CHALLENGE AND POTENTIAL OF LINKING MICRODATA IN BUSINESS STATISTICS

Dr. Sandra Jung, Stefanie Käuser

↘ **Keywords:** microdata – microdata linking – calibration weighting – business statistics – foreign trade statistics

ABSTRACT

Linking microdata from different statistics will be of increasing importance as a tool for obtaining new information without increasing the respondent burden. This article describes the methodology of creating a linked set of microdata from structural business statistics, business demography, foreign trade statistics and the business register. In addition to the data linking process, the article presents the basic validation procedures and discusses methods for calibrated weighting of linked microdata. The analytical potential of the data is examined using the example of employment trends in enterprises engaged and not engaged in foreign trade.

Dr. Sandra Jung

is an economist and directs the section "Development of Business Statistics" of the Federal Statistical Office. Her main areas of interest are SMEs, Inward- FATS and economic globalization.

Stefanie Käuser

is an economist and, since 2014, a research fellow in the section "Development of Business Statistics" of the Federal Statistical Office. Within the context of an EU grant she dealt with micro data linking and micro data analysis of different business statistics.

1

Introduction

For producers of official statistics, linking data at the micro level is an appropriate way to respond to evolving user needs. However, this method also presents new challenges. Official business statistics can be inflexible when it comes to providing data on new topics such as globalisation or the use of modern information technology in enterprises. Most surveys in business statistics are conducted separately by NACE sections and sometimes differ with regards to the survey content. Answering questions that require a comprehensive approach can therefore be difficult. For example, this can be the case when analysing the effect of involvement in foreign trade or the use of modern ICT technology on business performance and employment. Additional burden on respondents can be avoided by linking data from different statistical sources at the micro level.

The potential of micro data linking has long been recognised in the field of scientific research. With the support of the research data centres of the statistical offices, external scientists explore diverse new fields of analysis and methodology with the use of linked micro data (Voshage et al., 2015, here: p. 14).

These benefits are also recognised at the European level. Eurostat – the statistical office of the European Union – has been supporting official micro data linking projects for a number of years.[1] The construction of the linked micro data set presented in this article has also been supported by Eurostat. In the EU project "Microdata linking of Structural Business Statistics and other business statistics" (MDL), micro data from various areas of business statistics were linked. In addition to the Federal Statistical Office, eight other National Statistical Offices of the European Union (Denmark, Austria, Latvia, Finland, Ireland, the Netherlands, Portugal and Sweden) and of Norway participated in the project in 2014 and 2015. For this purpose, micro data from the Business Register, Structural Business Statistics, Business Demography, foreign trade statistics, Inward FATS and the survey on ICT usage and e-commerce in enterprises (ICT survey) were linked – first cross-sectionally for each reference year from 2008 to 2012 and then longitudinally across the entire reference period. Subsequently, analyses were carried out on the linked dataset.

The new data gathered by micro data linking also present new challenges for official statistics. Among other things, the requirements stated by the Federal Statistics Law regarding implementation and reporting obligations for linking micro data from different sources are important framework conditions that have to be respected.

Article 13a of the Federal Statistics Law: Matching of Data

Data records from statistics according to Article 13, para. 1, data from the statistical register, data specified in the Law on the Use of Administrative Data, and data obtained by the Federal Statistical Office and the statistical offices of the Länder from generally accessible sources may be matched, provided that matching is required in order to obtain statistical information without conducting additional statistical surveys.

Micro data linking also comes along with new methodological challenges. The following article presents basic aspects of validating the MDL dataset created and discusses methodological aspects related to linking mainly sample-based survey data. The analytical potential of the dataset is examined using the example of employment trends in enterprises with differing degrees of involvement in foreign trade.

2

Database and linking process

In order to create the MDL database, a number of different surveys were linked at the micro level. This requires a unique identifier that enables the linkage of data from different sources cross-sectionally and longitudinally. In the following section, we will briefly introduce the data sources that were linked and provide an overview of the matching rates of these data sources.

1 Examples are the following EU projects: ICT Impact Project, ESSLimit Project, ESSLait Project. For a report on the ICT Impact Project see Rauland/Bauer, 2010.

2.1 Included data sources

The backbone of the MDL database is the Structural Business Statistics (SBS) of the reference years 2008 – 2012. These annual surveys comprise industry, construction, domestic trade as well as relevant parts of services. The survey characteristics cover the most relevant areas of business activity of enterprises such as turnover and persons employed. In addition to that, a multitude of other characteristics are available (such as gross value added at factor costs, investments, wages and salaries). Most of the Structural Business Statistics are based on surveys. ↘ Summary table 1, p. 4

The statistical Business Register (URS) is another major data source especially for linking the data from all sources. The Business Register is a continuously updated database, comprising all active legal units and their local units. It holds administrative data and other business characteristics such as taxable turnover and the number of employees subject to social insurance. Furthermore, information on control links among enterprises in enterprise groups is available. Every legal unit in the Business Register has an ID number (URS ID). This ID serves as an identifier for linking the structural statistics (including Business Demography, Inward FATS and the ICT survey) among each other and to the URS.

Another data source that was included is foreign trade statistics which show cross-border foreign trade. It is divided into Intrastat statistics – which capture trade in goods between EU member states – and Extrastat which cover trade in goods with non-EU countries[2]. The administrative ID serves as an identifier between foreign trade statistics and the Business Register.

Business Demography is a source of information on enterprise births and closures as well as the population of active enterprises. Structural indicators such as the enterprise birth rate and the survival rate of newborn enterprises are calculated using these data. The main source of information on Business Demography is the statistical Business Register (see Rink et al., 2013).

Moreover, we included data from foreign affiliates statistics (so-called Inward FATS statistics; see Nahm/Söllner, 2014) and the survey on ICT usage and e-commerce in enterprises.[3]

2.2 Linking procedure

The data were linked in a two-step procedure. As a first step, micro data from Structural Business Statistics, Business Demography, Inward FATS and the ICT survey were linked with the Business Register using the URS ID and subsequently linked cross-sectionally for every reference year. The data from foreign trade statistics were linked to the Business Register via the administrative ID. In a second step, data from all reference years were linked via the URS ID to create a final longitudinal dataset. This longitudinal dataset contains micro data from all enterprises that were included in at least one survey (including the URS) in at least one reference year, comprising about 7.7 million units. ↘ Figure 1, p. 5

2.3 Matching rates

The matching rates that resulted from the two-step linking process differed between data sources. Achieving high matching rates is desirable when linking micro data because this increases the robustness of the analytical results.

↘ Table 1 presents the matching rates of each data source with the Business Register. Since the Business Register serves as a sampling frame for Structural Busi-

Table 1
Matching rates[1] of the individual surveys when linked with the statistical business register (URS)

	2008	2009	2010	2011	2012
	%				
Structural business statistics	99.4	99.5	99.6	99.6	99.6
Foreign trade statistics	99.6	99.6	99.0	100	99.7
Inward FATS	98.6	100	100	100	100
Business demography	100	100	100	100	100
ICT survey	98.8	99.0	99.1	98.9	99.1

1 Share of units in the data source which can be linked with the URS of the relevant reference year.

2 Detailed information on the methods and definitions applied and on the quality of the statistical results of foreign trade statistics are contained in the "Foreign trade" quality report (www.destatis.de).

3 Detailed information on the methods and definitions applied and on the quality of the statistical results of the ICT survey are contained in the German-language quality report "Nutzung von Informations- und Kommunikationstechnologien (IKT) in Unternehmen 2015" (www.destatis.de).

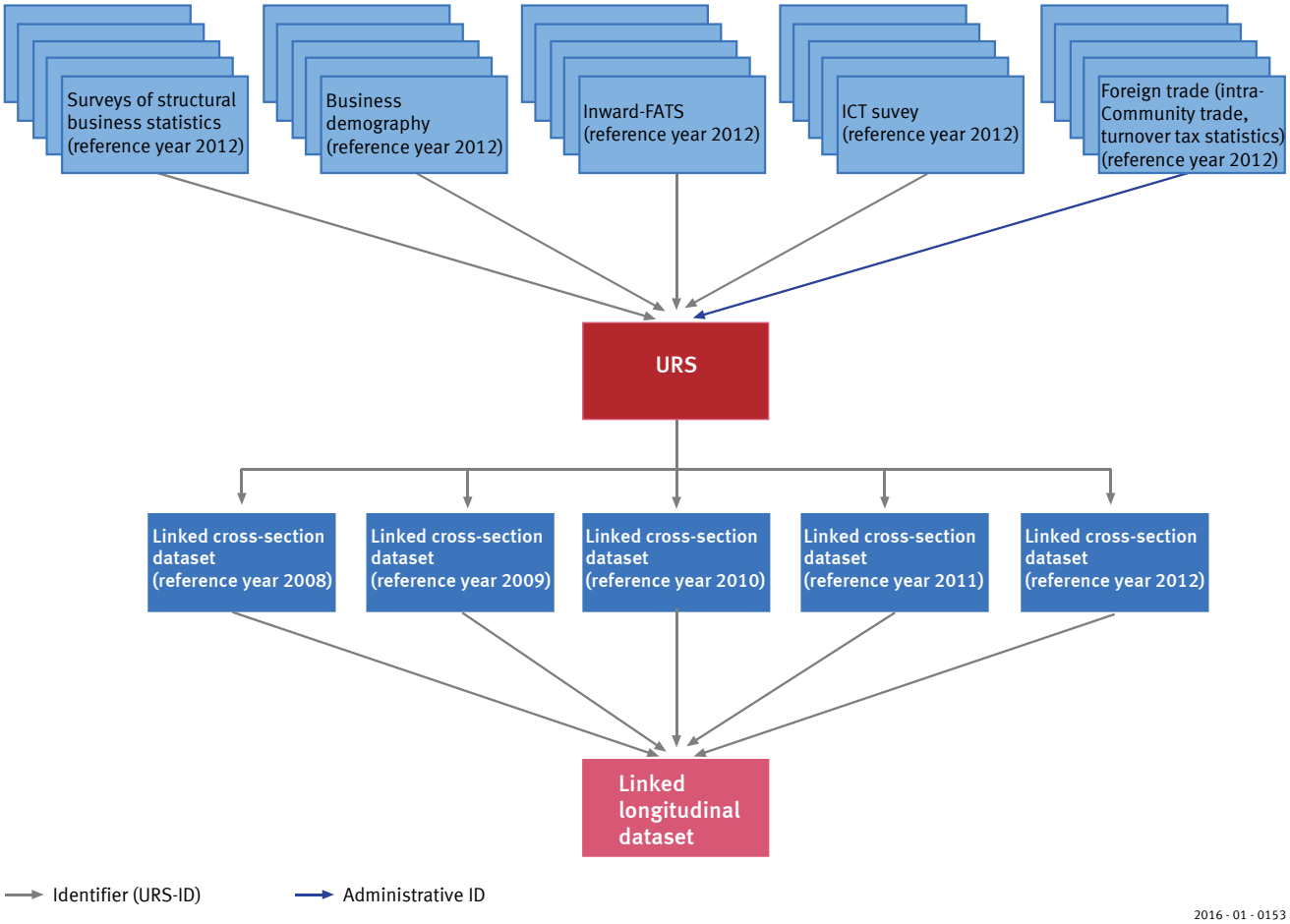
Summary table 1

Database of the MDL dataset

Section[1]	Survey	Type of survey	Cut-off limit	Coverage
	Annual structural business statistics (SBS)			
B, C	Cost structure survey in manufacturing, mining and quarrying	Sample survey	Enterprises with 20 or more persons employed	Roughly 18,000 enterprises
	Survey of investments in manufacturing, mining and quarrying	Complete count	Enterprises with 20 or more persons employed	Roughly 50,000 enterprises
	Structural survey of small enterprises in manufacturing, mining and quarrying	Sample survey	Enterprises with fewer than 20 persons employed	Roughly 6,000 enterprises
D, E	Cost structure survey in energy and water supply	Complete count	Enterprises in water supply with an annual water distribution of 200,000 m³ and over	3,000 enterprises in energy supply; 7,000 enterprises in water supply
	Survey of investments in energy and water supply	Complete count	Enterprises in water supply with an annual water distribution of 200,000 m³ and over	3,000 enterprises in energy supply; 7,000 enterprises in water supply
F	Cost structure survey in construction industry	Sample survey	Enterprises with 20 or more persons employed	6,000 enterprises
	Annual survey including survey of investments in main construction industry	Complete count	Enterprises with 20 or more persons employed	Roughly 7,500 enterprises
	Annual survey including survey of investments in building completion work	Complete count	Enterprises with 20 or more persons employed	Roughly 6,500 enterprises
	Structural survey of small enterprises in construction industry	Sample survey	Enterprises with fewer than 20 persons employed	6,000 enterprises
G	Annual survey in wholesale and retail trade, repair of motor vehicles and motorcycles and personal and household goods	Sample survey	None	Roughly 47,000 enterprises
I	Annual survey in accommodation and food services	Sample survey	None	Roughly 10,000 enterprises
H, J, L, M, N, S95	Structural survey in the service sector	Sample survey	None	Roughly 220,000 enterprises
	Business demography			
B to N, P to S	Enterprise births and closures	Register-based evaluation	Minimum taxable turnover of 17,500 euros or at least 1 person employed subject to social insurance	Roughly 3,700,000 enterprises (including surviving enterprises)
	ICT survey			
C to N	Usage of information and communication technologies in enterprises	Sample survey (voluntary response)	Microenterprises (fewer than 10 persons employed) receive reduced questionnaire	Roughly 20,000 enterprises (share of microenterprises between 25% and 30%)
	Inward FATS			
B to N, P to S	Foreign-controlled enterprises	Extraction of all foreign-controlled enterprises from the URS, linking with SBS data	Minimum taxable turnover of 17,500 euros or at least 1 person employed subject to social insurance	Roughly 21,000 to 27,000 enterprises
	Foreign trade statistics			
All	Exports and imports of goods of German enterprises	Complete count	Intra-Community trade: minimum value of goods of 500,000 euros per direction of trade in the reference year (compulsory response) (400,000 euros until 2011), extra-Community trade: none	Roughly 550,000 to 577,000 enterprises (including enterprises below the reporting threshold in intra-Community trade)

1 Classification of Economic Activities, 2008 edition (WZ 2008).

Figure 1
Procedure of linking microdata from different sets of business statistics



2016 - 01 - 0153

ness Statistics, the resulting matching rates are quite high at around 99 %. However, in some cases units that have been surveyed in Structural Business Statistics may not be in the Business Register of the same reference year (for more details, see also section 3.1).[4]

As Structural Business Statistics (SBS) are the backbone of the MDL database, we are interested in the matching rates between SBS and the other surveys. For foreign trade statistics, the matching rates are relatively low: they range from 11.2 % to 13.5 % p.a. The main reason for this is that foreign trade statistics are based on a full count with a cut-off limit whereas SBS are mainly based on sample surveys (see also section 3.1). This implies that an enterprise that has been covered in foreign trade statistics may not have been surveyed in SBS because it was not sampled. Alternatively, the foreign trade volume of an enterprise that has been surveyed in SBS may be below the cut-off limit for Intrastat and therefore the enterprise may not have been covered in foreign trade statistics. In both scenarios, data cannot be matched between the two data sources. The matching rates are highest for the ICT survey, ranging from 40.1 % to 45.4 %. ↘ Table 2

The second step of the linking process was the creation of a longitudinal dataset. In this context, the matching rates of Structural Business Statistics across all reference years are relevant because they show the share

4 The rather high matching rates for foreign trade statistics / business register refer to foreign trade data that have already been processed. For a first-time linking between foreign trade data and the business register, the rates are slightly lower (Allafi, 2001).

Table 2
Matching rates[1] when linked with structural business statistics

	2008	2009	2010	2011	2012
	%				
Foreign trade statistics	11.2	12.6	12.8	13. 4	13.5
Inward FATS	30.4	31.5	30.6	31.1	35.6
ICT survey	43.0	45.4	40.6	40.1	45.4

1 Share of units of the relevant survey which were surveyed in structural business statistics in the relevant reference year.

of units from reference year 2008 which can still be observed in subsequent years. ↘ Table 3

Table 3
Matching rates[1] of the micro data from structural business statistics, longitudinal[2]

	2009	2010	2011	2012
	%			
Structural business statistics	72.1	65.0	31.3	25.9

1 Share of units surveyed in structural business statistics in 2008, which were still observed in the subsequent years.
2 Referring to the 2008 cohort of structural business statistics.

Of all units surveyed in Structural Business Statistics in 2008, about 25.9 % were still contained in the MDL dataset in 2012. The decrease in matching rates over time has various reasons. The main reason is sample rotation: survey units are replaced by new units at different time intervals. Moreover, enterprise deaths can be another reason for this decrease (see also section 3.1).

3

Validation and extrapolation

3.1 Validation of the linked dataset

In order to ensure the quality and consistency of the MDL dataset for subsequent analyses, the linked dataset was subjected to an extensive validation process. One focal point was the investigation of causes for unmatchable units between data sources and within data sources over time. A second focal point of validation was to find a method for dealing with foreign trade data reported by tax groups[5]. Furthermore, we developed a strategy for identifying so-called demographic events in the linked dataset. Those three aspects will be discussed in more detail in the following sections.

No-match units

Considering the multitude of surveys included in the MDL dataset, a fairly high number of units that cannot be matched across all data sources is not unusual. During the investigation of the causes for these "no match" cases, we differentiated between those that are due to the survey methodology and design of the statistics and those that relate to technical errors during linking. The aim of this validation step was to identify and correct possible errors that occurred during matching.

Causes relating to survey design can for instance be differing statistical populations and modes of data collection. Since the survey data we linked were partly full-count data and partly sample data with different sampling fractions and cut-off limits, not all units could be matched across all data sources. Moreover, the surveys may be based on different statistical populations: the population of Structural Business Statistics is all active enterprises located in Germany within NACE sections B to N (without K) + S95 of the Classification of Economic Activities (2008 edition). In contrast, the population of foreign trade statistics is not limited to specific NACE sections but comprises all enterprises involved in cross-border trade in goods. Relatively low matching rates between SBS and foreign trade statistics should be interpreted with this context in mind (see table 2). No-match cases due to survey design have to be taken as given and cannot be corrected retrospectively.

Generally speaking, improved coordination of sample surveys and harmonisation of survey design would be beneficial for the purpose of micro data linking. This would help to increase matching rates of the sample surveys involved and help to increase the robustness of the results when extrapolating data from linked sample surveys. However, the coordination of the involved sample surveys can also not be altered retrospectively.

In addition to causes relating to survey design, technical errors when matching IDs can result in no-match cases. These errors can potentially be detected and mended afterwards. For instance, the same empirical unit can have different IDs in different data sources. This type of

5 A detailed explanation of tax groups is given in Wagner, 2004.

ID mismatch may occur between Business Register and SBS micro data when an enterprise moves to a different Land and thus receives a different ID in the Business Register whilst keeping the old ID in SBS because the sample for SBS was drawn prior to the change in ID. Discrepancies of this type between the datasets are among the reasons why matching rates between the BR and SBS never reach 100 % in any reference year (see table 1).

Furthermore, the same enterprise may have different IDs in different reference years in the same data source. When the micro data have been linked longitudinally, the enterprise cannot be tracked across all reference years. We investigated the dataset thoroughly to detect potential errors of this type. In the majority of cases when IDs were missing in certain reference years, we were able to identify sample rotation, changes in NACE activity or relocation to a different Land as causes for the absence.

Reallocation of foreign trade data within tax groups

When linking Structural Business Statistics data to foreign trade data, some consistency issues resulting from foreign trade data reported by controlling companies of tax groups had to be resolved in the MDL project. Tax groups are economically linked legal units which are treated as a single entity by German tax authorities and therefore pay their taxes jointly (e. g. their turnover tax; see Wagner, 2004). In foreign trade statistics, controlling companies of tax groups are obligated to report foreign trade activities for all members of the tax group as a whole. Thus, there are no reports available from individual legal units that are tax group members. In the data, this means that the trade volume listed under the ID of the controlling company in reality refers to the tax group as a whole – not just the individual legal unit. During the validation process, this manifested itself in the form of exports which considerably exceeded turnover from SBS. In order to mitigate this issue, we reallocated exports and imports reported by controlling companies to other tax group members within tax groups. For this purpose, we identified all tax group members belonging to a controlling company in the Business Register and added their IDs to the foreign trade datasets. Then, trading volumes (exports and imports) of controlling companies were redistributed within tax groups. As a distribution key for the allocation, we used redistributed turnover[6] from the Business Register. The following formula describes the reallocation procedure:

$$(1)\quad A_{ip} = \frac{\alpha_i}{\sum_{i=1}^{k} \alpha_i} \beta_p^{org} A^{org},$$

A_{ip} is the estimated trading volume of tax group member i with partner country p, while A^{org} is the total trading volume of a controlling company. The estimated turnover of tax group member i is α_i, which is divided by the sum of turnovers of all k members of the tax group: $\sum_{i=1}^{k} \alpha_i$. This determines the share of tax group member i of total tax group turnover. β_p^{org} represents the share of partner country p in the total trading volume of a particular controlling company.

Exports of tax group members whose main activity was not in manufacturing or domestic trade were exempted from the reallocation procedure since we wanted to include only those NACE sections in which goods exports of considerable scope were plausible. The reallocation procedure helped to increase the consistency of the combined dataset and improve the plausibility of the data.

Dealing with demographic events

The corporate landscape of an economy is not static but is in constant motion due to enterprise births and deaths as well as other demographic events (such as mergers, acquisitions, spin-offs, restructurings etc.)[7]. These demographic events can occur within the reference period and manifest themselves in different ways in the MDL dataset. For example, a takeover can result in the ID of the absorbed company vanishing from the dataset, whereas certain characteristics of the absorbing company such as turnover and persons employed exhibit unusually strong fluctuations in two consecutive time periods. These types of constellations can be problematic for longitudinal analysis of growth rates because "organic" growth from within the company cannot be distinguished from "artificial" growth resulting from mergers and acquisitions. In order to make the MDL dataset more consistent for longitudinal analyses, a variable was created to capture this phenomenon.

6 Reallocation of turnover reported by controlling companies within tax groups is regularly performed when producing the federal copy of the business register.

7 For a clear distinction between demographic events see Rink et al., 2013, pp. 425 f.

The procedure for identifying demographic events will be presented in more detail in this section. First, all units in Structural Business Statistics were categorised into four size classes based on their values for persons employed and turnover. Within these size classes, the average change in turnover and number of persons employed in full-time equivalents (FTE) between two consecutive years ($\overline{\Delta x}_{gt}$) was calculated. Furthermore, we took the standard deviation (σ_{gt}) of the respective variable within each size class for the same period. Based on these figures, we were able to select SBS units whose values for turnover or persons employed (FTE) deviated by more than three standard deviations of the average change within their respective size class. The following formula illustrates this:

(2) $|\Delta x_{it}| > |\overline{\Delta x}_{gt}| + 3\sigma_{gt}$,

Δx_{it} is the absolute change in the respective variable of unit *i* in time period *t*. This threshold was set to select units which exhibit unusually high fluctuations in turnover or persons employed (FTE).

In a second step, we checked whether changes had occurred in other variables from the Business Register (such as the name of the enterprise, group status, type of unit or NACE code) in the same period. We found that all pre-selected units did indeed exhibit fluctuations in the URS variable "type of unit" which contains information on whether the unit is a legal or a local unit. The majority of these units changed their status from legal unit to local unit during the time period in question. We concluded that these indications provided sufficient evidence that a demographic event had occurred. Those units were then marked accordingly in the MDL dataset.

3.2 Methods for extrapolation

When linking micro data from samples, consistency issues of marginal totals based on the linked data may arise with regards to benchmarks from existing publications referring to the individual surveys. In order to ensure the consistency of the results, an appropriate method for extrapolation has to be chosen. In the MDL project, the missing consistency of variable values was especially relevant with regards to important benchmarks of variables in Structural Business Statistics and Inward FATS. Therefore, in the following section, we will present possible solutions for extrapolating linked micro data from multiple sources while maintaining consistency with existing publications.

Calibrated weighting

Calibrated weighting is a specific type of extrapolation method. This type of procedure is similar to standard extrapolation but differs in the sense that correction factors are computed in order to calibrate the extrapolation factors. The aim of this calibration procedure is to adjust the extrapolation factors in such a way that a predefined benchmark value can be reached. The correction factors are computed with the following formula:

(3) $KF_j = \frac{Y_j}{\hat{Y}_j}$.

Here, $\hat{Y}_j = \sum_{i \in j} \hat{y}_i$ is the sum of variable *y* of all units *i* in NACE section *j*, extrapolated with SBS survey weights. The term $Y_j = \sum_{i \in j} y_i$ is the benchmark sum of variable *y* of all units *i* in NACE section *j* (as defined by existing published values). The correction factor is multiplied with the original survey weight to obtain the calibrated extrapolation factor.

For the purpose of the MDL project, we used the approach of calibrated weighting with correction factors in order to extrapolate the linked micro data. The objective was to ensure consistency with the most important benchmarks of Inward FATS subject to the constraint that deviations from the (important) marginal totals of Structural Business Statistics (in total) were to be minimised. These benchmarks were chosen as reference points because the focus of the publications from the project was intended to be on characteristics of foreign-controlled and small and medium-sized enterprises (SMEs). Applying this method made it clear that it is simple to implement but limited in terms of applicability. When the number of benchmarks is fairly low, consistency with existing publications can be achieved. However, this method is no longer applicable when the number of benchmarks to be met is larger. Due to this disadvantage, an improved method for extrapolation is desirable for future activities.

General Regression Estimator

As an alternative to calibrated weighting with correction factors, a regression-based calibration of the extrapolation factors with the General Regression Estimator (GREG) can be considered (see Deville/Särndal, 1992). With GREG, extrapolation factors can be calibrated using auxiliary variables in order to meet pre-defined benchmarks. The advantage of this method is that benchmarks from multiple statistics can be met simultaneously. Moreover, GREG provides an estimate of the accuracy of the estimation in the form of the relative standard error. This method of calibrated weighting is also used for the microcensus (Afentakis/Bihler, 2005). For an upcoming follow-up project, we plan on using GREG as the method for extrapolating the data.

4

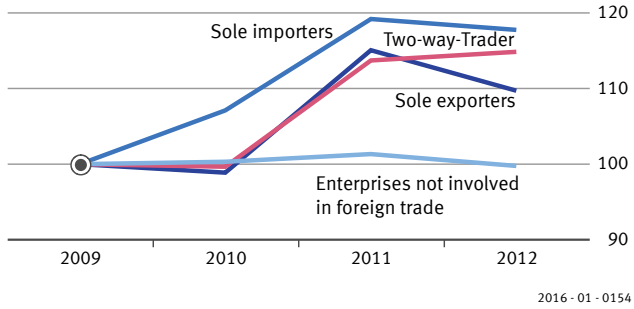
Analytical potential

By linking data from different areas of business statistics at the micro level, interdisciplinary questions can be answered which could thus far only be explored with limitations. For instance, linking data from foreign trade statistics with SBS and Inward FATS data can provide new insights on the structure and direction of global value-added chains. The relationship between cost structures, productivity, market exits and changes in ownership and control structures are further topics for empirical analyses that are possible through micro data linking.

Below we elucidate employment trends of trading and non-trading enterprises during the time period 2009 – 2012[8]. For this purpose, we depict the development of the number of persons employed per enterprise for importers, exporters, two-way traders[9] and for enterprises not involved in foreign trade.

The number of persons employed in enterprises not involved in foreign trade remained almost constant between 2009 and 2012. In contrast to that, enterprises active in foreign trade exhibited considerably more intense employment dynamics. Sole importers in particular saw a rise in persons employed of just under 20 % from 2009 to 2012. Employment gains in sole exporters and two-way traders were slightly lower at 10 % and 17 % respectively. ↘ Figure 2

Figure 2
Development of the number of persons employed per enterprises by involvement in foreign trade
2009 = 100



A possible explanation of the particularly strong surge in employment in sole importers could be that they were able to profit from lower import prices as a result of the global economic crisis.[10] In order to thoroughly analyse the relationship between foreign trade with companies abroad and employment effects, more detailed analyses than the scope of this article allows for are necessary.

5

Outlook

Linking micro data from multiple sources will be of increasing importance in the future of official statistics because this method provides a way to gain new information without adding to the respondent burden.

The EU project "Micro data linking of Structural Business Statistics and other business statistics" 2014/2015 helped to increase the analytical potential of existing business statistics and to highlight the possibilities and limitations of linking micro data from different sources. During the project, micro data of the reporting years

8 We did not include reference year 2008 because foreign trade data available from this year are not complete, which is due to the limited retention period for micro data in foreign trade statistics.

9 Enterprises involved in both importation and exportation.

10 Wagner and Weche Gelübcke (2014) detect higher probabilities of survival for importers than for enterprises not involved in foreign trade in the period of global recession. The reason they see is better conditions for importing enterprises in factor markets.

2008 to 2012 from six[11] different data sources were linked and statistically analysed. The results of the project will be published in three "Statistics explained" articles on the Eurostat homepage. To this date, one article presenting statistics for independent SMEs and large enterprises vs. SMEs and large enterprises belonging to an enterprise group has already been published (Airaksinen et al., 2015).

In a follow-up project, the MDL dataset will be extended with data from the reporting year 2013. Moreover, the follow-up project will also focus on analysing the economic activities of high-growth enterprises. In addition, we will use the opportunity to develop the methodology further – for instance by using the General Regression Estimator for extrapolating the micro data.

The micro data that were linked during the 2014/2015 MDL project offer opportunities for analysis that go beyond what individual statistics offer. Relationships between economic indicators such as employment development, productivity, profitability and other characteristics of businesses can be explored. Through linking with data from Business Demography, determinants of survival in the market can be determined. Currently, there are plans to publish further selected analyses with data from the MDL project in this journal.

11 The tender specifications permitted to include up to four additional data sources into the MDL dataset, such as Outward FATS statistics or the Community Innovation Survey (CIS). This is why additional data from other sources are available for some participating countries.

REFERENCES

Afentakis, Anja/Bihler, Wolf. *Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005*. In: Wirtschaft und Statistik 10/2005, pp. 1039 ff.

Airaksinen, Aarno/Alajääskö, Pekka/Luomaranta, Henri/Roodhuijzen, Anton. (2015). *Dependent and independent SMEs and large enterprises*. Web document [Access date: 15.01.2016]. Available from: http://ec.europa.eu

Allafi, Sabine. *Intrahandelsergebnisse nach Wirtschaftszweigen*. In: Wirtschaft und Statistik 6/2011, pp. 546 ff.

Deville, Jean-Claude/Särndal, Carl-Erik. *Calibration Estimators in Survey Sampling*. In: Journal of the American Statistical Association, 87, 1992, pp. 376 ff.

Voshage, Ramona/Drafz, Ute/ Habla, Heike/Klumpe, Matthias/Meisdrock, Claudia/ Nowak, Kristin/Raab, Stefanie/Richter, Alexander/Rößner, Michael/Schmidtke, Kerstin. *Formal, faktisch oder absolut nachgefragt? Die Entwicklung der Nachfrage in den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder*. FDZ-Arbeitspapier No. 49, pp. 14 ff., Wiesbaden, 2015.

Nahm, Matthias/Söllner, René. *Die wirtschaftliche Bedeutung auslandskontrollierter Unternehmen in Deutschland 2012*. In: Wirtschaft und Statistik 12/2014, pp. 754 ff.

Rauland, Jan/Bauer, Oliver. *Statistische Analyse des Einflusses von Informations- und Kommunikationstechnologien auf die Produktivität von Unternehmen*. In: Wirtschaft und Statistik 12/2010, pp. 1066 ff.

Rink, Anke/Seiwert, Ines/Opfermann, Rainer. *Unternehmensdemografie: methodischer Ansatz und Ergebnisse 2005 bis 2010*. In: Wirtschaft und Statistik. 6/2013, pp. 422 ff.

Statistisches Bundesamt. *Qualitätsbericht Außenhandel*. Wiesbaden 2014. [Access date: 29.01.2016]. Available from: www.destatis.de

Statistisches Bundesamt. *Qualitätsbericht Nutzung von Informations- und Kommunikationstechnologien in Unternehmen 2015*. Statistisches Bundesamt, Wiesbaden 2015. [Access date: 27.01.2016]. Available from: www.destatis.de

Wagner, Ingo. *Schätzung fehlender Umsatzangaben für Organschaften im Unternehmensregister*. In: Wirtschaft und Statistik 9/2004, pp. 1001 ff.

Wagner, Joachim/Weche Gelübcke, John P. *Risk or Resilience? The Role of Trade Integration and Foreign Ownership for the Survival of German Enterprises during the Crisis 2008-2010*. In: Jahrbücher für Nationalökonomie und Statistik, Bd. 234, 6/2014, pp. 757 ff.

Extract from the journal WISTA Wirtschaft und Statistik

Published by:
Statistisches Bundesamt (Federal Statistical Office),
Wiesbaden

www.destatis.de

You may contact us at
www.destatis.de/kontakt

Abbriviations

WISTA	=	Wirtschaft und Statistik
JD	=	annual average
D	=	average (for values which cannot be added up)
Vj	=	quarter of a year
Hj	=	half-year
a. n. g.	=	not elsewhere classified
o. a. S.	=	no main economic activity
St	=	piece
Mill.	=	million
Mrd.	=	billion

Explanation of symbols

–	=	no figures or magnitude zero
0	=	less than half of 1 in the last digit occupied, but more than zero
.	=	numerical value unknown or not to be disclosed
. . .	=	data will be available later
X	=	cell blocked for logical reasons
I or —	=	fundamental change within a series affecting comparisons over time
/	=	no data because the numerical value is not sufficiently reliable
( )	=	limited informational value because numerical value is of limited statistical reliability