Federal Statistical Office, Germany
E101/E102 – Business Register

17.08.2011

Joint UNECE/OECD/Eurostat Meeting of the Group of Experts on Business Registers, 14-15 September 2011, OECD, Paris, France

Session 3 "Reducing respondent burden - the role of the business registers"

The role of the Business Register in the use of Admin Data for Short Term Statistics

by Robin Lorenz / Roland Sturm

1. Use of Admin Data for STS in Germany

In 2003 the "Administrative Data Use Act" was passed in Germany. , It obliged the Federal Statistical Office of Germany (FSO) and the statistical offices of the *Länder* to examine the suitability of administrative data for different purposes of economic statistics. The focus was on tests in the field of short term statistics (STS) with the aim to reduce the response burden of the enterprises and improve the quality, if possible. For this purpose the statistical offices in Germany got monthly access to two administrative data sources: monthly VAT-turnover files held by the tax authorities of the German states (*Länder*) and monthly employment data from the Federal Employment Office on persons in employment liable to pay social insurance contributions and on persons with minor employment (further on referred to as "Admin Data").

Extensive testing showed that because of the peculiarities of each economic sector examined with respect to the structure of the enterprises and because of the different demands (international and national) in terms of timeliness and periodicity a general suitability of the examined Admin Data for STS could not be stated. Instead individual models of using Admin Data for the various economic sectors have been suggested. Table 1 gives an overview about the outcome of the tests and the current or intended use of the Admin Data in the various economic sectors examined:

Table 1: Use of Admin data for STS: Current situation in Germany

Economic sector	NACE Rev. 2	Time-liness	Perio-dicity	Usabi-lity	Additional survey	Begin of use
Transport, storage, communication and other business services	H, J, parts of M and N	t+60	quarterly	yes	TO* › 15 Mill. € or PE** › 250	2007
Crafts	----***	t+60	quarterly	yes	Solely Admin data	2008
Building installation and completion	432, 433 (part of F)	t+60	quarterly	yes	PE › 20	2011
Trade of motor vehicles	45	t+60	monthly	yes	TO › 10 Mill. € or PE › 100	2012
Wholesale trade	46	t+60	monthly	yes	TO › 20 Mill. € or PE › 100	2012
Accommodation and food service activities	I	t+60	monthly	no	----	----
Retail trade	47	T+30	Month	No	----	----

* Turnover per year; ** Persons employed; *** Not defined by the NACE, but by the enrolment in the chamber of crafts;

The table shows that the Admin Data can be used for a wide range of STS-purposes in Germany. The implementation of Admin Data-based statistics started in 2007 with the economic sectors transport and storage, communication and other business services, for which a survey for the largest enterprises is combined with Admin Data for the small and mid-sized enterprises. One year later the STS-survey in the field of crafts was fully replaced by the use of Admin Data. In both cases quarterly results with a timeliness of t +60 are required. The same timeliness and periodicity is valid for the sector building completion and installation, where at present a combination of a survey of the large local units and Admin Data use is put into practice. Finally in 2012 another two models with mixed sources for the sectors wholesale trade and trade of motor vehicles are intended, but with a timeliness of t+60 and a monthly periodicity, which raises the additional issue of dealing with the VAT- turnover of quarterly payers. In the sector accommodation and food service activities, where quarterly payers have a substantial proportion of the turnover, attempts to break the quarterly turnover down to months failed so far, so the Admin Data will not be used in the NACE section I. Also in the retail trade sector Admin Data can not be used, as 30 days after the end of a month the VAT data is too incomplete.

Even though in most applications of the Admin data the initially hoped-for full replacement of the survey was not possible, a substantial reduction of the response burden in particular of small and mid sized enterprises could be achieved. In the service sector the number of surveyed units decreased from 37 000 to 4 500 units per quarter, in the craft sector all the 41 000 units per quarter could be unburdened and in the sectors wholesale trade/trade of motor vehicles 6000/3000 enterprises per month will be released from the surveys. Meanwhile, in view of this considerable reduction of bureaucracy the "Admin Data Use Act" has been revised to allow for the Admin Data to be used on a permanent basis for purposes of STS and other economic statistics.

2. The role of the Business Register (BR)

Admin Data in general is collected for non-statistical purposes (administrative purposes) and therefore potentially inadequate for statistical use. This was proved in the case of the examined Admin Data in the tests described in chapter 1. Due to deviations from required statistical concepts in the Admin Data the STS-results can not be generated by using solely the data of the two Admin Data sources.

Here the Business Register (BR) comes into play: It helps employing the Admin Data in two ways. Firstly the BR serves as central tool which enables the linking of the administrative data from the two sources with the data from surveys of the large enterprises. Furthermore the Business Register helps to deal with some deficiencies in the Admin Data concerning the activity code and the definition of statistical units with respect to the requirements of STS.

2.1 The BR as central tool in the editing process of the Admin Data

The BR is updated on a yearly basis with Admin Data from the same sources which are used for STS-purposes - VAT data from the tax authorities and employment data from the Federal Employment Office. In addition, structural data from sources such as the Chambers of Commerce and the Chambers of Crafts and information from different surveys is also integrated into the BR. As almost each of the various sources has its own numbering system and therefore no common identifier exists, the linking of the records of all these sources on micro level requires automated record linkage techniques using names and addresses as main identification variables. However, this is a time-consuming process, which must be accompanied by manual interventions to a large extent.

Apart from the integration of the data of several sources the BR also has to deal with different unit concepts among administrations and the concepts of statistical units employed by official statistics. The VAT declarations refer either to enterprises[1] or to VAT groups, a combination of several enterprises for reasons of simplification of the taxation. In turn, the Federal Employment Office delivers the data on the level of local units[2]. The BR staff expends big efforts with manual interventions to detect the relations between VAT groups, enterprises and local units. Because of these labour-consuming processes and the use of annual data the complete picture is finished about 18 months after the end of the reference year. The result is once a year a so called time slice of the BR which is a nearly complete representation of the German enterprise landscape referring to a certain reference year. In the monthly editing process of the Admin data for STS always the most up to date time slice is used for the generation of STS-results within one calendar year.

The fact that the BR contains data from the same two Admin Data sources which are used for STS, enables the automated linking of the current monthly Admin Data with the BR. In case of the VAT data the key variable is the tax number, for the employment data it is the local unit number. The automated linking of both current Admin Data sources with the BR implicitly results in linking the current VAT data and the employment data on micro level, which otherwise would be impossible because of the lack of common identifiers. However, due to the time lag of the BR the linking of the monthly Admin Data with the BR succeeds only for those units of the current Admin Data delivery, which have already been active in the reference year of the BR and still have the same identification number. Relatively new records in the current delivery can not be found in the BR, because they are not included in the BR yet. The successful linking of the majority of the current records in the Admin data is essential when it comes to the generation of STS- results in the different economic sectors.

Two special cases shall be tackled now, as the BR plays the decisive role when it comes to the determination of the units which are incorporated in the results. This illustrates the decisive role of the BR as the central tool in the editing process of the Admin Data:

- *Mixed Models:* In the mixed models, which combine a survey of large enterprises and Admin Data for the rest, the selection of the surveyed enterprises is based on the yearly turnover value and the number of persons employed respectively, in the reference year of the most up to date time slice of the BR. Each enterprise which exceeds the specific threshold (see table 1) and, hence, is selected for the survey, is flagged in the BR. In course of the editing process the current Admin Data is linked to the respective time slice of the BR. If a record of the Admin data delivery is linked to a flagged enterprise in the BR, then the values from the Admin Data sources are excluded from the statistical calculation, and the values of the survey are taken instead.

- *Crafts sector:* Also in the crafts statistics, the only economic statistics where solely Admin Data is used, the BR plays an essential role. The information whether an enterprise belongs to the crafts sector is not delivered by the two Admin Data sources. The information stems from the Chambers of Crafts, one of the sources which is processed in the BR. Only the enterprises which are flagged in the BR as crafts enterprises and which have current values in the Admin Data are incorporated in the STS-results for crafts. Thus, in case of the crafts statistics exclusively the BR determines the scope of the potential reporting units. New enterprises in the current Admin Data deliveries, even though they might be crafts enterprises, can not be

[1] Enterprise in the sense of the tax administration means taxpayer and does not necessarily correspond to the statistical enterprise concept in the BR.
[2] Local unit in this context means the unit which pays the social insurance contributions of the employees and does not necessarily correspond to the local unit concept in the BR which applies to units where employees actually work.

incorporated in the STS-results for crafts as they are not integrated in the BR yet, and hence, the information about their belonging to the crafts sector is not available.

2.2 The BR as source of additional information to reduce deficiencies in the Admin data

Apart from the linking issue the data delivered by the Admin sources is not fully in line with the statistical requirements according to the STS-regulation. The weaknesses include fundamental aspects as the concepts of the variables turnover and employment, the assignment of units to economic activities and deviations between administrative and statistical units. The BR contains additional information, which helps to compensate at least partially some of these deficiencies and increase the quality of the STS-results.

- *Assignment of the economic activity:* Both the tax agencies and the Federal Employment Office assign an activity code to each unit. However, these activity codes do not entirely meet the statistical requirements. It is less a problem of standardisation of the classification, because the classifications used are more or less the same and diverge only at some 5-digit positions. The difficulties statisticians have to deal with arise from the procedures how the activity codes are attributed by the administrative bodies and from the fact that the frequency of updating them does not correspond to the statistical requirements. High rates of misclassification and differing codes between the Admin sources are a consequence. A comparison between the surveyed industrial classification codes (NACE) - predominantly from annual statistics - and those from the administrative sources showed considerable discrepancies. On the two-digit level of NACE about 20% of the units have different activity codes in the administrative sources and in the surveys. In some fields of economy, in particular on lower level of aggregation, significantly higher shares of units with deviant activity codes can be observed taking the surveyed codes as reference.
  As mentioned before, the BR processes data from various sources, of which the most also include the information on the economic activity. Among all the different activity codes available for an enterprise one is selected to be the correct economic activity which is then stored in the BR. In general the activity code from a survey is considered to be the most reliable information. Due to the sampling schemes, bigger enterprises are usually covered to a higher extent by statistical surveys than small and medium-sized enterprises, so that surveyed activity codes are available for a high percentage of the turnover and the persons employed. If no surveyed activity code is available, the BR relies on the Admin data sources. Of these the quality of the activity codes from the Federal Employment Office is considered to be slightly higher than the one from the tax office. In any case it is assumed that the BR contains the activity code with the highest quality under the given information.
  When editing the monthly Admin Data for STS-purposes the activity code of the BR is assigned to each unit which can be linked with the BR. If the linking fails there is no alternative but to use the activity code of the respective Admin source for the generation of the STS-results. This leaves a certain quality issue in the Admin Data with respect to the activity code and restricts the level of disaggregation for potential publications. Even so, due to the adoption of the BR activity, the quality of the activity code in the current Admin Data is no vital obstacle for the use of Admin Data for STS-purposes.

- *Deviations in the Admin Data from required statistical units:* In the economic sectors where Admin Data is used for STS the results should be generated on the basis of enterprises according to the STS-regulation. Both Admin Data sources, however, exhibit deviations from the enterprise concept.

In the case of the already mentioned VAT groups only the controlling unit declares the total turnover to the fiscal agency, but no information about the turnover of the individual members of the VAT-group is available. In order to avoid that the total turnover of a VAT group is assigned to the economic activity of the controlling unit, the total turnover must be broken down to the individual members of the VAT group. The BR deals with this problem by estimating the turnover of the reference year of VAT group members by means of a multiple regression model. This is a complex procedure including manual control mechanisms, which can be applied in the longer time horizon of the BR, but not in the short time frame between the delivery of the current Admin Data (t+50) and the generation of the STS-results (t+60). Therefore in the Admin Data use a different procedure is in practice: An indicator is derived from the BR estimations of the annual turnover of the VAT group members. This indicator gives the percentage of the turnover that is allocated to each VAT group member and is applied to the current monthly or quarterly turnover of the VAT group. Hence, the estimations in the BR are the input for further estimations in the current Admin Data for STS. A weakness of the splitting method is that the indicator refers to the reference year of the BR, which is two years old at the time of editing the Admin Data. Changes in the composition of the VAT group since the end of the BR reference year are not (or only to a small extent) taken into account. This solution is not perfect, but considering the extent of the problem – at the federal level over all branches VAT group members make up 45% of the turnover, but for example in the fields of air transport and communications their share can reach up to 85-90% - it is a quick and a completely automated method which leads to current turnover values for the VAT group members in sufficient quality.

3. Conclusion

The BR is the crucial instrument in the editing process of Admin Data for STS-purposes. Having the monthly VAT data and employment data as main data sources the additional benefit of the BR compared to the pure Admin data sources is that the BR contains information from various sources (Admin data, surveys including best activity codes available and data from Chambers) matched together on micro level with an adequate treatment of the different unit concepts. By linking the current Admin Data to the BR these additional benefits of the BR are directly transferred to the current Admin Data. Thus the STS-results generated with the current Admin Data correspond to a large extent to the required statistical concepts for STS. By this means the BR contributes to the quality of the STS-results.

However, the advantages of the BR compared to the pure Admin data sources can only be achieved by time-consuming processes, which is the reason why the reference year of the current BR is past for 18 months. Due to this time-lag the additional BR information can only be transferred to those units which are already integrated in the BR. That means current developments with respect to changes in the composition of VAT group or in the relations of enterprises and local units can not be taken into account. This reduces the benefit the BR adds to the current Admin Data and leaves some uncertainty in the Admin Data.
In the near future the further development of the BR is intended. A new BR data bank software will be implemented and the processing routines will be changed to a sub-annual integration of data sources. It can be expected that this will reduce the time-lag of the BR and make the BR an even more valuable tool in the editing process of the Admin Data.