

17th International Roundtable on Business Survey Frames
Rome – Italy 26 – 31 October 2003
Istat day's Session - Paper n°

Author: Paolo Consolini

Administrative data based statistics: the case of non-pension cash benefits (NPCB)

Abstract

This paper explores the possibility to collect the information on the non-pension cash benefits by using the administrative data sources. Firstly, we describe the framework of the Italian ITS on social protection and define the units of analysis. Afterwards, we review the Bodies holding sources of statistical information and assess the prominence of each data source in terms of the rate of NCPB expenditure coverage. A crucial point of this process is in combining the information from different sources in order to produce statistical data on NPCB expenditure and on the respective recipients. As an example, we analyse the exact matching between DM10 database and 770-tax register. Technical aspects are not taken into account because beyond the scope of this analysis.

1. The Information Technology System on Social Protection Schemes in Italy: SISAP

The Italian ITS on Social Protection (SISAP) is aimed at providing detailed and up-to-date information on the organisation of the Italian social protection system. The SISAP is set up by integrating various administrative data-sources and uses as a framework of basic concepts and classification of the European Classification Systems ESSPROS96 and ESA95.

The SISAP covers four main fields: 1) The Characteristics of institutional units providing social protection (Public Funds, Private Funds, employers, non profit sectors); 2) The information on expenditures and receipts of the social protection systems; 3) The characteristics of social benefit's recipients; 4) The normative, economic, and demographic framework (rules and regulations, sustainability of social protection systems, demographics trends, etc).

The SISAP is a modular system that includes three blocks (a fourth module D is additional):

- A) statistics on basic and supplementary pension schemes;
- B) statistics on social protection *schemes* providing non-pension cash benefits (NPCB);
- C) statistics on private and public institutions providing goods and services directly to protected people
- D) micro and macro simulation models (MODSIM/P; MARSS; MASTRICT)

We focus on the module B "statistics on social protection *schemes* providing non-pension cash benefits (NPCB)". This module deals with two units of analysis: social benefits and recipients. These units refer to two aspects. The first one is quantifying the finances used for non-pension cash benefits, which can be used to reply to the following question: how much did families effectively receive from social protection schemes during the year?

The second one is identifying the effective recipients of social protection actions, which can be used to reply to the following question: how many people receive "non-pension cash benefits" during the year and what are their characteristics?

In reference with the former unit of analysis (i.e. social benefit) we can observe three phenomena or attributes: the amount of expenditure for the benefits paid; the period covered (the period of absence at work during in which the worker is still paid) and the number of payments. On the other hand, the main characteristics of the latter unit (recipients) are: the age, the gender, the residence, the occupation, and composition of own household.

2. Definitions and conceptual features

In this context, we define “Non-Pensions Cash Benefits” as unilateral temporary transfers leading to the flow of cash resources from public and private institutions, insurance companies, and employers to families in order to relieve them of the burden of a defined set of risks and needs involved in social protection.

The risks and needs involved in social protection are classified into the following seven functional categories: sickness and health care, disability, old-age, survivors, family/children, unemployment, housing and social exclusion not elsewhere classified. Contrary to that set forth in ESSPROS'96, housing is not included as this only involves services in kind.

The recipient of non-pension cash benefits is someone who has received one or more non-pension cash benefits for the same type of risk or need or for the basic purposes of social protection over a certain period.

The units providing Non Pensions Cash Benefits (NPBC) can be classified as the following:

- Social security funds (they account for 44.3% of NPCB expenditure in the year 2000);
- Private Employers providing benefits to their current and former employees directly (43.2%);
- Public Employers providing benefits to their current and former employees directly (8.3%);
- Autonomous and self-administrated pension funds (lump sum payments (0.5 %);
- Local government providing public assistance by means of cash benefits (3.7%).

As we can see, there is a plurality of Bodies responsible for paying non pension cash benefits. That is a consequence of current legislation on welfare, in particular social security and employment laws. Furthermore, it may occur that a specific risk is covered by a plurality of institutional units. An example of this plurality is the income maintenance benefit for maternity leave. On the one hand, welfare legislation involves the contemporary presence of a number of social protection and insurance Bodies (INPS, IPSEMA, Professional Funds, etc.) to protect people for the loss of earnings due to the momentary absence from work. Each of these Bodies has a specific competence to protect certain categories of workers, such as public servants, self-employed, etc. On the other hand, legislation on employment establishes that employers ensure better levels of coverage in according to Collective Employment Agreements. This heterogeneity causes a multiplicity of data-sources.

3. Data Sources on NPCB

Now we review a list of the main administrative data sources currently available on *Non-Pensions Cash Benefits* schemes. Each data source is classified according to the institutional unit responsible for the collection of the data and the level of information. The expenditure coverage rate of data sources is put in

brackets. The first group of data sources containing information on a *micro* level (the detail of information is referred to the recipient) includes:

- The 770 Form tax register of the Ministry of Economy and Finance. It contains information regarding social benefits provided directly by employers to their employees: amount of expenditure and characteristics of recipients (the coverage rate of this source is about 41.3%);
- The databank of payment of INPS. It provides information on the social benefits directly paid by this institute (17.0%);
- The databanks of other social security funds that will be acquired in the next future: INPDAP - National Institute of Social Protection for public servants - (11.5%), INAIL - National Institute for Insurance against Industrial Injuries (2.0%);
- The Civil Service databank of General Accounts Office: it comprises around 1,200,000 units referring to the individual information regarding a set of benefits provided by public employers (4.2%);

The second list includes data sources recording information only at a more aggregated level (institutional unit that provides social protection) as following:

- The DM10/2- Form INPS register: it provides information regarding the amount of expenditure for social benefits paid in advance by employers on behalf of INPS - National Institute of Social security for private sector (14.5%);
- The ISTAT survey on the Balance Sheets of Italian Social Security Funds: it provides aggregate information on expenditures for social benefits paid by the remaining social protection and insurance Bodies (2.4%);
- The Personnel Annual Account of the General Accounts Office: it records the expenditure of social protection provided by Public Administration to employees (4.3%);
- The databanks of Social Affairs Department: it includes data on expenditure for the social assistance benefits provided by the Central government (0.3%);
- National Account data: it includes data on expenditure for the social assistance benefits provided by the Local government (2.4%)

In this context, there are two main problems. The first is concerned with the integration of different sources in order to obtain complete or representative information on benefits and recipients broken down by social protection schemes. The latter refers to the implementation of a “linking procedure” between the DM10-Form INPS register and the “770 form tax register. The purpose is to distribute the social benefits paid in advance by the employers on behalf of INPS among all the employees.

By using the first four sources (individual-level) and the DM10-Form database matched with 770-Form tax-register, we expect to produce statistics on benefits and recipient with a coverage rate of nearly 90% of total NPBC expenditure.

4. The “770-Form” tax register

In Italy the employers - as withholding agents - are obliged to return information on “wages and social benefits” paid to their employees by means of the “770 Form” of the Ministry of Economy and Finance. Therefore, the “770 Form” provides information on the social protection schemes run by employers. In particular, the 770 Form includes two sections of interest: Section SA and Section SB.

Section SA records information on:

- i) social benefits paid by employer to employees as integration of social security benefits. These are payments made by employers to integrate the social security benefits and are referred to Collective Employment Agreements;
- ii) the characteristics of recipients (age, gender, residence);
- iii) the total number of weeks they received benefits paid in advance by employer of behalf of INPS;
- iv) the family allowances;
- v) the wages;

Section SB contains information on: the lump sum retirement allowances (retirement bonus, severance pay, cancelling price, retirement incentive, etc) and the characteristics of relative recipients.

The information in this database can therefore be used to distribute the amounts paid in advance by the employers - information recorded in the “DM10-Form” INPS register - among the respective beneficiaries through an appropriate matching operation.

5. Data acquisition of a sample of 770-form: data reading, check and correction procedure

Istat carried out an experimental analysis on a sample of 22.446 firms drawn by ASIA - Business Register frame, the year of reference is 1999. The aims of this activity were:

- to test the data quality (reliability) of the variables of interest (social benefits and recipients) recorded in 770-form register;
- to implement the procedures for the detection and correction of errors. Procedures that will be applied on the register of 770-Form as whole;
- to estimate the amount of retirement benefits and the number of recipients involved;
- to evaluate the feasibility of matching “Dm10-Form register with 770-Form database” through the selection (correction) of an appropriate matching-key.

For this scope, it was designed one stage stratified sample of about 22 thousand enterprises with “a sample rate” of 2%. The stratification variables are 1) the number of employees (three size groups: 0-9, 10-99, over 100); 2) the economic activities (23 groups). The total number of strata is 69.

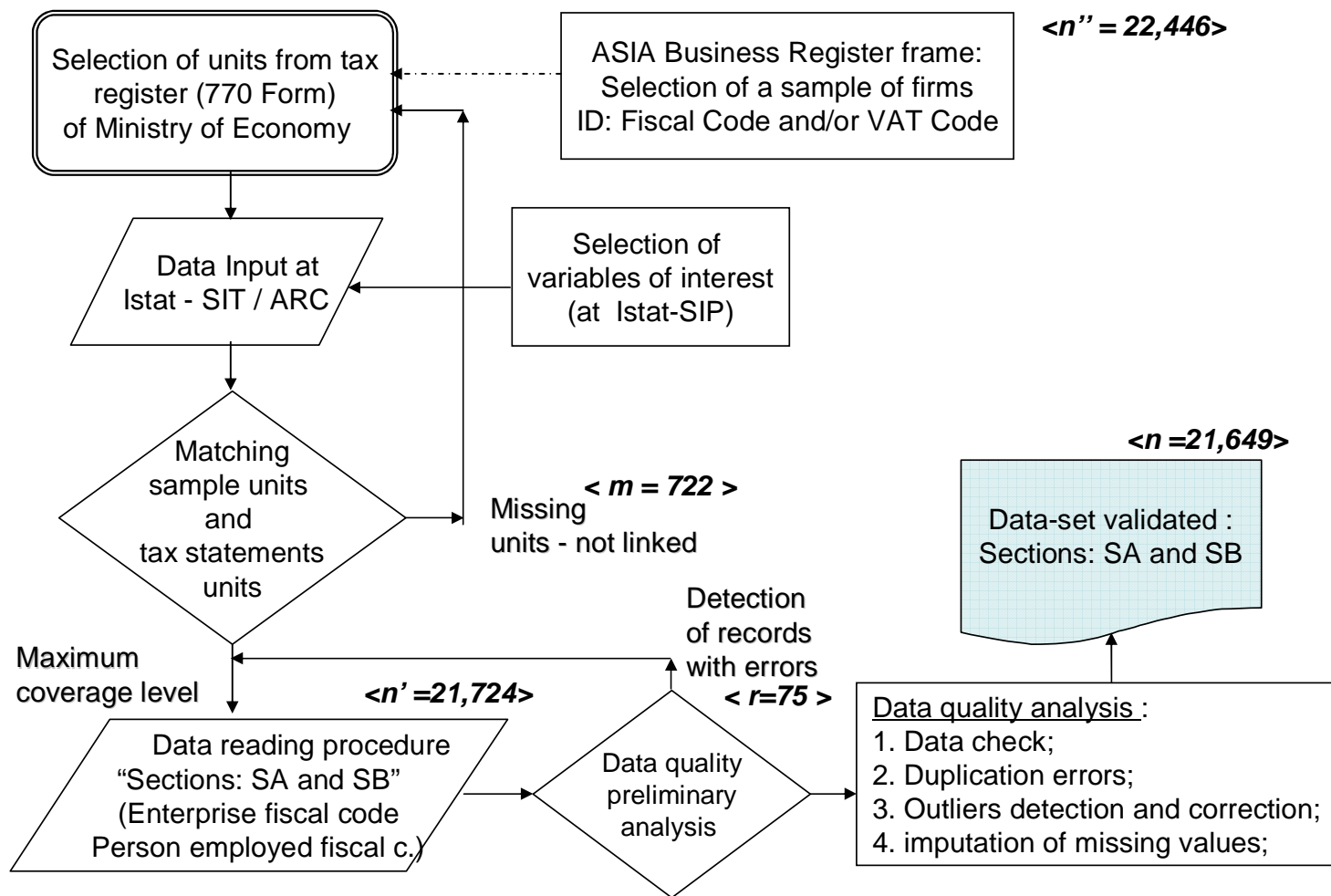
Hence, a list of 22.446 fiscal/VAT codes was delivered to the IT (Information technology) system of the Ministry of Economy and Finance, for the selection of the units from 770-Form tax register.

Afterwards, the Ministry of Economy and Finance sent to Istat the database of tax statements. The Istat Census and Archives Division (ARC) controlled the list of the units selected with the units sampled. The maximum coverage level reached was 21.724 enterprises, hence 722 firms didn't match. Therefore, we can say that the non response rate was about 3%.

The SIP (Statistics on Public and Private Institutions) Office implemented the data reading procedure for the Sections SA and SB, and the data quality preliminary analysis, which led to refuse 75 enterprises.

Finally, we arrive at the data quality analysis. This stage is organised into four steps. The first step concerns with the data check; the second step overcomes the problem of the duplication errors; the third step refers to outlier detection and correction; the fourth step regards the imputation of missing value. As output of this process there is the data-set validated for section SA-SB

**Data acquisition of tax statements (770 form) - year of reference 1999:
Data reading, check and correction procedure**



6. A simplified scheme of exact matching between DM10 and 770 register

The objective of matching the DM10 database with the 770-tax register is to disaggregate the information on expenditure for benefits recorded in DM10-Form at individual level. As a result, we increase by 14.2 percent the expenditure coverage rate of data sources at micro level.

The matching procedure between the two data sources includes four main steps. At the outset, the validated data-set SA comes into our analysis.

The first step is to select a matching-key able to link the two databases. Now, the choice is between the INPS code¹ and the fiscal/VAT code. The INPS code revealed to be the best solution because the variable “amount of social expenditures” from this source is correctly referred to this unit. When this matching-key is not available in the fiscal database, we use alternatively the fiscal/VAT code as a second best solution.

The second step is aimed at deriving a rough estimation of the NPCB expenditure at individual level from a set of auxiliary variables in the fiscal database. The list of social benefits includes: a) family allowances; b) income maintenance benefit for maternity leave; c) paid sick leave; d) unemployment benefits.

The third step aggregates the individual expenditures, as calculated in the previous step, according to the positional unit code;

The fourth step is to compare the aggregate expenditures mentioned above only for the matched enterprises. We distinguish the comparable units from the not-comparable ones according to a minimum and a maximum exclusion threshold applied to the relative differences distribution. The relative differences are calculated respect to DM10 expenditure that represents our benchmark. Therefore, the amount of social expenditures of DM10 form database is assumed to be correct by definition. The amount of social benefits calculated on fiscal database, when not coincident, must be adjusted according to an algorithm: we impose as constrain that the sum of individual benefits for each enterprise must be equal to DM expenditure. We implement the correction of matching-key or micro-data only for the largest enterprises.

This experimental study should be considered a starting point for the implementation of a consolidate procedure working on entire Archives.

¹ It represents a positional unit of the enterprise for INPS.

A simplified scheme of Linking/Matching “DM10_register and 770_tax register”

