EUROPEAN COMMISSION
EUROSTAT

Directorate G: Global Business Statistics
Unit G2: Structural business statistics and global value chains

UNSD Friends of the Chair Group

Paper on a new data infrastructure: the Micro data linking approach in European business statistics

1. Introduction

European business statistics compilers often face a dilemma: On the one hand, user and policy requirements demand additional information on the structure and development of European enterprises. On the other hand, budget constraints and the reluctance to increase the burden on survey respondents and national statistical institutes put tight restraints on the enlargement of data requirements.

Micro data linking can provide an opportunity to discover new information and to develop new statistics and indicators both when using existing data sets but also when combined with new data collections. In the European Statistical System (ESS)[1] one of the initiatives to investigate into this opportunity for business statistics was a project concentrating on linking the international sourcing survey with Structural Business Statistics (SBS) and International Trade in Goods Statistics (ITGS) in 2010. This was succeeded by a follow-up exercise launched in late 2012 which broadened the scope of the 2010 exercise.

This paper will discuss in detail the approach of the 2010 exercise and will use it as an example for requirements and expectations of using microdata linking in the context of international business statistics.

2. Linking of International Sourcing (IS) Survey data with SBS and ITGS data

The IS was a survey conducted in 2009 focusing on establishing the extent of the movement of business functions (core or support) which were previously performed in-house to a foreign location. The questions asked were mostly qualitative in nature focusing on whether the enterprises had outsourced/insourced during the period in question, to which region they had outsourced and what their motivations had been. By linking the resulting data to the SBS and ITGS statistics, it was made possible to profile the enterprises engaging in

---

[1] The ESS is the partnership between Eurostat and the national statistical institutes (NSIs) and other national authorities responsible in each EU Member State for the development, production and dissemination of European statistics. This Partnership also includes the EEA and EFTA countries.

international sourcing by looking at the employment development in, or the value added by those enterprises in contrast to a control group of enterprises not engaging in international sourcing.

When attempting to link two or more sets of data, two things are vital for a successful exercise:

1. There has to be a large enough intersection of responding units. In the case of the IS – SBS exercise, this was achieved, because the IS sample was selected as a subset of the SBS population, which means that in theory there should have been a matching SBS data set for every IS data set.

2. There has to be a unique identifier or at least a very reliable matching approach. Since both the IS and the SBS use national business registers (NBR) as a frame, this was the case for the IS-SBS linking. However, ITGS is not directly based on NBRs, so VAT numbers were used to link ITGS data sets to the NBR and thereby to IS/SBS.

The IS data was not only linked to the latest SBS data (2008 data at the time of the exercise) but also to earlier records going back to 2000. Since business demography events like the creation, termination or acquisition of enterprises may affect the NBR populations over time, the matching rate was the highest for 2008 data (>95% in all participating countries) and decreased for each step back in time (65% - 85% for 2000) for the group of enterprises engaging in international sourcing.

To be able to benchmark the stratum of enterprises engaging in international sourcing, it was important to designate a control group. A simple approach would have been to use all the enterprises in the linked dataset not engaging in international sourcing, however, this may have led to a skewed result, as the structure of the control group in regard to other variables like enterprise size or turnover may have been very different from the structure of the group of sourcing enterprises. The approach taken by the taskforce was therefore to construct a control group of non-sourcing enterprises from the survey population with the same characteristics as the sourcing one in terms of economic activity.

Since confidentiality is an important issue in the field of European business statistics, it is important to note that the linked micro data files were stored locally at national statistical offices throughout the project and were not shared with third parties. It was therefore important to decide on concrete research objectives before the exercise and to design the linked data set accordingly (i.e. decide on the variables to retain and on the aggregates to be produced). To achieve a harmonised output, the participants were provided with precise data requirements and standardised guidelines explaining in detail how the linked datasets in each country were to be structured. The computer code to produce the tables was provided as well, to ensure that identical tables were constructed in all countries.

The exercise provided the ESS with important insight into the impact of international sourcing on employment, economic performance and international trade patterns and was therefore considered a success. It was repeated with a wider scope in 2013, including linking to inward foreign affiliates statistics (iFATS). The basic approach remained unchanged from the one described here.

3. Other MDL projects

Following the successful MDL exercises in 2009 and 2013, a new exercise will be carried out in 2014-2015 by 9 member states, using input data for the reference period of 2008-2012 from Structural Business Statistics, International Trade in Goods Statistics, International Trade in Services Statistics, Community Innovation Survey, ICT usage and e-Commerce in enterprises Survey, Foreign Affiliate Statistics (Inward and Outward), Business Demography Statistics, International Organization and Sourcing Survey and the national Business Registers. The SBS enterprise population will be analysed and classified into sub-populations according to their involvement in international trade in goods and services, how innovative they are, how intensive their ICT usage is, whether they are active in GVC/IS, whether they are foreign affiliates or whether they control affiliates abroad etc. This will help to better understand drivers for business competitiveness and productivity.

Micro data linking can also be used to improve the quality of existing statistics. In 2013, under the umbrella of the ESSNet on Measuring Global Value Chains, the NSIs of Denmark, Norway and Finland have linked statistics on the activities of an economy's affiliates based abroad (oFATS) with statistics on foreign affiliates resident in the local economy (iFATS). Since iFATS is based on administrative data while oFATS information is collected by a survey, iFATS quality is generally assumed to be superior. The approach taken was to mirror iFATS and oFATS data sets between the countries, where control was exerted from an enterprise resident in one of the three countries and the affiliate was located in another. In theory, this approach should have resulted in an identical set of affiliates in iFATS and oFATS; however the exercise showed that there are some discrepancies between the two statistics and gave important leads for the improvement of FATS data quality[2].

Apart from the three GVC/IS MDL exercises and the FATS linking project described above, the ESS has engaged in several other MDL exercises. There has been an ESSNet on Linking of Micro data on ICT Usage, where business registers were linked with SBS and ICT usage and e-commerce data[3]. Furthermore, several trade related exercises, namely on trade in goods and trade in services were conducted. For these exercises, trade data was linked with business registers in order to create a breakdown by enterprise characteristics. Finally there has been an ESSNet on data warehousing and MDL which touched more on theoretical

---

[2] More information can be found at http://www.cros-portal.eu/content/mirroring-foreign-affiliate-statistics
[3] See http://www.cros-portal.eu/content/esslait and http://www.cros-portal.eu/content/esslimit-finished

aspects[4]. The following table provides an overview and shows which countries were involved in each round:

	ICT Impact (LAIT/LIMIT)			International sourcing			Trade					DWH	FATS	Count	
	2006-08	2010-12	2013	2010-11	2012-13	2014-15	TEC regular	TEC Meets 2009	TEC Meets 2011	TEC Meets 2013	sTEC voluntary	DWH	iFATS vs oFATS		
BE							x							1	
BG							x							1	
CZ	x			x			x				x			4	
DK	x	x	x	x	x	x	x		x	x	x		x	11	
DE	x	x	x			x	x	x						6	
EE							x		x	x		x		4	
IE	x	x	x		x		x				x			6	
GR							x	x						2	
ES							x							1	
FR	x	x	x		x		x							5	
HR														0	
IT	x	x	x	x			x		x			x		7	
CY							x							1	
LT					x		x					x		3	
LV						x	x	x						3	
LU		x	x				x				x			4	
HU							x	x						2	
MT							x							1	
NL	x	x	x	x	x	x	x			x	x	x		10	
AT	x	x	x			x	x			x	x			7	
PL		x	x				x			x	x			5	
PT				x		x						x		4	
RO		x		x	x		x							4	
SI	x	x	x	x			x							5	
SK							x		x					2	
FI	x	x	x	x	x	x	x						x	8	
SE	x	x	x		x	x	x					x		7	
UK	x	x	x				x				x	x		6	
NO	x	x	x	x	x	x	x			x			x	9	
TOT	**13**	**15**	**14**	**9**	**9**	**9**	**28**	**4**	**3**	**6**	**9**	**7**	**3**		

4. Conclusions

- Micro data linking has significant potential as a way of gathering new statistical evidence without increasing the burden placed on respondents;

- it can be an additional way of ensuring data quality and consistency between related data sets;

- a coordinated approach to micro data linking is a cost-effective way for national statistical offices to undertake micro data linking and ensures harmonised, comparable results across countries;

---

[4] See http://www.cros-portal.eu/content/micro-data-linking-and-data-warehousing

- micro data linking gives new insights into the behaviour of enterprises engaging in international sourcing and the impact of international sourcing on European economies;

- by linking various data sources, it is possible to conduct regression analysis, and by extension, other types of statistical analysis such as multivariate analysis and cluster analysis;

- The potential for micro data linking to increase the usefulness of a data collection should be factored in when planning the exercise. MDL is a useful and powerful tool in combination with new surveys; if it is factored in in advance that survey results will also be used for micro data linking, the survey questionnaire can be limited to questions for which data is not available from existing sources.