover tax statistics | 39,000 | 78.6% |
| Cost structure survey in manufacturing, mining and quarrying | 17,558 | 35.4% |
| Monthly report incl. survey of orders received for local units in manufacturing, mining and quarrying | 17,460 | 35.2% |
| Annual report in manufacturing, mining and quarrying | 17,378 | 35.0% |
| Survey of investment in manufacturing, mining and quarrying | 17,269 | 34.8% |
| Structural survey in services sector | 15,590 | 31.4% |
| Annual survey in wholesale and retail trade | 10,787 | 21.8% |
| Annual survey incl. survey of investments in the building industry and in the finishing trade | 4,032 | 8.1% |
| Cost structure survey in building industry | 2,776 | 5.6% |
| Structure of earnings survey | 1,625 | 3.3% |
| Number of all OEF in the KombiFiD dataset | 49.613 | 100 % |