Improving the coverage of the Economic Census by integrating the Business Register: a method to measure under-over coverage in the two sources

C. Abbate, D. Filipponi, C. Viviano
ISTAT (Italien Statistical Institut), Rom

Abstract: The coverage survey carried out after the 1991 Italian Industrial Census (CIS) showed the limits of the traditional survey technique by estimating an under-coverage error of about 200,000 local units and 400,000 employees, mainly concentrated in the area of small enterprises. The availability of the Italian Business Register (ASIA) made possible to introduce, in the 2001 Census, important innovations, aimed to improve the effectiveness of the survey technique and to achieve an optimal coverage of the observation field: a correction of the Census coverage error through a comparison between the Census outcome and the set of units estimated active in the register at the same date.


Keywords: Census, coverage, record linkage, latent classes model.

1 Introduction

The structure of the Italian economic system in terms of number of active enterprises and the related characters like economic activity, size and geographic distribution is strictly dependent on the tool used to enumerate the units. A coverage error, that occur when a unit is missed by the survey or enumerated in error, is one of the most relevant type since it affects not only the count of universe but also the accuracy of data in relation to the characteristics of such universe.

A Business Register provides a frame describing the universe of active enterprises. The active statistical units and related characters are defined as a result of a physical integration process, using administrative sources covering business information, and of methodologies developed to estimate and choose the characters of statistical units.
Over-coverage errors are a typical feature of data collected through administrative sources and they are generally due to the time lag between the formal registration of a unit and its actual state of activity (either for delays in cancellation of ceased units or for the actual start-up of an unit after its registration). The cross matching of several sources and the use of updated and monitored archives, especially with regard to the activity status of units, can contribute to reduce such errors. As opposite, under-coverage errors relates to the missing recording of legal subjects due to evasion, delays and the existence of thresholds in some administrative sources.

On the other hand Census is a survey aiming to know the universe of enterprises on a territory. Under-coverage errors are a typical feature of territorial surveys, which require complete coverage of all streets and identification of even less visible units. For an economic Census an under-coverage error is mainly due to the problem of the exact identification of units. High under-coverage error rates are observed for the "hidden" local units such as workers at home, workers with no fixed place of work and the related branches of economical activities (building, transport and communication, professional and personal services) and finally for the coexisting local units. Moreover, the units that are more likely to be omitted in the Census are the small units and the units located in the biggest cities where some survey districts are not covered by the enumerator. An over-coverage error, usually much smaller then the first one, is due to double enumeration of the same local unit at two different address, or to enumeration of local unit not belonging to the Census observation field.

The availability of the Italian Business Register (ASIA) made possible to introduce, in the 2001 Italian Economic Census (CIS), two important innovations with respect to traditional technique. The main innovation was an “archive-assisted” survey, carried out by enumerators equipped with lists of local units, provided by the Business Register. Moreover, instead of carrying a post enumeration survey, Business Register data have been used to measure the quality and the coverage of the Census through a comparison between the Census outcome and the Business Register active units at the same reference time. Comparison between CIS and ASIA allows to estimate under-coverage, that generally affects territorial surveys, and over-coverage that affects a statistical register.

2 The coverage analysis

For comparing the enterprises surveyed with the Census and the enterprises defined as active in ASIA a common observation field should be preliminarily identified. In fact, there are two main differences between the two data sources. The Census data gives stock information, i.e. the number of enterprises active at a certain date, while ASIA gives the number of enterprises active during the year. Generally speaking, for the Census the active units are those operating at the date of the survey while, for ASIA an active enterprises is a unit that has paid the yearly registration duty to the Chamber of Commerce, Industry, Crafts and Agriculture (CCIAA) or which has filed an income tax return for the year of reference, even when the unit have worked for one month only. Moreover, the units surveyed with the Census are all the enterprises engaged in industrial activities and services plus public and non-profit institutions, while the units
considered in the Business Register, ASIA, are all the enterprises engaged in industrial activities and services.

With the limits of the features characterizing both survey techniques the common observation field has been defined as the one covered by both the surveyed, i.e. enterprises operating at the date of the Census and engaged in industrial activities and services, excluding public institutions, non-profit institutions and farm businesses.

2.1 Matching process

To carry out the coverage analysis between the two data sources, and therefore to identify the enterprises surveyed by the Census and not active in ASIA and the enterprises not surveyed by the Census and active in ASIA, the first step is the linkage between the units surveyed on the territory and the units present in the administrative files used to build up Asia. It is important to underline that, even if the coverage analysis is carried out at the enterprises level, the matching process is performed linking the local units (the local unit is an enterprise or part thereof situated in geographically identified place), which are the Census surveyed units and involves on one hand about 3.4 million records, i.e. all the local units collected by the Census and on the other over 30 million records, i.e. all the local units recorded in all the basic administrative data (ASIA-BASE), that correspond to about 10 million enterprises.

The steps for matching Census data and the aforementioned archive are:

- matching between CIS and ASIA-BASE for reliable keys (fiscal code and address);
- probabilistic linkage between residual records of CIS and ASIA-BASE carried out with a software (RIDA) and based on the linkage of the enterprise denomination, address, economic activity and legal status codes;
- matching between the residuals of ASIA-BASE and information annotates by enumerators regarding not-delivered questionnaires.

Once linked the local units the comparison between the two data sources is then carried out at enterprise level and enables us to identify:

- enterprises surveyed by the Census and currently present in the ASIA-BASE (i.e. the enterprises for which at least one local units belonging to the enterprise is present in both sources).
- enterprises with at least one local unit surveyed by the Census and not present in the ASIA-BASE.
- enterprises that the Census has not surveyed but with local units present in the ASIA-BASE.

In addition, for the enterprises present in the ASIA-BASE it is available an activity status indicator, which consequently makes a further subdivision possible:

- the unit defined as active in ASIA,
- the unit defined as not active in ASIA.
The following table 1 shows results of the ASIA-CIS comparison.

Table 1: Number of enterprises by presence in ASIA administrative archives, ASIA activity status and surveyed condition

<table>
<thead>
<tr>
<th></th>
<th>ASIA- BASE</th>
<th></th>
<th>CIS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACTIVITY STATUS</td>
<td>IN</td>
<td>OUT</td>
<td>TOTAL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3,141,83</td>
<td>1,149,58</td>
<td>4,291,422</td>
</tr>
<tr>
<td>IN</td>
<td>ACTIVE</td>
<td>3,141,83</td>
<td>1,149,58</td>
<td>4,291,422</td>
</tr>
<tr>
<td></td>
<td>NOT ACTIVE</td>
<td>234,572</td>
<td>2,222,52</td>
<td>2,457,099</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>3,376,41</td>
<td>3,372,11</td>
<td>6,748,521</td>
</tr>
<tr>
<td>OUT</td>
<td></td>
<td>27,472</td>
<td>0</td>
<td>50.0%</td>
</tr>
</tbody>
</table>

Before applying the coverage analysis’ methodology, an explanation of the disagreements between CIS and ASIA status of activity should be given. The presence of units in the cell E, i.e. units surveyed by the Census and not present in the administrative archives (ASIA-BASE), can be explained as:

- units that have not been matched due to erroneous or missing values in the fiscal code present in the questionnaire, even though they may be present in ASIA-BASE;
- units that are not part of the observation field but that have been erroneously classified as enterprises in Census –over-coverage of the Census--;

The presence of units in the cell C, i.e. units surveyed by the Census and not active in ASIA-BASE, can be explained as:

- units involved in structural changes such as an event that determines a change of the fiscal code (both codes are present in the administrative archive but the unit surveyed in Census has been matched the old fiscal code - cell C – while the new one is unmatched - cell B -);
- units not active with reference to the date of Census but whose questionnaire has been returned to the enumerator by mistake –over-coverage of the Census--;
- units with a wrong state of activity in the Business Register –under-coverage of ASIA--;
The presence of units in the cell B, i.e. units that the Census has not surveyed and that are active in ASIA-BASE, can be explained as:

- units involved in structural changes such as an event that determines a change of the fiscal code (both codes are present in the fiscal archive but unit surveyed in Census has been matched with the old fiscal code- cell C – while the new one is unmatched - cell B -);
- units no longer active before the data in reference or units that have started-up after the date in reference –difference in the time reference--;
- units missed by the enumerators–under-coverage of the Census--; 
- units with a wrong state of activity in the Business Register –over-coverage of ASIA-.

Disagreements classified in cells C and B are the most consistent and relevant to solve and therefore the following analysis will focus on those cells. Cells E will be taken into consideration in the last part of the coverage process.

### 2.2 Identification of the links between units to reduce CIS-ASIA disagreements

The cells C and B in table 1 indicate considerable discordance between the Census and ASIA activity status. It has been underlined that the disagreement between the two data sources cannot be immediately interpreted as under and over-coverage in the individual sources as it is necessary to quantify the effect due to the different concept of activity that has been used - active on a certain date in the case of the Census (stock) and active during the year for ASIA (flow)- and to remove the linkage errors between the units. In effect, matching of units, carried out via fiscal code (FC) and address, can generate false links when it refers to units that have changed their FC during the survey period. In particular, the unit which has answered the Census questionnaire providing its name, address and activities could have omitted to provide the change of FC; consequently matching was carried out on the basis of the previous-old code reported as no more active in ASIA and not on the basis of the new one, already active in ASIA.

To reduce the number of discordances due to enterprises that have changed their FC, the following two techniques have been used:

- identification of enterprises involved in structural changes or demographic events (cessations, starting-up and juridical changes), through the information recorded in the Tax register database;
- identification of enterprises involved in demographic events through record linkage techniques

According to the changes recorded by the Ministry of Finance, between 2000 and 2002 the number of enterprises involved in simple structural changes, that is when a new fiscal code takes the place of an old one, amount to 4,916. The two linked enterprises can be re-classified for the purpose of reducing the discordances. For example, in the case of two units that are linked by a cessation and start-up due to a change in legal
status, if the first unit is classified in C and the second unit in B then they will be reclassified in cells A and D respectively. In this case, the first one is active for both the Census and ASIA and the second is no longer active for both.

Because not all the demographic events are recorded in the tax files, RL technique is applied to identify the enterprises that change of FC. Comparison is made taking into consideration three main matching variables: enterprise name, its address and economic activity code. To establish the agreement/disagreement between the matching variables appropriate standardization and parsing of the name and the address have been developed and the decision rule used is a deterministic one, i.e. if a pair agrees on two characteristics within the set of three characteristics, the pair is designed as a link. The FC and legal status code are used as supporting elements to establish the link/non link rules between the units to be compared.

The units linked through this technique, about 13,000, are re-classified according to the cell to which they belong as previously described in the cases of changes to legal status registered with the Ministry of Finance.

Results obtained using both techniques do not help so much in reducing disagreements between the two sources even though they allow to correctly identify linkages at micro level.

3 An estimate of the enterprise activity status at 22nd October 2001.

The estimate of the number N of enterprises active at October 2001, once the two surveys have been cross tabulated, cannot be obtained using the classic Dual-System method (Wolter, 1986), as this method is only applicable in cases where under-coverage is present and over-coverage of units is absent in both sources (in addition to the assumption of independence between the two estimation techniques).

In the specific case both CIS and ASIA sources can be affected by over-under coverage.

<table>
<thead>
<tr>
<th>Table2: The Dual System model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Indeed, not all the n12 units non-surveyed, but active in ASIA can be considered as Census under-coverage, since it is necessary to eliminate ASIA’s over-coverage due to (i) difference in time reference, in fact the Census gives stock information, which is the number of enterprises active at a certain date, while ASIA gives the number of enterprises active during the year; (ii) the way activity status of enterprises is
determined. In fact, in a Business Register it is preferable to include units having a certain degree of uncertainty of existence rather than miss them.

Likewise, not all the n21 units that have been surveyed and are non-active in ASIA can be considered as ASIA’s under-coverage, since it is necessary to check whether over-coverage in the Census exists. Census’s over-coverage can be attribute to (i) the possibility of having false questionnaire due to the survey “archive-assisted” technique; (ii) the classical error of double surveyed units.

3.1 The latent class model

To recognize into the business frame the units under-covered by the Census from the ones over-covered by the Business Register, it has been used a latent class model. The basic idea underlying a latent class analysis is to explain the distribution of observed variables \( \mathbf{Y} \) according to a latent variable X not directly observed. Here, X is defined as a dummy variable (X= 1 if the unit is active and X=0 if the unit has no signal of activity at the Census date) and \( \mathbf{Y} \) is a vector of variables available from both the administrative files and the survey.

Let assume that N units have been classified according to L categorical variables \( \mathbf{Y} = (Y_1, \Lambda, Y_L) \) with \((K_1, \Lambda, K_L)\) modalities respectively. Data can be represented via a L-dimensional contingency table with \( J = \prod_{l=1}^{L} k_l \) cells. Let assume X be a latent variable and define \( \vartheta_c = P(X = c) \), \( c = 0, K, C \).

Assuming that \( P(\mathbf{Y} = \mathbf{y}) = \sum_{c=1}^{C} \vartheta_c P(\mathbf{Y} = \mathbf{y} / X = c) \) and under the hypothesis of local independence (the L observed variables are independent conditionally to the classes they belong), it follows:

\[
P(\mathbf{Y} = \mathbf{y}) = \sum_{c=1}^{C} \vartheta_c \prod_{l=1}^{L} P(Y_l = y_l / X = c)
\]

The probability to belong to a class c can be calculated as:

\[
P(X = c / \mathbf{Y} = \mathbf{y}) = \frac{\vartheta_c P(\mathbf{Y} = \mathbf{y} / X = c)}{\sum_{c=1}^{C} \vartheta_c P(\mathbf{Y} = \mathbf{y} / X = c)}
\]

A latent class model can be specified as a log-linear model for contingency tables with latent variables. Even under the hypothesis of local independence the log-linear model is specified as:

\[
P(\mathbf{Y} = \mathbf{y}, X = c) = \beta + \beta_c + \sum_{l=1}^{L} \beta_{y_l}^{l} + \sum_{l=1}^{L} \beta_{c,y_l}^{l}
\]
This model includes a parameter for the latent variable $\beta_c$, a parameter for each observed variable $\beta_{y_l}$ and a parameter describing the interactions between observed and the latent variable $\beta_{c,y_l}$. The link between the log-linear model and the conditional probability is:

$$P(Y = y | X = c) = \frac{\exp(\beta_{y_l} + \beta_{c,y_l})}{\sum_{r=1}^{K_c} \exp(\beta_{y_l} + \beta_{c,y_r})}$$

The formalization of a latent class model through a log-linear model (Haberman, 1974) is useful to overcome the local independence hypothesis, by introducing in the model interactions parameters between the observed variables. Parameters are generally estimated through the EM algorithm, where in the maximization step is used the Newton-Raphson algorithm.

The aforementioned model has been applied to detect those units without signals of activity at the 22nd October 2001 in the cell n12 (over-coverage of ASIA) and in the cell n21 (over-coverage of Census). Moreover, assuming that both the sources are affected by over-coverage it must be assumed the presence of over-coverage even in cell n11.

The latent variable $X$ is binary, that means to consider a model with two latent classes with a priori known meaning ($X=1$ if the unit is active and $X=0$ if it has no signals of activity at the date). Observed variables $Y$ have been chosen among the available information in the Census and in administrative sources, taking into account the ones having the highest discriminant power.

In the model have been included in the following variables:

- telephone lines declared in year 2001
  0 absence
  1 presence
- turnover declared by businesses in year 2001. This variable is broken down in 5 classes built according to quartiles of the different economic activities distribution:
  0 for units without turnover,
  1 for units with turnover less than the 1st quartile,
  2 for units with turnover between the 1st and 2nd quartile,
  3 for units with turnover between the 2nd and 3rd quartile,
  4 for units with turnover above the 3rd quartile;
- regional tax on productive activities (IRAP) paid in 2001
  0 not paid
  1 paid
- Employees, 2001 yearly average:
  0 zero employees
  1 1 employee
  0 >1 employees
- Payment of the annual registration fee to the Chambers of Commerce:
  0 not paid
  1 paid
• Employees at October 2001:
  0 zero employees
  1 >=1 employees
• Information on the unit’s activity status recorded by the Census:
  1 active
  2 inactive
  3 ceased
  4 unknown
  5 without any kind of information
• Date of start-up or cessation of the unit:
  1 unit ceased before or started after the 22\textsuperscript{nd} October 2001 (not active at the Census date),
  2 unit ceased right after or started right before the 22\textsuperscript{nd} October 2001
  3 otherwise.
• Typology of unit:
  1 free professional
  2 other

In table 3 are shown some examples of probabilities of activity $P(1/Y)$ assigned by the model to different vectors of explanatory variables; $N$ is the number of units present for each combination of such variables.

<table>
<thead>
<tr>
<th>Teleph. line</th>
<th>C.of Comm.</th>
<th>Turnover</th>
<th>Employees</th>
<th>Employees October</th>
<th>CIS info</th>
<th>Date</th>
<th>Irap</th>
<th>Typology unit</th>
<th>$N$ of units</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>78930</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>120005</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>127216</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>46024</td>
<td>1.000</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>13570</td>
<td>1.000</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>49458</td>
<td>0.5059</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>10932</td>
<td>0.2316</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>36514</td>
<td>0.0195</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>94049</td>
<td>0.0001</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>56386</td>
<td>0.0000</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>74292</td>
<td>0.0000</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>383530</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Results highlights that the model assigns probabilities close to 1/0 at all those units with strong/weak signals of activity from the administrative sources whether the unit has been surveyed or not. Instead the Census response is very useful to determine the
activity status of those units having ambiguous combinations of signals of activity from the administrative files.

3.2 Over and under covered units

After having estimated the probabilities of belonging to a latent class, we must define a threshold value \( p_0 \) so that units with posterior probabilities \( p (1/Y) < p_0 \) are considered as non-activity at the Census date. The threshold value is chosen with the aim of maximizing the concordance between the model, CIS and ASIA activity status for the units present in cells \( n_{11} \) and \( n_{22} \), which are respectively the units active and the units not active for CIS and ASIA. Choosing a threshold value equal to 0.5 we obtain 2,699,141 non-active enterprises and 4,049,380 active enterprises.

To obtain a measure of the coverage error of the two data sources the following two tables are analyzed.

Table 4: Number of enterprises estimated as not active, by ASIA and CIS activity status

<table>
<thead>
<tr>
<th></th>
<th>CIS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In</td>
</tr>
<tr>
<td>ASIA</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>Out</td>
</tr>
<tr>
<td>Total</td>
<td>56,204</td>
</tr>
</tbody>
</table>

Table 5: Number of enterprises estimated as active, by ASIA and CIS activity status

<table>
<thead>
<tr>
<th></th>
<th>CIS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In</td>
</tr>
<tr>
<td>ASIA</td>
<td>In</td>
</tr>
<tr>
<td></td>
<td>Out</td>
</tr>
<tr>
<td>Total</td>
<td>3,343,317</td>
</tr>
</tbody>
</table>

Table 4 provides an estimation of over-coverage error made by ASIA and CIS. In particular (i) about 373,000 units considered as active in ASIA, have been classified as not active at the day of the Census, according to the model (ASIA over-coverage); (ii)
about 56,000 units, surveyed by the Census, have been classified as not active at the day of the Census (Census over-coverage).

Again, it is underlined that the 373,000 units cannot be considered all over-coverage of the Business Register. In fact, taking into account the information present in the administrative archives or the information of the enumerators, 50% of those units are units no longer active at the data in reference or units that have started-up after the date in reference, but active for at least a month in the year. Moreover, the analysis of the main features of those enterprises, such as employees and turnover, shows that they are mainly enterprises with zero employees and low turnover.

Table 5 provides an estimation of under-coverage error made by ASIA and CIS. In particular (i) about 706,000 units are indicated as active in ASIA and have not been surveyed in the Census (Census under-coverage); (ii) about 131,000 units are surveyed by the Census and are not indicated as active in ASIA (ASIA under-coverage). Overall, the active enterprises at 22nd October, 2001 amount to 4,049,380.

In the following tables 6 and 7 are shown the distributions by class of employees and economic activity sectors of the units estimated as active at 22nd October, 2001, compared with the units active in ASIA.

Table 6: N° of unity by economic activity sectors

<table>
<thead>
<tr>
<th>Economic activity sector</th>
<th>N° of enterprises in ASIA</th>
<th>%</th>
<th>N° of enterprise the 22nd October 2001</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry</td>
<td>575484</td>
<td>13,41</td>
<td>551365</td>
<td>13,62</td>
</tr>
<tr>
<td>Construction</td>
<td>541203</td>
<td>12,61</td>
<td>513251</td>
<td>12,67</td>
</tr>
<tr>
<td>Commerce</td>
<td>1335749</td>
<td>31,13</td>
<td>1230325</td>
<td>30,38</td>
</tr>
<tr>
<td>Service</td>
<td>1838986</td>
<td>42,85</td>
<td>1754439</td>
<td>43,33</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4291422</strong></td>
<td><strong>100,0</strong></td>
<td><strong>4049380</strong></td>
<td><strong>100,0</strong></td>
</tr>
</tbody>
</table>

Table 7: N° of unity by class of employees

<table>
<thead>
<tr>
<th>Class of employees</th>
<th>N° of enterprises in ASIA</th>
<th>%</th>
<th>N° of enterprise the 22nd October 2001</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3151238</td>
<td>73,43</td>
<td>2908011</td>
<td>71,81</td>
</tr>
<tr>
<td>1--9</td>
<td>980543</td>
<td>22,85</td>
<td>981461</td>
<td>24,24</td>
</tr>
<tr>
<td>10--49</td>
<td>137642</td>
<td>3,21</td>
<td>137853</td>
<td>3,40</td>
</tr>
<tr>
<td>50--99</td>
<td>12312</td>
<td>0,29</td>
<td>12347</td>
<td>0,30</td>
</tr>
<tr>
<td>100--249</td>
<td>6481</td>
<td>0,15</td>
<td>6497</td>
<td>0,16</td>
</tr>
<tr>
<td>&gt;249</td>
<td>3206</td>
<td>0,07</td>
<td>3211</td>
<td>0,08</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4291422</strong></td>
<td><strong>100,00</strong></td>
<td><strong>4049380</strong></td>
<td><strong>100,00</strong></td>
</tr>
</tbody>
</table>
4 Final remarks

It is important to underline that the methodology shown in this paper has been developed with the aim to measure the under-coverage of Census. The correction of Census results without going to a post enumeration survey is based on the use of administrative information with the advantage to reduce costs and the possibility to integrate individual units and related characters. In fact this method allows to integrate missed units taking all the information from the Business Register ASIA, leaving to a statistical model the purpose to estimate the “true” number of active units at the Census date.

The result of having estimated around 700 thousands units under-covered by the Census (against the 200,000 of the 1991 Census) can be, in part explained by the different estimation technique, but mainly by the fact that the economic sectors and the typology of enterprises, usually under-covered by the Census, are the one that have grown up more in the last 10 years (e.g. professional and personal services). Moreover, the high number of under-covered units, as shown in this exercise, highlight that the traditional door to door survey technique have become an inappropriate and obsolete tool to describe the universe of active enterprises because of the structure of the economic system. Instead, monitored registers supply a complete and update information of the universe of interest and sample surveys can be thought only to validate the register.

Of course, the result of having around 370 thousands of units active for Asia but not taken into consideration for the integration in the Census show the problem of over-coverage of the Business Register. The difference in scope of the two data sources, active units at a given date for the Census and active units in the year for the Business Register, explain part of the difference. However it is accepted the evidence that a part of active units in the Business Register could be false active units; those are, in general, units without employees, with a turnover less than 10 thousands euros and therefore units that lie in an area of uncertainty with regards to the activity status.

References


G. Garofalo, M. Porcu and C. Viviano. The use of administrative sources to model the
activity status of the enterprises. *16th Roundtable on Business Survey frames*, 20-25
October, Lisbon, 2002.

H. Haberman. Log-Linear Models for Frequency Tables Derived by Indirect
5 (Sep.), Chicago, p. 911-924, 1974.


M. Martini. Necessità e possibilità di un registro statistico delle imprese in Italia. *Il
Registro Statistico Europeo delle Imprese*, a cura di S. Biffignandi e M. Martini,

C. Viviano. Demografia d’impresa: l’utilizzo di tecniche di abbinamento per l’analisi


Author’s address:

Dr. Corrado Abbate
Istat - Istituto Nazionale di Statistica
Via Rava 150
00142 Roma
Italy
Tel: +39 06 5952 / 4676
Elec. Mail: abbate@istat.it
http://www.istat.it

Dr. Danila Filliponi
Istat - Istituto Nazionale di Statistica
Via Rava 150
00142 Roma
Italy
Tel. +39 06 5952 / 4662
Elec. Mail: dafilipp@istat.it
http://www.istat.at

Dr. Caterina Viviano
Istat - Istituto Nazionale di Statistica
Via Tuscolana 1776
Roma