**Coherence – Accuracy – Flexibility**

Carsten Schumann, Federal Statistical Office of Germany (Destatis)

carsten.schumann@destatis.de

**Abstract**

*In European Business Statistics the “number of active enterprises” and the “number of persons employed” are provided from two sources (“Structural Business Statistics” (SBS) and “Business Demography” (BD)) with equal populations and similar definitions. Consequently, users expect the same results. However, deviations occur because most Member States use sample surveys for SBS and administrative data processed in the Business Register (BR) for BD. In the past some deviations have been severe and Eurostat launched initiatives to remove the gap.*

*Coherence can be improved to a certain degree by calibrating results. However, it is difficult to decide on a benchmark: primary surveys have a sophisticated quality assurance for high reliability of each characteristic while the BR respectively the administrative data has better coverage and allows the publication of small-area data. Unfortunately, the NACE-code – a very crucial characteristic for business statistics – is not always reliable from administrative sources and SBS has to cope with non-responses. Furthermore, there is a limit to the amount of calibration constraints and not calibrated variables and breakdowns can be biased.*

*Clearly, this results in a delicate trade-off between credibility of published results, their accuracy and flexibility in promptly producing additional results on newly arising user demand. Recent initiatives promoted “from stove pipe to data warehouse” and “reuse of existing data” to enhance flexibility and improve responsiveness while reducing response burden. Pilot projects have proven the vast potential of micro-data-linking and are still just tapping the surface. Ex-ante restrictions imposed on results before publication might hamper these possibilities.*

*The presentation to this paper wants to trigger a discussion about goals and limitations of initiatives to improve coherence. Should we bend results to enforce coherence? Is coherence needed on every level of aggregation? How much gap can be explained to users? How will calibrations affect accuracy and flexibility?*

**Keywords:** Coherence, Accuracy, Flexibility, Consistency, Credibility, FRIBS

**1. Background**

In European Business Statistics the “number of active enterprises” and the “number of employees and self-employed persons in active enterprises” can be retrieved from two statistical sources: Business Demography (BD) and Structural Business Statistics (SBS).

*1.1 Business Demography (BD)*

The main objective of BD is to track the evolution of individual businesses in the enterprise sector based on panel analysis. Published results include “birth-rate”, “death-rate” and “survival-rate”. Under the heading “employer-demography” the same ratios are compiled for enterprises that create jobs. The data requirements contain several (small) subsets, such as “High-Growth-Enterprises [HGE]” (annual growth rate of 10% or more for a period of three years) or “Gazelles” (which are young HGEs).

To identify these small subsets and track them over time, it is crucial to have a comprehensive data set for each reference year that can be linked with specific identifiers. For this reason the member states use administrative data processed in the national business registers (BR) as source for business demography. BD is not sample based, but full registers are processed (complete census). Because of this, it also allows for regional analysis down to the NUTS-3 level.

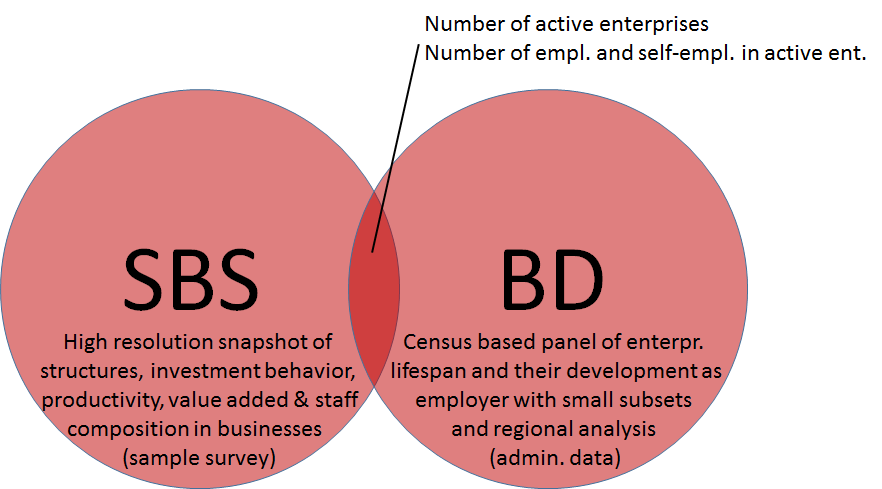
*1.2 Structural Business Statistics (SBS)*

The main objective of SBS is to provide a high resolution snapshot of the structure of the enterprise sector. This concerns size-classes and NACE-codes, but also cost structures, assets, investment behaviour, staff composition and the distribution of value added. In fact the range of deliverables for SBS covers 74 variables that can be cross-tabulated into various dimensions.

*1.3 Cross-domain coherence and internal consistency*

Both domains provide the two variables “number of active enterprises” and “number of employees and self-employed persons in active enterprises” with equal populations and similar definitions. However, in publications this is most commonly used as reference point for ratios: e.g. share of [employees and self-employed persons in] newly born enterprises (BD) or share of [employment in] small-and-medium-sized enterprises (SBS). In this respect, it is of course very important, that all sub-aggregates add up to the total. Also, in BD the saldo of enterprise births and deaths should lead to the next total to comply with the panel style. Currently, both domains transmit these variables under different codes and the focus is on internal consistency. With the upcoming Framework Regulation Integrating Business Statistics (FRIBS) only one transmission of these variables is foreseen, which will be used for both sources. In this respect there would be a new priority setting: from internal consistency to cross-domain coherence.

Figure 1: Objectives and overlap of SBS and BD



source 1: self-made

**2. Possible sources of incoherence**

The Federal Statistical Office of Germany has acknowledged the need for action to reduce the gap between BD and SBS concerning the two shared variables. A Task Force has been set up in 2015 to analyse reasons for these deviations and propose solutions for better coherence.

*2.1 Outdated sampling frames*

The main impact on the deviating “number of active enterprises” (respectively employment) is caused by over-coverage in the sampling frame. The reference year of administrative data in the German Business Register (BR) used to be t-2 years[[1]](#footnote-1) at the time of sample design for SBS, meaning that some sample units have ceased to exist when the survey questionnaires were sent out in year t+1 (2.5 years after the reference year of the sampling frame). This has been cured by an acceleration of the BR production process to reflect year t at the time of sample design in t+9 months. Starting with reference year 2017 this acceleration should have a visible effect and reduce deviations.

*2.2 Un-harmonized sampling frames*

One – in quantitative terms minor – reason for deviating results was caused by the practice of extracting the sample frame from the BR at different points in time for each domain (i.e. “industry”, “construction”, “trade” and “services”). If for example industry statistics extract a sample frame in April and trade statistics extract their sample frame in September, some units might have been reassigned to a new activity-code (NACE) in the meantime. This can lead to double counting or omittance of the unit.

*2.3 Different legal basis and/or IT-environment for each domain (stovepipes)*

In Germany the legal basis for statistics is traditionally domain oriented. One specialized law is regulating trade statistics, another law is regulating services statistics and yet another is regulating industry statistics. If a unit is requested to participate in one statistic with a legal obligation to report, but it turns out that its main activity (NACE) has shifted to another domain, the obligation to report is not valid anymore because it was referencing to the wrong legal basis.

In some cases re-assigned units cannot even be passed on, if both domain statistics are based on the same legal basis, but they are produced in different IT-environments.

*2.4 Treatment of unit-non-response*

If a unit is not responding to the questionnaire, it could have several different reasons, for example:

* The postal address is wrong (obligation to report cannot be delivered),
* the enterprise has ceased to exist,
* the administrative data in the sampling frame was based on a group-tax-declaration and the unit is not a single enterprise,
* the unit is a miss-match of administrative sources (doublet),
* the unit is not an enterprise, but merely a local unit of an enterprise,
* the activity code (NACE) was outdated and the unit is outside the legal basis.

In earlier years these cases were sometimes treated as over-coverage in the sampling frame and the results were corrected for the unit concerned and its expansion factor. Starting with reference year 2014, the treatment of unit-non-response is harmonized among all domains and the reference is always the finalized BR (respectively identified errors are corrected in the BR) to improve coherence between SBS and BD. If a unit is active in the BR, it is recognized in SBS and BD equally. The only exception to this rule is a wrong NACE-code (see chapter 2.7)

*2.5 Treatment of private energy production*

By far the greatest deviation on NACE-section-level (letters) occurred in section D “Electricity, Gas, Steam and Air conditioning supply”. That is mainly based on small and micro-enterprises in production of electricity. If a private household has a solar-panel on his roof or a wind-mill in the garden, it has to register an enterprise in order to receive subsidies for the investment. This leads to the creation of many micro-enterprises that actually produce less electricity than they consume themselves (netto-consumers). These units used to be excluded from SBS. Starting with reference year 2017, these micro-enterprises will be imputed into SBS from administrative data to ensure coherence with BD.

*2.6 Treatment of private landlords*

Another striking deviation on NACE-section-level occurs in section L “Real estate activities”. Concerning the “number of active enterprises” the figures from BD are in all cases larger than in SBS. The only exception is in section L, where again the micro-enterprises make the difference. This is caused by the fact that private landlords may choose between two taxing schemes. Either they choose turnover-tax (as businesses) or income-tax (as household performing private wealth management). Therefore, since reference year 2015 also households opting for turnover tax are equally excluded in both, SBS and BD.

*2.7 Misclassifications respectively outdated NACE codes (over-coverage)*

Despite the accelerated production of the sampling frame (BR) and the harmonised treatment of unit-non-responses mentioned above, an unsolved issue remains in the treatment of misclassifications in the sampling frame. The NACE code in the administrative data, which is the source for the national BR, is not always up to date. Most enterprises announce an activity at the time of their registration, but do not declare any changes to their activity portfolio in the course of their existence. Therefore, changes in the main activity of enterprises are sometimes not detected until a survey questionnaire is rejected. This results in a number of justified unit-non-responses, which are classified as over-coverage in the sampling frame. The unit itself is reassigned to the new NACE-code in the BR and consequently analysed in BD. Yet an incoherence between SBS and BD results from the fact that all “represented enterprises” (represented by the expansion factor of the sample unit) cannot be followed up in the BR. Also, if the new NACE-code was outside the respective domain statistic, it could not be forwarded to the new domain, because it had received the wrong questionnaire. Consequently, these justified non-responses (due to over-coverage of the sampling frame) were taken out of the sample without appearing anywhere else. This led to a widening gap between SBS and BD.

**3. Further improvements**

For most of the above mentioned causes of incoherence a solution is agreed and their implementation has progressed well (for example the acceleration of the BR).

Furthermore, two ongoing projects in Germany will have an additional substantial impact on reducing the deviation: The implementation of the upcoming Framework Regulation Integrating Business Statistics (FRIBS) and the full implementation of the Statistical Units Regulation 696/93 (SU-Regulation).

*3.1 FRIBS*

According to the current plan, the implementation of FRIBS in Germany will lead to a reduction of legal acts and stovepipes. The national legal environment will then consist of only two legal acts for all business statistic surveys (including short term statistics): one for industry (NACE B to F) and one for services (NACE G to S, excluding[[2]](#footnote-2) K64, K65, O and S94). Also the IT-production environments will be harmonized within these two legal environments. Therefore, a misclassification in the administrative data will not necessarily result in a unit-non-response anymore.

Another benefit of FRIBS is that it will explicitly exclude non-market units from the scope of both, SBS and BD.

*3.2 SU-Regulation*

In German business statistics the *legal unit* is currently used as equivalent to the *enterprise*. Starting with reference year 2018 this will be substituted by the definition of an enterprise in the SU-Regulation 696/93: *“The enterprise is the smallest combination of legal units that is an organizational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources. An enterprise carries out one or more activities at one or more locations. An enterprise may be a sole legal unit.”*

To apply this definition, the enterprise will be implemented as presentation unit, while continuing to address legal units in the surveys. In order to get from the legal unit to the enterprise two additional steps are included in the production system: “imputation” of legal units that are not surveyed, but belong to a complex enterprise, and “consolidation” of all legal units that form a complex enterprise together. For these two steps all domain statistics deliver validated micro-data to one central data warehouse. All following compilation steps will be based on this consolidated data set. These steps will either include a re-assignment of expansions factors and grossing up or they include an exclusive imputation of SBS variables into the full set of micro-data in the BR.

Furthermore, all domain statistics will extract their sample frames at the same date and a unified sample is drawn by one central unit. Altogether, this will effectively overcome the current stovepipe system and improve coherence between BD and SBS.

**4. Remaining issues**

By the various initiatives, the current deviations should be reduced substantially in the upcoming years. On the top level aggregates they will most certainly be close to zero. However, some remaining issues will be discussed in this chapter.

*4.1 Misclassifications*

Now that legal and practical issues of misclassifications will hopefully be solved in the mid-term future (only two legal acts and only two production environments) a methodological question arises on how to treat misclassifications.

It is easy to re-assign a sample unit and follow up on this in the BR, respectively in BD. But the sample unit has an expansion factor that represents a certain number of enterprises, which are not specifically appointed. Should the reassigned enterprise take its expansion factor along? If so, which enterprises should be reassigned? And given the fact, that BD is publishing results down to the NUTS 3 level, how do you ensure the quality of regional results, which are below the sample stratification?

Even more complicated is a situation, where the expansion factor is very different between the two sample strata concerned. If for example a unit with an expansion factor of 200 is moved to a sample stratum with an expansion factor of 10.

There is no solution in Germany to the treatment of misclassifications in this respect yet. Consequently, the gap will not be removed completely on all breakdowns by the range of initiatives presented so far.

*4.2 Calibration*

Certainly any gaps between two data sources can be removed by calibration of results. It also improves the quality of the sample. In general there are two categories of calibration methods: regression based methods that find specific adjustment factors for each observation and ratio estimation methods that find common adjustment factors for all units in a certain aggregate (for example sample strata or publication breakdowns). Both categories provide benefits and drawbacks. But in any case they fully rely on the accuracy of the benchmark.

Ratio estimations preserve the internal structure of the sample, but are either crude or very labor intensive (depending on the granularity of the adjustment). Furthermore, if the internal structure is to be strictly preserved, these methods calibrate only one variable. This is very problematic, if some other variables are not correlated.

Regression based approaches can handle more calibration variables (though the amount of benchmarks is still limited). But they change the internal structure of the sample. This can cause a bias and lead to implausible results of secondary analysis (i. e. any results that were not calibrated). Also, to improve the timeliness of our publications, most validation procedures focus on the influential units. These used to be identified according to their size relative to the lowest publication breakdown, they were included in. If the calibration is applied afterwards, these intensively validated units might not have any weight anymore (in fact their weight could even be negative in some methods) and other – in foresight considered neglectable – units might weigh tenfold.

Concerning the issue of coherence between SBS and BD, it is very difficult to decide on the better benchmark: Calibrating SBS on BD-benchmarks will bias the NACE-structure of the sample (because it is calibrated on the partly outdated NACE codes in the administrative data). This means that valuable insight into structures of the business economy would be blurred again. Furthermore, BD (respectively the BR) can only provide turnover and employment of the units. Unfortunately, these two variables are not fully correlated with all other variables of interest in SBS (for example investment). Next to that, once the calibration is performed, secondary use of the microdata might be biased. This clearly reduces flexibility in promptly reacting to newly arising user demand.

Using the SBS-results as benchmark for BD is difficult, because BD needs to provide results on a much lower regional breakdown (NUTS 3). The same applies for the very small subsets that are detected (Gazelles etc.). Both aspects (regional and functional breakdown) cannot be benchmarked on SBS. Also, in both cases the resulting aggregates (small as they may be) must contain full units. A calibration according to SBS-benchmarks might lead to decimals and make additional rounding necessary. Furthermore, the above mentioned question on how to choose “represented” enterprises in the case of outdated NACE-codes in the sample is unsolved.

Finally, the sheer amount of un-comparable breakdowns in SBS and BD certainly poses a problem for any calibration methods and/or approaches.

**5. Questions for discussion at the Q2018**

* How are your experiences with calibration?
* How do you treat misclassifications?
* Can we have it all: minimal bias, maximal flexibility and identical results in SBS and BD on every breakdown?

1. t = reference period [↑](#footnote-ref-1)
2. Section O “public administration and defence” and Division S94 “activities of membership organisations” are excluded from the SBS scope. “Financial and insurance activities” in divisions K64 and K65 are outside these legal acts and production environments because administrative data from bank supervision (Deutsche Bundesbank) and insurance supervision (BaFin) can be used for SBS. [↑](#footnote-ref-2)