

“STATISTICAL USE OF ADMINISTRATIVE RECORDS: A REAL TIME EVALUATION OF LABOUR DEMAND”

Manlio Calzaroni-Roberta Rizzi

Istat- calzaron@istat.it

Istat- rizzi@istat.it

In Italy, as in the other European countries, information from administrative records have allowed for a significant improvement in the production and dissemination of statistical data. Such data attends to the needs of a growing demand for reliable, prompt economic data on the production structure and economic results of businesses. Moreover, obviously, it reduces the costs and the statistical burden for businesses.

1. The legal framework

To use administrative data for statistical purposes, a legal framework must be in operation. Every NSI in Europe is legitimated by statistical laws on international and national levels.

On the international level, various regulations enacted by Community Agencies have established the foundations for the sound use of information sources of varying origin. Below are some of the sections in EU Regulation 322/97 concerning Community statistics, and EU Regulation 58/97 concerning structural statistics of businesses.

Subsection 1 of sec. 16 of Regulation 322/97 sets forth that:

“In order to reduce the burden on respondents...the national authorities and the Community authority shall have access to administrative data sources, each in the fields of activity of their own public administrations, to the extent that these data are necessary for the production of Community statistics”.

While subsection 2 points out that:

“The practical arrangements and the limits and conditions for achieving effective access shall be determined, where necessary, by each Member State and the Commission in their respective spheres of competence”.

Finally, subsection 3 of sec. 6 of Regulation 58/97 states that:

“...the national authorities and the Community authority shall...have access to administrative data sources covering the fields of activity of their own public administrations...”.

On the national level, legislation was introduced in order to recognise the right to full access to all administrative sources by the Statistical Institute. Istat has been assigned particular importance in the management and control of public administration records.

The previously synthesised legislative context represents a necessary prerequisite, but not enough to achieve appropriate use of administrative sources. Starting from this regulative framework, it becomes essential to define the means with which the NSI relates to the agency producing administrative data. Based on whether or not the NSI may intervene during the phase of formulating the data, there are three methods with which the same data is made available to the NSI:

I) **“foundation stage”**, during which:

- national statistics institutes must acquire the power to access and use data bases of the various controllers of public agencies;
- their independence and impartiality must be guaranteed;
- they must stick to the rules that safeguard the privacy and confidentiality of the data in cases of treating sensitive information;

II) **“consolidation stage”**, during which the NSIs use the legislative instruments already acquired in order to consolidate the administrative sources into a resource available for the production of statistics;

III) **“evolution stage”**, during which a profound degree of cooperation among administrative agency and statistical institute is created, so as to allow the latter to influence the form and content of administrative data in such a way as to make it more suitable to its subsequent statistical usage. This last phase is of crucial importance, since it can contribute to removing stumbling blocks that exist in using administrative data, above all:

- by reviewing the composition of concepts, definitions and administrative classifications in order to make them as coherent as possible with those used in statistics (compatible with the institutional aims from which the administrative data is derived);
- by defining common standards for data processing treatment of the information;

- by foreseeing the possibility that the Agency, driven by requests from the statistical institute, may ask the persons in charge of the administrative documents for new information, the purposes of which differ from purely administrative purposes.

Istat is currently in the “consolidation stage” and is beginning to get down to the problem of starting up the third stage (evolution stage). The endeavour achieved by Istat is that of setting up “permanent” structures of comparison between Istat itself and the Agency who supplies administrative information, the object being to apply the guidelines that characterise the evolution stage.

2. Problems linked to the use of administrative sources

Clearly the use of information acquired for administrative purposes in statistics is not a zero cost operation.

Literature on this subject identifies four elements useful in synthesising the methodological problems when using information with an administrative origin: 1) the subjects registered in administrative sources are different from statistical units; 2) data is collected for non-statistical purposes; 3) coverage of the reference population is usually different; 4) the control of methods for data collection is done entirely within the administrations.

Using the point of view of statistics on enterprises, we can say that enterprises systematically produce administrative acts during their lives: they pay taxes, stipulate telephone and electrical energy services contracts, insure the employees against accidents at work. All the aforementioned administrative acts are potential sources of useful information to describe economic activity from a statistical point of view.

Every administrative body has its own function to collect data and manage the corresponding records, under specific legislation and rules which govern relations between various individuals and between them and the public administration. The administrative body defines, classifies, collects and records information on economic agents and their characteristics that, in the strict sense of word, do not have statistical validity. In other words using administrative data causes a problem for statisticians: the inconsistency of data.

The use of administrative data for statistical purposes imposes the necessity to solve the usual problems of a statistical survey - accuracy, completeness, timeliness – with a new conceptual and methodological approach.

Within a survey, consistency is a problem evaluated ex-ante as well as it is strongly linked to the process of microdata collection and macrodata production. When we want to use data stored in non-statistical (administrative) databases, for which statisticians do not have any control of the production process, the problem of consistency is set in a different context and it is resolvable only ex-post.

The main problem that arises using administrative sources for statistical purposes is to identify the correspondences between the statistical concepts and the administrative rules through which those sources observe the population of reference. It is therefore necessary to handle the administrative sources in order to align them with the statistical concepts and definitions. This is possible if, on the one hand, we have an in depth knowledge of the sources to be used and, on the other, suitable statistical methodologies are available.

It is possible to synthesise the logical process for the use of the information derived from administrative sources according to the following three conceptual steps:

1. definition of the reference conceptual frame: statistical definitions and classifications;
2. knowledge of the observed universe (administrative files) referring to coverage, definition of the units and characters, classification used, time and modalities of updating;
3. identification of the rules to convert administrative data into statistical data.

The use of an exclusively administrative source, for example the fiscal one used in some European countries as basis for statistical registers, could cause serious problems. Referring to a defined statistical universe, the typologies of errors generated in the use of only one administrative source for statistical purposes, can be summarised as follows:

- | | |
|-------------------------------------|--|
| E1 – error of under-recording | a) missing records of legal subjects due to delays, etc...;
b) misrepresentation of legal subjects not obliged to the registration. |
| E2 – error of over-recording | a) registration of not active legal subjects due to duplications, delays or cessation recording;
b) registration of legal subjects without any enterprise features. |
| E3 – error assignment of characters | a) incorrect recording due to delays in variations acquired or to errors in declarations, in recording, in checking;
b) incorrect recording due to different definitions and classifications. |
| E4 – missing assignment of char | a) partial or total lack of attribution of a character. |

For the above mentioned reason together with the conceptual steps previously illustrated, it is necessary to develop a further function of "identification of rules for the integration of data coming from more administrative sources". After the implementation of such appropriate statistical methodologies, the integration process is a tool useful to ensure the

exhaustiveness of units and of the characters of the units, obtaining in such way a reduction of type E1 and E4 errors. Such process must be less useful for the reduction of over recording errors and of wrong character attribution. In fact, using more sources can cause an increase of type E2 error; while if each source is really and considerably better than the other, further information for imputation of statistical characters would cause troubles. Besides, the presence of unchecked matching procedures among sources could cause record duplications and therefore an overestimation of units and statistical aggregates.

In other words, a standardisation rule converts administrative concepts and classifications into statistical ones. This rule, generally deterministic, can be divided into three types:

- coding rules: which convert coding (e.g. economic activity, legal form, and location) into statistical classifications (Nace, Nuts, etc.);
- link rules: by which the different records corresponding to legal or administrative units in one source can be combined to define one statistical unit (enterprise or local unit) ;
- conversion rules: to obtain statistical variables from administrative characters

The methodologies adopted in the imputation of characters must be based on the concept “choice among alternative values” and not on “combining the available values”, when there are more values of an attribute for the same unit (Abbate Garofalo, 1998). This means to build a statistical business register.

3. “Nominative Declaration Of Insured Parties (D.N.A.)”

3.1 Characteristics

Legislative aspects

Starting from March 16, 2000, Legislative Decree no. 38 of February 23, 2000 set down the obligation for employers to communicate to Assurance Body (Inail) the starting an/or termination date of an employment, within 24 hours of its occurrence, on pain of an administrative penalty for each omitted or erroneous declaration.

Field of observation

The obligation to “immediately” supply the information to Inail concerns businesses that set up subordinate employment relations in activities that the law identifies as hazardous.

Variables

In order to reduce to a minimum the material burden of fulfilling requests made to businesses by the DNA, a very limited amount of information is requested, such as: the taxpayer’s code of the business, the taxpayer’s code of the employee, the date of hire, the date of termination and the date the declaration was made.

In particular, the business must indicate whether the job is an open-ended or fixed contract, in the latter case it must be specified the termination date of the employment. However, there is no obligation to indicate it at the time the hiring takes place, even if the employment is a fixed contract. This apparently takes into account the fact that, at the time of hiring, the business does not always know with certainty when the employment will be terminated.

Tools of transmission - Time recording

Aware of the fact that this obligation is added to others that businesses have to discharge over the course of the year, Inail has tried to facilitate this task, by arranging for a series of channels for communicating the declarations. The transmission tools are: in written form, the call center, e-mail, normal post, fax, computer support (floppy, CD).

The data supplied is transmitted via computer to the agency that keys it into a single file. This file is then updated and available to users in real time.

Unit of analysis

For the purpose of a correct economic interpretation of data that can be withdrawn from the DNA, it is necessary to specify the unit of analysis of the source, which is dictated by the individual employment records that are set up at a particular time. This means that the occupation is not disclosed (as the number of employees), but rather the working positions related to the hazardous activities that begin and end, and that refer to a specific employee. The taxpayer’s code of the employee allows for tracing the working positions to the individual employment (or, better said, to the variation in the employment position of each individual). The structure of this information does not represent a limitation, but rather an enhancement of information potential: it is actually possible to work up this data in terms of both positions and employees.

3.2 Information potential

The choice of surveying a limited set of information is strongly influenced by the necessity to simplify administrative duties for businesses, thus guaranteeing the quality of the information transmitted. In itself, the slight amount of information requested contains a considerable informative power. The taxpayer’s code of the business (that represents the survey unit) allows for integration with information contained in other databases (statistical record of active

businesses – ASIA), and, hence, to describe the businesses that hire and fire employees, analysed according to the economic activity sector, location, company size, and all the information that the link with Istat records allows to be ascribed to each individual business. On the other hand, the use of the taxpayer's code of the employee permits adding other information that may characterise the employees involved in the various actions, such as age, sex, nationality.

The time recording of the data (communication with 24 hours of the occurrence) makes this information source especially valuable in the short-term analysis of the dynamics of the labour market. The thoroughness of the information offered by the DNA regarding the dynamics of the labour market is potentially such as to grant levels of awareness and trends, which were previously unknown in Italy, once this administrative database has been opportunely screened and transformed into a reliable statistical source.

Suitable use and treatment of the DNA would allow for the production of indicators useful in the awareness of the labour market, some of which are: the average time-span of labour relations in progress and the ratio between fixed and open-ended contracts, which would give an account of the trends, on one hand of company seniority and, indirectly, of the survival capacity of the businesses, and on the other hand of the flexibility/stability of labour relations.

Moreover, this information represents a valuable documentation of flows that occur on the Italian labour market and also concerns matters of labour mobility that are recorded among the employed insured by Inail.

Another potential use of DNA is derived from the prospect of linking it to the accidents file (DNI) and other INPS files, in order to combine individual personal data with company data, especially regarding size and activity sector. This way one could obtain an estimate of accident probability that reconnects the personal characteristics of the injured employees to the economic activity sector and company size.

The speedy updating (in real time) of the DNA makes it possible to use it as an ideal forecasting tool for employment variations. The main problem connected with the use of DNA in conceiving employment forecasts is composed of the lack of employee stock: the DNA only records movements about the beginning and end of labour relations and, at the moment, there is no link between the flows and an administrative Inail file, which records all of the insured parties. However, the DNA does report the identifying code of businesses that leads to labour relations (taxpayer's code), which allows for a linkage to any administrative or statistical record that contains all businesses identified in that manner. To date, ASIA records seem to be the most suitable for use in combination with the DNA. So, with timely stratagems, it is possible to estimate employment variations on absolute levels existing between two opportunely chosen dates.

The basic idea of the proposed outline is to estimate, by means of the DNA, some employment variations (more or less collective) that can be logically compared to those revealed through the quarterly survey on the labour force, conducted by Istat. According to this inquiry, a person is considered working during a given reference week, if, in that week, he/she carries out an employment activity for at least one hour. Thus, two periods of one week, chosen within a time span of an arbitrary number of months, are taken into account and all labour relations indicated by the DNA during those two periods are "counted" one by one, the completely realistic hypothesis being that any labour relation whatsoever detected in the DNA considers at least one hour of labour during the observed week.

One can deduce a complete case record of the movements:

- Some movements are not detected in the DNA (they are "invisible") and thus do not lead to any employment variation ($\Delta = 0$). They include those relations existing before the beginning of the observation period and still in progress after the end of the period; those existing, as well as terminated, before the beginning of the observation period; those that began and ended within the period elapsing between the two observation periods; those that began after the end of the week ending the observation period; and, finally and obviously, those that never existed before nor after.
- Other movements do not lead to any employment variation ($\Delta = 0$), but are "visible" and concern all of the relations existing during both the first and final observation weeks.
- Movements exist that lead to a positive employment variation ($\Delta = + 1$); these are relations that begin during the intermediate period and still exist during the final observation week as well as those that begin during the final observation week, regardless of when they end.
- Finally, there are relations that lead to a negative employment variation ($\Delta = - 1$); these are all relations that begin before the first observation week and end before the final observation week; there are also those that begin during the first observation week and end before the final observation week.

This logical outline exhausts all possibilities, which make up a labour relation that spans the two weeks that begin and end the observation periods. In order to "repeat" the estimation of the employment variations revealed in the quarterly Istat survey of the labour force, two weeks would have to be chosen between those in which the inquiry is carried out.

Despite its being a logic of extreme simplicity, the algorithm is relatively demanding from a computational point of view. The algebraic sum of all positive and negative variations detected during the interval of observation supplies the estimate of the employment variation during that period.

The proposed outline estimates the variations of the number of labour relations during a given time span, not the number of persons employed, and certainly not the equivalent labour units; jobs related to the same person can be traced

back to the unit, so as to reach an evaluation of the number of persons employed. The problem with estimating units of equivalent labour must be solved after this 'recompression'.

Hence, by means of the DNA, the proposed method leads to an estimate of the variations of the total employment rate, according to some important separations, in the same way and comparable to those obtained in the quarterly survey of the labour force.

4. Problematic aspects

As previously stated, the use of administrative sources for statistical purposes is neither easy nor automatic; it requires a great deal of caution and, in any case, may only occur after timely verifications and statistical treatments that allow for the transformation of generic information into statistical data. One of the essential requirements is the accuracy of the data, or rather, the exactness of each piece of information contained, for example, in each declaration of the DNA. Analysis of the data has pointed out a series of more or less formal critical points, which are outlined below:

- a. Inconsistencies. Taxpayer codes, formally correct, but not exact, or the VAT registration number is used instead; hire and termination dates not admissible or missing.
- b. Duplications of declarations with the same key (company taxpayer's code–employee taxpayer's code). In case of corrections or erroneous registrations.
- c. Overlapping of records with the same key (company taxpayer's code–employee taxpayer's code). In case of corrections or erroneous registrations.
- d. Overlapping of records with the same employee taxpayer's code, but at different companies. Possibility of two jobs in case of two overlapping records.

Apart from the more formal problems of accuracy of the data, aspects more connected to the interpretation of the data itself have arisen. One problem, for example, is that which arose from the statement that, during the first few days of work, there is a certain intensification of accidents; this would appear to indicate, though indirectly, the existence of a possible evasion of the obligatory record of employment. If the declaration is only made for employees who have been injured and not when they are actually hired, this would not be a case of loss of information, but of incorrect temporal placement of the event, which would simply lead to the arrangement of a correction procedure. Whereas, with regard to non-declared or non-injured employees, this would be a real loss of information. Therefore it would be necessary to prepare a procedure to detect these cases.

More complex problems also exist, which are not easily resolved by a plan of checks, but rather by paying particular attention to the field of observation of the source of information.

The distinct uncertainties about the subjects included, the particular vastness of the field of observation and frequent changes in the theoretical reference population make identification of the aggregate both imprecise and variable over time, and thus not very reliable for statistical purposes, despite its unquestionable, significant relevance. In fact, one should consider that a legislator could reach new categories of subjects to be insured by Inail and that the judiciary itself could go so far as to see persons in charge of some kinds of activities as subject to insurance obligations, hereby extending the field of observation.

References

Accornero A., Contini B., Filippucci C., *Relazione finale progetto “Utilizzazione statistica della Denuncia Nominativa Assicurati”* – First draft- September 2002 .

Calzaroni M., *Informazioni amministrative e dati statistici. L'esperienza in corso sulla nuova fonte Inail: “Denuncia Nominativa Assicurati”* – V Conferenza Nazionale di Statistica- Innovazione tecnologica e informazione statistica. Rome, November 2000.

Garofano G. *To exploit administrative sources: a framework of concepts*. In 16th International Roundtable on Business Survey Frames, Lisbon 20-25 October 2002.

ISTAT. *Rapporto annuale: la situazione del Paese nel 2001*, Rome: Istat, 2002.

ISTAT- INAIL. *Contributo per una rilevazione in tempo reale dei flussi occupazionali*- September 2001

STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES, *Use of administrative Sources for business Statistics Purposes* – Handbook of good practises (Version 1)