

USE OF THE ADMINISTRATIVE DATA IN STRUCTURAL BUSINESS STATISTICS IN POLAND

Paper prepared by
Jacek Kotowski and Mirosław Stepień
Central Statistical Office of Poland

Summary

In the paper we present basic information on the current sources for Structural Business Statistics, as well as on the main administrative registers in Poland and their usefulness for SBS. Additionally, we present some results of methodological works recently undertaken in the Office, and their impact on improvement of final results of the survey, by using data from administrative sources.

Introduction

The issue of the structural business statistics covers a broad area, which characterizes the economic results of the national economy's entities operating on the market commercially. A number of requirements defined by the Commission's regulations, regarding the scope as well as characteristics, inflicts both the choice of units for observation (certain types of statistical units, such as enterprise, local unit or kind of activity unit) and the aggregates to be elaborated and delivered to Eurostat by the members states. This requires an application of appropriate classifications describing e.g. types of business activities or enabling to group the results according to the size classes of enterprises, the size classes of the store's selling area, etc.

A primary requirement for every statistics, including the European structural business statistics, is its good quality. One of the elements, and at the same time the factor conditioning a success of the entire project at the European level, is the states' assurance of a cohesive scope and content of this statistics as well as an application of the same definitions while considering specific variables, which allows for comparing the data between individual member states.

The European law does not impose on the member states the way they should collect and elaborate data that are essential for the preparation of the files to be delivered to Eurostat. It only suggests that, as the experience of many states shows, a traditional statistical reporting may be in many cases substituted by various types of administrative sources. It is often emphasized that such an approach allows for diminishing the costs of the statistical survey, and at the same for decreasing the effort that respondents must put in the statistical reporting. A primary question that should be asked, when considering the adoption of such data collection system, deals with conditions that should be met in order for it to be implemented and become the best solution from the point of view of the statistical tasks.

The subject of this paper is a discussion of the ability to use administrative sources for Polish statistics, exemplified by particularly broad structural business statistics as well as the experience of the Central Statistical Office (GUS).

1. The Sources of Information for the Structural Business Statistics

Up to now, the primary source of data for Polish structural business statistics has been the statistical surveys conducted by the Central Statistical Office (GUS). The basic annual information on the range and effects of the business activity of enterprises are collected by means of exhaustive surveys for the entities with 10 or more employed persons, and the representative surveys of micro-enterprises, i.e. those with 9 or less employed persons. The range of the collected information in the former type of the above surveys is very detailed, in the latter one - quite limited.

As a result, a complete – accordingly with the needs of structural statistics – use of data from the representative survey requires an application of the optimal stratification of the sample estimate as well as of the advanced statistical methods, in order to develop necessary data, estimate the generalized data into the required detailed aggregates and select more specific data impossible to obtain (due to the existing regulations regarding book-keeping) directly from the units.

Moreover, the data coming from the number of other statistical surveys and constituting a complement source for the required information, should be added to the above mentioned surveys (including the more detailed exhaustive survey for larger entities).

Another operation that must be conducted on so collected data aims at their cohesion and unification, which in turn should allow for collecting all data in one database of the structural business statistics being the basis for the calculation of the data files required by Eurostat.

As it stems from the above, Polish structural business statistics is developed in whole on the basis of various statistical surveys, without reaching for the administrative resources. It should be emphasized here that the law on public statistics, in force from 1995, constitutes a sufficient legal basis for collecting the data from various administrative resources by the statistics. In practice, until now, there have been no regular exchange of information, and the works conducted by the statistics were methodological works, within which the potential sources of administrative data and possibilities of their use for the need of the enterprise statistics were defined. The effects of these works and the conclusions are presented below.

2. Administrative Registers in Poland

It should first of all be observed that in Poland, until now, various institutions have created separate registers. They use them for their own needs and these registers have never been established as one integrated system of administrative registers, where every national economy's entity would be treated in the same way. This resulted in a situation where it is hard to identify the same units which may be found in different registers. What is more, the unit functioning as an entity in all registers is defined individually in each of them. In consequence, it may happen that in order to obtain information for one entity registered in certain register, it is sometimes necessary to accumulate data for several units registered in other register. Hence, in spite of the fact that the system of administrative registers in Poland is a rich source of information on the national economy's entities registered there, it is frequently not possible to directly use it for the need of statistics. As it has already been mentioned, the works in the statistics have recently been intensified both over the development of tables of correspondence between specific registers, so the information could be easily assigned to the same entities coming from different registers, as well as over the methodology on the development of methods of combining data and processing them in a proper way, so they could be further used by statisticians.

In this paper, selected register systems, the most important for enterprise statistics due to the range of collected data, will be discussed. The systems are as follows:

1) The Register of the National Economy's Entities

A primary administrative register used by Polish statistics for updating its resources (including the statistical register) is conducted by the President of the Central Statistical Office (GUS) national official register of the entities of the national economy – REGON. It is obligatory for all legal persons, organizational units without legal status and natural persons conducting business activity to register. Each of the registered entities, as well as the local entities where the business activity is conducted by at least one person, are assigned REGON number - a unique identification number - and legal attributes defined by law. This register does not cover typical statistical units, such as enterprise or kind of activity unit which should be the entities of the statistical register. At present, the REGON register is closely integrated with statistical systems and constitutes the primary source for the statistical register's update within, for example, newly established economic units, liquidation of business activity or changes of the type of conducted business activity.

On the basis of information collected in the REGON register, it is possible to integrate the whole systems of registers kept in Poland, as the REGON number is, next to the NIP (Tax Identification Number), obligatory for all registers.

2) Register of Taxpayers

The objective of the system of registering taxpayers is collection, verification and processing of data on the entities of the tax system run for the needs of the public administration, and specifically of the tax administration. The system functions mainly in the form of regional registers run by regional tax offices, and covering the taxpayers residing in the area specific for them. Basic databases on taxpayers function as central registers. One of them is the National Taxpayers Evidence (KEP) run by the Minister of Finance, as the law states. The KEP register's role, apart from its primary registering function, is to compare data submitted on the identification applications of taxpayers with other official registers in order to select those taxpayers who violated the law's provisions.

The register covers, among others:

- natural persons conducting independently business activity;
- legal persons;
- organizational units without legal status;
- the establishments of legal persons who are treated as independent taxpayers in separate regulations.

The register includes information allowing for identification of the income tax payer by assigning each of them the NIP number (Tax Identification Number) and, additionally, the REGON number, as well as by entering identification data. As to the establishments of legal persons, the identification data of the mother enterprise are also recorded. The data range of the fiscal system is in accordance with the data collected on the tax forms, but categories are frequently not cohesive as to definitions.

The VAT payers' register, particularly important for enterprise statistics, is closely connected to the National Taxpayers Evidence (KEP).

3) The Register of the Insurance Premium Payers

The register covers all social insurance payers. The statistics deals with the data on the number of the insurance premium payers running business activity. The information in the

system is usually updated on a current basis by entrepreneur directly via the program distributed by the Social Insurance Company (ZUS) which runs this register.

The database includes information allowing for the payer's identification, such as: abbreviated name, ID type and its number, first name, second name and surname, place and date of birth, citizenship (for natural persons) or the name/company compliant with the constitutional act of an entity, and the name of the founding body (for legal persons). Moreover, in the register, there are collected data which help in the payer's identification in other registers, such as NIP, REGON and PESEL (National Register of Population) Id numbers. However, these fields in the forms are not always filled in.

The statistics is interested in obtaining information helpful when determining the activity state of a payer, i.e.: the start date of a business activity, the start date of the obligation to pay social security premiums, the reason and date of the company's opting out of the register, as well as the information regarding the amounts of paid insurance premiums, the number of employees employed by the payer and the employment period.

3. The Administrative Registers as a Source of Information for the Enterprise Statistics

After the analysis of the registers being the source of information for the enterprise statistics (particularly for the structural statistics), two types of the key issues, which occur when exploiting the administrative data, may be specified:

- 1) the range of the units for observation and the database of their variables are different in each administrative register;
- 2) information on registered units, particularly regarding their activity level, type of the business activity they run, and the basic data allowing for defining the size of the unit, are updated in each register independently, which leads to them not being unified.

The Range of Variables

The range of variables collected within each of the register systems for specific entity is in close connection with the objectives of certain register. It should also be observed that the range of data collected in each register is frequently adjusted to legal requirements that the

entities in the register are subject to. A good example here may be the resources of the Ministry of Finance, divided into the number of databases covering only selected types of units, VAT payers: legal persons and natural persons (with tax declarations of various types). Another problem is, from the point of view of statistics, the data deficiency in some resources. For example, the information on employees by their main workplace, or on all remunerations paid out to the employees in certain enterprise, may not be retrieved from the ZUS register. The ZUS only keeps information on those employees who pay obligatory social benefits. As to the tax registers, the statistics cannot retrieve from there any complete balance sheet data and the loss and profit account, as they are required in statistics.

The Units for Observation

When comparing the register system of public statistics with administrative registers, it should be taken into consideration that in every register the same unit is covered in a different way than it is in the statistics. The reason for such a diverse approach to registering units is the objective each register has. As an example of different approach to registering a unit in various register systems may be the civil partnership of natural persons. In the REGON register, it will cover the business activity conducted by all partners in this partnership, while in the tax registers of the Ministry of Finance, it will focus on either one: as to the income tax, all partners are obliged to individual tax settlement; as to the VAT tax, the partnership is treated as one economic entity.

The difficulties that may come across when using the administrative data by the statistics stem from:

- not unified classifications of specific legal forms;
- differences in identification of the registered entity;
- lack of information on the type of conducted business activity or the business activity currently conducted.

For example, the differences regarding the way a natural person is registered in specific registers originate from the fact that each register registers the same natural person under different feature or function. In the KEP register, this natural person is called “a natural person running a business activity independently”, in the ZUS register – simply a “natural person”, and in the REGON register – “a natural person running a business activity”.

Another example regards the rules defining legal persons in the tax systems:

- In the National Taxpayers Evidence (KED), a legal person, being an equivalent of enterprise as understood in statistics and with an obligation to deliver consolidated financial results, does not occur. Sometimes, under certain legal person, the results are delivered by "the establishment of a legal person" or "the individual internal unit of entity". Such a unit may only be considered when its mother unit is a legal person or an organizational unit with no legal status.

The Data Up-To-Dateness in the Registers

Analyzing registers as the source of information for statistics, including structural business statistics, the quality of these data should be most of all taken into consideration.

The updating of the administrative registers, in spite of legal regulations, is far from perfect. This results in the inability to define, on the basis of administrative data, an actual number of active entities.

One of the essential issues that statistics encounters is an out-of-date classification of types of the business activity in the registers, when compared with the business activity types conducted currently by the economic entities. It occurs quite frequently that the only information on the type of business activity in the register is an activity declared by an enterprise at the moment of launching this business activity, i.e. the original entry in the register.

As a complete survey covers larger enterprises and thus the database is updated annually, a representative survey, in turn, covers micro-enterprises, and the scale of such problems is increasing. Such situation stems from the fact that each change entered into the administrative system must be preceded by formal, usually paid for, submitting of an application for such a change by an entity, i.e. an entrepreneur. The entrepreneur is usually not interested in visiting the office again until he is told by some institution to provide an updated copy of an entry from the register. Such a situation makes it difficult for the statistics to obtain high quality results with further specification into the types of business activity. At the level of NACE section such results could be acceptable, but at the level of divisions or classes, the flow of units between the registered activity and the one actually conducted is quite essential. In effect, one of the primary requirements regarding proper data aggregation imposed by the regulations on the structural statistics cannot be met.

Similar difficulties refer to the variable of the employment size, which in the REGON register is not included to be obligatorily updated at all. This enables to use only some, limited information for statistics.

4. Current Exploitation of the Registered Data in the Structural Statistics

As it has already been mentioned above, until now the administrative data have not been used in the structural business statistics. The start of regular exchange of such information, particularly between the statistics and the tax and social insurance systems, must be preceded by multiple preparation works.

For this group, one of the basic problems that may occur when elaborating data for structural business statistics is the necessity to prepare the results in aggregates impossible to be obtained directly by means of the representative survey. In that field, an estimation of the results without any additional information, enabling to develop appropriate statistical models, is extremely difficult, and the mistakes that could possibly be made in such a situation are impossible to assess. The entities may change significantly in the aspect of statistics. This results, among others, from the asymmetrical grouping according to such features as income, costs and even remuneration, but also from their considerable changeability due to the size of the entity or its legal form.

Throughout the work process, the similarities and differences as to the scope and content of specific registers were recognized as well as the adopted rules for their updating and actual up-to-dateness of specific rules. This latter issue refers specifically to the micro-enterprises, covered by the representative survey.

Therefore, an idea has come up to use the tax data as a source of data supporting the small area statistics. Thus, the survey was conducted which was an attempt to use the administrative data from the tax system of the Ministry of Finance for legal persons and natural persons running the business activity as a source of additional data (complementary, also called covariates) in order to better estimate the parameters characteristic for small economic entities than it could be done on the basis of the statistical surveys. When applying the traditional methods of estimation, which exclude additional information from outside the sample, it is difficult to develop satisfactory results in aggregates required by the structural statistics. In most cases, such samples would give estimators of to high bias.

In the process of works, there was also developed an outline for methods that can be applied when exploiting the tax data in the indirect estimation technique for small domains. A

primary estimation with the tax system data was also executed and the precision of results was assessed on the basis of the analysis of the estimation variances for simple samples. The results so achieved were assessed, and in conclusion, the target size of the sample for future statistical surveys of micro-enterprises was suggested. The results of these works may be applied in practice the moment the statistics obtains regular access to the tax data.

SUMMARY

Summarizing the above, it should be stated that Polish statistics should and will continue, by conducting surveys, the works over the use of the data collected in various administrative systems. The analyzes that have been made so far indicate, however, that the move towards the administrative sources should not fully eliminate the statistical reporting, cannot be done automatically and the implementation of such approach may only take place after certain requirements have been met. The primary condition is to obtain by statistics fully identified individual data files. Even if this condition is met, it should be kept in mind that the use of administrative data may be limited. They stem from, among others, different subjectivity and acquired definitions of individual variables, and sometimes from the classification used, which usually derives from the objectives for which certain administrative register was created.

As it has been observed earlier, it occurs quite frequently that the administrative registers do not contain such data as the type of conducted business activity or the size class of the entity. Even if such information is available in the register, then, as to the largest units being of key importance to the national economy, it does not concern the entire statistical unit but only its selected elements treated as separate entities of administrative registers. Thus, a development of methods allowing for the unification of information collected on the statistical forms with the information from the administrative sources for one statistical unit requires much effort and work at the level of the unit data coming from different registers.

It should also be emphasized that even in the target system, the administrative data shall not substitute all information required by the structural business statistics. In case of Poland, these registers do not collect a number of detailed data, which is needed for the statistical system. It should particularly not be expected that the complete balance sheets of enterprises, or their loss and profit accounts, be possible to obtain. Here too, combining the administrative and statistical data, their control and correction, may turn out to be the time- and effort-

consuming process. The non-unification regarding the entities that may be observed in the registries may result in the situation where, for example, detailed results of consolidated balance sheet, collected during the statistical survey, may be different than the results being the total for selected elements of this enterprise available in the administrative registers. It will stem from the way the data are registered by the reporting unit, which in the statistics is an enterprise and in the tax system – a taxpayer, company, i.e. the unit producing the balance sheet. For example, part of the flow between local units and enterprises is registered as the semi-products' flow in the statistics, while in the situation where the “company” submits partial balance sheet, these operations may be registered as the products sale, goods sale, or other.

Another issue that should be given attention is the time frame for delivering data to Eurostat by individual countries. Meeting the deadline also determines the possibility to use the administrative data for the structural business statistics. Individual administrative systems have their own internal requirements to which they adjust their work schedule, including the deadline for registration of the data which could be used by the statistics. The necessity to complete all necessary preparation works on obtained files, including the comparison of administrative data with the statistical data as well as clearing up some particularly significant discrepancies and agreeing (from the point of view of the structural statistics) on the correct data, is a time-consuming operation that could negatively affect the data availability within the deadline dates.

For example, the data processing for outliers units, which may be observed every year in the reports, cannot be automated. An illustration here may be the enterprises which declare itself insolvent, or the ones which ceased up their activity in the reporting year.

Majority of the problems presented above regards large units. In case of smaller units, the main problem that comes up when developing the structural business statistics for them regard the statistical register being out of date, as to the activity state of the surveyed entities, the type of their business activity, and their size. In such a case, the data from administrative registers may be used when calculating values at the high aggregation levels and as the data for creating models for estimation within smaller areas.