STATISTICAL PROCESSING OF ADMINISTRATIVE FILES

THE FRENCH EXPERIENCE

Philippe TROGAN, INSEE

USE OF ADMINISTRATIVE SOURCES IN FRANCE

Many services in the civil service gather information on various aspects of the activity of produc' on units. Indeed, economic and financial administrations have many contacts with production units for collecting taxes and contributions to Social Security (pension fund and national health insurance) or for securing the settlement of economical and social regulations. An important quantitative information is thus collected: enterprise accounts, but also data on salaries and employement, financial data, data on sectors which are experiencing very strict regulation or supervision by the Government services (banks, insurance companies but also enterprises providing temporary personnel, security activities, ...) [1, page 11].

ADVANTAGES AND DISADVANTAGES OF ADMINISTRATIVE SOURCES

Administrative sources have the **advantage** to be quasi exhaustive. Their use was facilitated, in France,by a set of various favorable circumstances.

First, the objective of limitating the statistical burden put on enterprises permitted to overcome many oppositions.

Secondly, on a technical point of view, three factors were very favorable:

- the use of the SIRENE indentifier (which consists of an enterprise identity code unchanged over time and an activity code called APE) in all administrative files;

- the standardization of enterprise accounts;

- the development of the use of computers.

At last, the presence of statisticians in the services which process the administrative files has contributed much for the actual mobilization of these files.

There are indeed advantages in using administrative files, but there are also some **weaknesses**.

The scopes are often partial. Exhaustivity of the theoretical scope is rarely guaranteed, either because late declarations are not included (fiscal declarations) or because absence of declaration does not result in a sanction.

The main codifications, essential for the statistician, either are rather of poor quality (principal activity, juridical status) or do not exist at all and are to be made.

Information not directly connected with the application of the regulations is generally not reliable. The number of wage earners in fiscal declarations and the non taxable turnover in VAT declarations are two of the most obvious examples.

At last, the size of the file may be a handicap: in order to curb costs and increase timeliness, statisticians are sometimes obliged to use sampling procedures.

Because of these difficulties, it is indispensable that statisticians intervene in the processing of administrative documents, in order to check the files, to correct them, to make certain codifications and estimates, to complete the files, ... After these

interventions, the quality of the results is usually comparable to that of the survey results [2, page 137].

THE USE OF ADMINISTRATIVE FILES

Administrative files can be used in three ways:

- direct processing;

- processing by matching with statistical sources;

- processing using a supplementary ad-hoc statistical survey.

Direct processing

Direct processing is the most frequent use of administrative files. The statistical source is of sufficient interest to justify a direct processing. This happens frequently in numerous processings on employment made by the Ministry of labor. This is also the case for the annual declarations of social data (DDAS = déclarations annuelles de données sociales), one of the main source for detailed statistics on wages, the data of which are provided to INSEE which makes a statistical processing. Similarly, the declarations for unemployment insurance and for Social Security are used for many processings, and UNEDIC (= union nationale pour l'emploi dans l'industrie et le commerce), one of the collector organizations, issues a monthly bulletin of statistics which containss extremely rich results.

Activities which are closely regulated or looked after, either for safety reasons (banking, insurance activities) or for social reasons (provision of temporary personnel, surveillance, building cleaning, ...) are liable to numerous administrative declarations which are often the subject of statistical processings.

This explains why the scope of annual surveys on enterprises was not extended to the financial sector: the Bank of France and the Service of insurances of the Department of treasury have at their disposal the administrative sources which permit to gather an information comparable to that of the annual surveys, provided only that the basic data be re-organized in a simple manner.

As to provision of temporary personnel, the Ministry of labor publishes many statistics on the number of contracts, their duration, their localization, the repartition of "work-weeks" by sectors of activity, which allow national accountants to make approximations of the turnover of the enterprises by sectors of users and spares the statistician the carrying out a specific statistical survey. On the opposite, statistics on leasing by sector of users do not exist in administrative files and a specific survey has to be carried out in order to be able to break down investments by sectors of users.

In fact, it is regrettable that certain administrative sources dealing with very precise domains (surveillance, for example) are little, badly or not enough processed regularly, while the cost of such processings would be extremely small.

Processing by matching with statistical sources

The best example of this kind of processing is the unified system of enterprise statistics (SUSE = système unifié de statistiques d'entreprises), where annual fiscal declarations by enterprises (approximately 2,000,000 annual

declarations) are matched with the files of the annual surveys on enterprises (approximately 200,000 records).

SUSE aims at establishing a consistent set of statistics on enterprises. It was created because of the difficulties encountered by national accountants, when they carried out the 1962 base of national accounts, to make use altogether of data in the fiscal declarations and of the responses to the questionnaires of the annual surveys on enterprises.

It is "a central system of structured statistics on enterprises, which presents enough internal consistency so as to elaborate national accounts from it".

In order to reconcile data coming from various sources, it seems generally preferable to act at the elementary unit level, at least for the most important of these units ... thus, the idea of designing a system which ensure the consistency of individual data.

The general principles upon which the present system relies are consistency, comprehensiveness, selective treatment of anomalies and observed discrepancies, and mutual enrichment of sources.

Several types of **consistencies** are treated:

- internal consistency of each source for a given fiscal year
(balance of accounts, likehood of figures, ...);

- temporal (pluri-annual) consistency (identity of "hooking"
variable concepts such as inventories, likehood of time-to-time
changes, ...).

The **comprehensiveness** of each data source, fiscal declarations and annual surveys, is not ideally perfect. Some enterprises do not appear in the fiscal files (late declaration, enterprises not liable to taxation); similarly, some enterprises are absent from the survey files (smaller enterprises are interrogated on a sampling basis; although all important enterprises are questioned, there may be non responses and observation errors). There is therefore a need for procedures to check for, at least, the presence of bigger units by comparing the two sources and examining the previous years files.

Anomalies and observed descrepancies are treated at the individual level, at least for the bigger enterprises (in general, enterprises with 20 or more wage-earners), using all available sources (in particular, Assembly reports).

The matched sources do not contain the same data. There is therefore mutual **enrichment**. For example, the survey questionnaire does not include questions about short term data such as cash accounts, short term liabilities, ... On the other hand, the surveys on enterprises include numerous questions on employment and on the breakdown of sales, which are not available from the fiscal documents.

Annual survey information, made consistent with the fiscal declarations, is used for estimating many data on fixed capital formation (which requires a breakdown by type of capital formations) as well as, for example, the "transfer matrix sectors-to-branches" which permits the calculation of profit-and-loss accounts by branches [3, pages 4 and 5].

Processing using a supplementary ad-hoc statistical survey

The calculation of the monthly indexes of sales by enterprises in internal trade and market services sectors is a good example.

Industrial and commercial enterprises the turnover of which is higher than 3 millions francs and service providers whose turnover is higher than 900,000 francs are liable to the so-called "actual profit" fiscal regime, and must fill up, according to their specific situation, a monthly or a trimestrial declaration of turnover, within the framework of the VAT (value added tax) payment regulation.

Since 1976, INSEE processes these declarations in the manufacturing sectors, using a sample of enterprises. It is thus possible to measure the time-to-time changes of turnover at "constant scope of enterprises".

Since the beginning of the eighties, the processing has been extended to internal trade and market services sectors. The method was presented during the fourth meeting of the Voorburg Group [4] and will not be presented here again. About 12,000 enterprises in the internal trade sectors and approximately as many in the market services sectors are now processed each month.

However a problem arose, which did not exist for the manufacturing sectors. In the internal trade and market services sectors, or, at least, for some of the sectors (retail trade and household services), small enterprises, not liable to the "actual profit" regime and thus not obliged to fill a monthly or trimestrial VAT declaration, are numerous and represent an important part of the turnover in their sector of activity.

Thus, if the "coverage rate" (share of the total turnover of a given sector of activity corresponding to the sales of the enterprises liable to a monthly declaration for VAT) is examined, one arrives at rates close to 100% for manufacturing and wholesale trade sectors; as to retail trade and household services sectors, the percentages are often much lower. For example, the coverage rate is only 35% for bakers or 33% for hairdressers.

The resulting fact is that, for those activities, the turnover changes obtained from the use of VAT declarations are not representative of the sector as a whole.

Two solutions were then considered:

- abandon, for those sectors, the processings of monthly VAT declarations and carry out a direct statistical survey;

- carry over the processings of the administative source, but supplement it by an ad-hoc statistical survey on small enterprises.

The latter solution, less burdensome for enterprises, was selected. A supplementary survey was then launched, the "monthly survey on small enterprises in the distributive trade and services sectors" (enquête mensuelle auprès des petites entreprises du commerce et des services). Each month, 6,000 enterprises are interrogated on their last month turnover and investments. The survey concerns 46 activity groups in the French activity classification, while the processing of VAT declarations deals with 150 activity groups for internal trade and market services.

The survey is used to correct the lag in the time-to-time changes between the sub-population of enterprises liable to monthly VAT declarations and the total population. But it does not aim at producing turnover indexes for small enterprises,

which would be disseminated at the detailed levels of the activity classification. No monthly publication of the results of this survey are made [5].

-O-O-O-O-O-

BIBLIOGRAPHY

[1] "Les sources statistiques sur les entreprises", les Collections de l'INSEE, série E Entreprises, n°117, septembre 1988.

[2] M.QUELLENNEC - "Les statistiques d'entreprises", les Collections de l'INSEE, série E Entreprises, n°101, juin 1986.

[3] Ph.TROGAN - "Transfer of data by kind of business to data by homogeneous production units" - 4th meeting of the Voorburg Group, Ottawa, september 1989.

[4] N.ACTIF, R.DEPOUTOT and M.-L.HONNIBALL - "The use of value added tax files and of supplementary direct survey on firms for calculating short and medium term indicators for sales and investments in service industry" - 4th meeting of the Voorburg Group, Ottawa, september 1989.

[5] R.DEPOUTOT - "Indices mensuels de chiffre d'affaires dans le commerce et les services", INSEE-Méthodes, n°11, avril 1991.