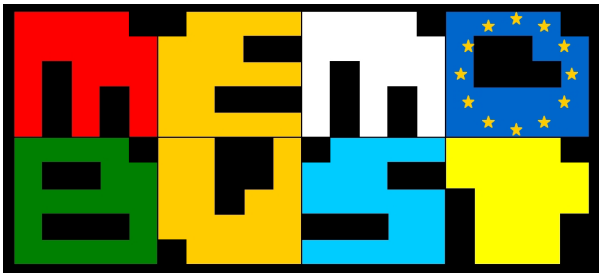


This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Theme: Business Demography

Contents

General section

1. Summary

The term business demography is used to cover a set of variables which explain the characteristics and demography of the business population. The creation of new enterprises and the closure of unproductive ones are considered important indicators of the business dynamics.

There is a large demand for information on business demography both at national and international level. At European level, demands are for coherent and comparable data across the members of the European Statistical System (ESS). The European Commission, as key customer, has assured its commitment to a policy that promotes entrepreneurship as an essential instrument for improving competitiveness and generating economic growth and job opportunities since its communication to the Council on 'Promoting Entrepreneurship and Competitiveness'. The support of entrepreneurship and entrepreneurial dynamics, the presence of which can be revealed by the analysis of business demography statistics over time. As a consequence, there is high demand for comparable data on business demography for the purposes of monitoring and policy formulation.

This module aims to provide theoretical guidance in the production of data on business demography. It has been developed taking Eurostat-OECD Manual on Business Demography Statistics into account, used in the EU business demography harmonised data collection.

After a short introduction on the importance to collect business demography data, in the second section we give a general description of two main topics: first we describe the methodology and indicators for business demography statistics such as: Enterprise Births, Enterprise Deaths and Surviving Enterprises; second we describe methodology and indicators for entrepreneurship such as: Employer Enterprise Births, Employer Enterprise Deaths, Employer Surviving Enterprises, High-Growth Enterprises and Gazelle Enterprises.

2. General description

2.1 Introduction

The term business demography is used to cover a set of variables which explain the characteristics and demography of the business population. The creation of new enterprises and the closure of unproductive ones are considered important indicators of the business dynamics.

Business demography data are currently produced to fulfil the European regulation, for the European Union (EU) and European Free Trade Association (EFTA) members; it is used to satisfy the requirements for producing the Structural Indicators used for monitoring progress of the Lisbon process, regarding business births, deaths and survival. It also provides key data for the joint OECD-Eurostat "Entrepreneurship Indicators Programme".

A new methodology has been developed for the production of data on enterprise births (and deaths), that is, enterprise creations (cessations) that amount to the creation (dissolution) of a combination of production factors and where no other enterprises are involved. This methodology aims to harmonise how business demography is computed within the ESS. The present module is an overview of this methodology fully described in the Eurostat-OECD Manual on Business Demography Statistics. The

reader who is interested in more details and background can read this manual that is available from the Eurostat website.

The methodology and definitions adopted in this module are also based on the Business Registers Recommendations Manual and indications, as the Business Registers is the source for the Business Demography data. In addition we focus on business demography output on a yearly basis.

2.2 Main Components of Business Demography

The real enterprise birth and death definitions must take into consideration the enterprise definition. A fundamental requirement in measuring business entries (creation) and exits (destruction) concerns the definition of a **business** itself. Statistical System distinguishes a number of unit type such as: establishment, enterprises, legal units, local kind of activity units, etc. (See the module "Statistical Registers and Frames – The Statistical Units and the Business Register".) In order to harmonise the business demography data collections, the statistical unit to be used is the enterprise. According to the statistical units Regulation (Council Regulation (EEC) No 696/93 of 15 March 1993) "*The enterprise is the smallest combination of legal units that is an organisational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources. An enterprise carries out one or more activities at one or more locations. An enterprise may be a sole legal unit.*"

In the next subsections we give an overview of the main Business Demography components: Enterprise Births, Enterprise Deaths and Surviving Enterprises. In addition we give an overview of the main Entrepreneurship components: Employer Enterprise Births, Employer Enterprise Deaths, Employer Surviving Enterprises, High-Growth Enterprises and Gazelle Enterprises.

For each of these indicators we give the definition, a description of identification process and, where it was possible, a case study.

2.3 New Enterprises (Births)

The number of Enterprise Births (*RB*) is a key variable in the analysis of Business Demography; other variables such as the survival and growth of newly born enterprises are relevant and related to this concept. The production of statistics on births must be based on a clear and acceptable definition and interpretation.

2.3.1 Concept

According to the Commission Regulation No 2700/98, we define a new Enterprise (Birth) as:

A birth amounts to the creation of a combination of production factors with the restriction that no other enterprises are involved in the event. Births do not include entries into the population due to: mergers, break-ups, split-off or restructuring of a set of enterprises. It does not include entries into a sub-population resulting only from a change of activity. An enterprise creation can be considered an enterprise birth if new production factors, new jobs in particular, are created.

Inclusions

Enterprises started by a person who previously performed the same activity, but as an employee should be included in the statistics on enterprise births.

Exclusions

Events leading to a creation of a new enterprise, but which should be excluded from the statistics on enterprise births are:

1. Enterprises that are created by merging production factors or by splitting them into two (or more) enterprises (breakups, mergers, split-offs, restructuring)[1];

2. Newly created enterprises that simply take over the activity of a previously created enterprise (take-over)[2];

3. Any creations of additional legal units/enterprises solely for the purpose of providing a single production factor (e.g., the real estate or personnel) or an ancillary activity (see note below) for an existing enterprise.

4. An enterprise that is registered when an existing enterprise changes legal form. E.g., a successful sole proprietor moves operations from his home to another location and at the same time changes the legal form of the enterprise to a limited liability company.

5. Reactivated enterprises if they restart activity within two calendar years[3].

2.3.2 The identification process

Users that want to know how many new enterprises have been created in a specific year usually compare populations of active enterprises referred to two adjacent periods (t, t-1).

According to the Eurostat-OECD Manual on Business Demography Statistics the Business register serves as primary and preferred source of information for business demography statistics. The main reasons that we choice this source are: -there is a degree of harmonisation of statistical business registers in EU Member States following the adoption of the business statistical Regulation (Council Regulation (EEC) No 2186/93); - under the EU Regulation, Member States are required to hold data on the enterprise, a harmonised statistical unit that removes the impact of different legal and organisational infrastructures; - using data from business registers minimised the burden on businesses. The BR contents the population of active enterprises. This population consists of all enterprises that had either turnover or employment at any time during the reference period. A unit present into the BR is tested according to a methodology determining whether it is active or not active on the basis of these signals of activity (presence of turnover or employment); on the basis of this information entries can be identified in a population of active units.

The identification of real enterprise births is based on the application of a complex and expensive procedure that is build up on a set of automated and manual steps aiming to eliminate (i.e., to identify) the "non-real" components from the set of new enterprises (entries) units into the BR for a reference period t.

A formalisation of the demographic flow is given by the following equation:

$$N_t = N_{(t-1)} + E_t - U_{(t-1)} = A_{t-1,t} + E_t \qquad (1)$$

[1] For more details see the Glossary of this module.
[2] For more details see the Glossary of this module
[3] For more details see you Eurostat-OECD Manual of Business Demography Statistics pp.34-35

where:

N_t = population of active enterprises in a reference period *t*

$N_{(t-1)}$ = population of active enterprises in a reference period (*t*-1)

E_t = Entries in a reference period *t*. We define Entries in year *t* as the subset of the population of active enterprises in year *t*, which have taken up economic activity between 01.01 and 31.12. These new enterprises are identified as enterprises that are only present in year *t* and not in year (*t*-1).

$U_{(t-1)}$ = Exits from the reference period (*t*-1). We define Exits as the subset of the population of active enterprises in year (*t*-1), which have ceased their economic activity in year (*t*-1). These enterprises are identified as enterprises that are only present in year (*t*-1) and not in year *t*.

$A_{t-1,t}$ = active enterprises both in the period *t* and (*t*-1)

Entries (and exits) are affected only by the structure of signals of activity (presence/absence of turnover or employment) or by other national methodology establishing if the enterprise is active or not.

Regarding the new enterprises (entries), some steps are followed to reach the final result, i.e., the real enterprise births (data on enterprise deaths are produced with a "mirror" process). These steps are performed for the whole population of active enterprises.

We summarise all the process into three steps each determining a relative subset of data.

Step 1 – The first step consists in comparing three subsequent populations of active enterprises (N_t, N_{t-1} and N_{t-2}). This operation allows to eliminate reactivated units.

A merge by identification code (or fiscal code) of the three subsequent populations of active enterprises determines the following pattern:

***t*-2**	***t*-1**	***t***	**Output**	
N_{t-2}	Missing	N_t	E_r	Reactivations
Missing	Missing	N_t	E_1	Entries

In detail, it is necessary to split the Entries (E_t) into two components: 1) E_r the subset of reactivations; 2) E_1 the subset of entries cleaned from reactivations.

The resulting equation (1) becomes:

$$N_t = A_{t-1,t} + [E_1 + E_r] \qquad (2)$$

Step 2 – Mergers and other events of structural changes involving enterprises' re-organisation cause creations of new units into the register. The identification of births is carried out by eliminating creations due to these events (break-ups, split-offs, mergers and take-overs). It possible to have information about these events from different sources (pilot studies, administrative sources, statistical sources, survey etc…).

In the previous equation E_1 is split into two components according to the following formula:

$$N_t = A_{t-1,t} + [(E_2 + E_{ev}) + E_r] \quad (3)$$

where:

E_2 = entries not due to events of structural changes

E_{ev} = entries due to events of structural changes

Step 3 – The further step consists in identifying and excluding those false entries because continuing the activity of some other units. Continuity rules are assigned through a general **matching process**, that matches units according to economic activity and location, name and location, economic activity and name, and the use of other nationally available information, such as telephone number, date of registration/deregistration at the administrative source, employer/employee links etc. This step contains also the results of manual controls for large enterprise births. E_2 component is split as follows:

$$N_t = A_{t-1,t} + [(RB_t + E_{cont}) + E_{ev} + E_r] \quad (4)$$

where:

RB_t = Real births in the year *t*

E_{cont} =entries due to the continuity rules application

The RB_t component identifies the subset of real enterprise births.

2.3.2.1 The notion of continuity

Within the study of enterprises population development it is relevant to identify whenever a change happens whether it produces a discontinuity of the unit: an enterprise is considered to be continued if it modifies without any significant change in its identity, in terms of its production factors. The production factors include the set of means (employment, machines, raw material, capital management, buildings) that the enterprise uses in its production process and leading to the output of goods and services. It is clear that measuring the continuity of all production factors and weighting them can be quite difficult and costly. For those reasons Eurostat suggests, as a practical criterion, to use precise variables available in the register that are correlated to the most important production factors for identifying the enterprise: the basic hypothesis is that a change in such variables would stand for a change in the production factors. The variables (characteristics) considered are:

⇒ The controlling legal unit of the enterprise (**N**) - The continuity of the management of the enterprise may be assumed to be positively correlated with the continuity of the controlling legal unit. The same may be assumed for some immaterial assets.

⇒ The economic activities carried out (**S**) - Continuity of the four-digit NACE Rev.2 code of the principal activity may be assumed to be positively correlated with the continuity of the production factors, especially employment, machines and equipment, land and buildings.

⇒ The locations where the activities are carried out (**L**) - The continuity of the locations where the activities are carried out is of course closely linked to the continuity of the land and buildings used by the enterprise.

The empirical rule suggested is that an enterprise is considered to be discontinued if at least two over three modifications in the previous factors occur. Continuity rules reflect a notion of identity based on the consideration that the enterprise is a set of specific resources, procedures and relationships with the environment. In the suggested rules an element of discontinuity is introduced when changes are "of great extent" and quick. The concrete applicability of such rules must be evaluated according to the economic structure in which they have to perform, because of the peculiarity of each country. For instance, for some domains of study, as for demography of very small enterprises, it does not make sense to separate the legal subject (the entrepreneur) from the statistical subject (the enterprise); for such cases a new controlling legal unit becomes a factor producing discontinuity even if it is the only one to change.

For the continuity rules, several software and standard matching systems can be applied according to the national sources and experiences.

2.3.3 Case study: the matching process in the production of Italian Business Demography data

In Italy, the identification of the "real" components is based on the application of Record Linkage (RL) techniques. The matching process matches on name, economic activity and location of enterprise.

A reminder of the main elements for a record linkage is needed, a detailed description can be found in the module "Micro-Fusion – Object Matching (Record Linkage)". Let *A* and *B* be two files respectively containing records *na* and *nb*. The record linkage procedures compare each unit in *A* with each unit in *B*. The object of interest of the record linkage problem is the pair of units (*a*,*b*) of a set $A \times B$. This set can be partitioned in a set M of pairs representing the same business entity and a set U of pairs representing different entities. Record linkage methods aim at determining which pairs belong to the set M. We can distinguish two approaches: deterministic and probabilistic. For both these two approaches some preliminary steps are necessary.

1) Identification of populations (file *A* and file *B*) and units.
2) Selection of match characteristics (components). This selection should be based on the quality of the available data, their discriminating power and the purpose of the study.
3) Standardisation, parsing and string comparison of match variables.
4) Blocking (comparison reduction). The size of the files usually considered does not often allow explicit consideration of comparison of all pairs of records, and usually only pairs with some common characteristics are actually compared, by using blocking criteria.
5) Agreement/disagreement rules to evaluate the similarity of records and weighting system to take into account that some information is more important than other; for example, an agreement on economic activity can contributes less than an agreement on enterprise names.
6) Determination of thresholds. A match is accepted only if its level of agreement is higher than a designated "threshold" level.

For Business Demography purpose, a record linkage technique is applied to find pairs of records across files that correspond to the same entities and to perform matching to identify "real" births and deaths according to the continuity rules.

2.3.3.1 Identification of populations (file A and file B) and units

If the purpose is the identification of **Real Births** the populations compared are:

File A: Stock of enterprises in the year *t*

File B: Entries of enterprises in the year *t*

and

File A: Exits of enterprises in the year (*t*-1)

File B: Entries of enterprises in the year (*t*-1)

If the purpose is the identification of **Real Deaths** the populations compared are:

File A: Stock of enterprises in the year *t*

File B: Exits of enterprises in the year (*t*-1)

and

File A: Exits of enterprises in the year (*t*-1)

File B: Entries of enterprises in the year *t*

2.3.3.2 Selection of match characteristics (variables)

After having identified the populations, it is necessary to select the matching variables (components). According to **continuity rules**, for Business Demography the ***matching variables*** are: 1) Enterprise name; 2) Address; 3) Economic activity code (Nace Rev.2 at 4 digits). Other variables are used as support for a better specification of the enterprise name: the Fiscal code and the legal form. When matched variables are identified, their standardisation, parsing and string comparison functions follow.

2.3.3.3 Standardisation, parsing and string comparison of match variables

For Business Demography appropriate standardisation and parsing of enterprise name and address components is crucial for computerised probabilistic or deterministic record linkage.

Standardisation of **enterprise name** consists in three main activities:

1) Parsing: to divide the name in sub-components;

2) Editing of each sub-component of the name;

3) Use of the dictionary to attribute to each sub-component its meaning.

The main difficulty with enterprise name is that even when they are properly parsed, the identifying information may be indeterminate.

Standardisation of **enterprise address** consists in two main activities:

1) Parsing: to divide the free form address in three sub-components toponymic (T) (address type as: square, road, avenue etc...), street name (sNa) and street number (sNb);

2) Editing of each sub-component of the address.

2.3.3.4 Blocking (comparison reduction)

According to blocking criteria, two alternative blocks are used: **municipality** and **postal code +economic activity code (NACE Rev.2 - 3 digits)**.

After blocking criteria, agreement-disagreement rules on enterprise name and address are applied. These rules came from a set of specific values.

2.3.3.5 Agreement rules

Agreement rules for enterprise name

According to the Italian enterprise structure, enterprise name takes different formats according to its legal form. We aggregate the legal form (J) into 3 classes: (**I=** Sole proprietorship; **Sp=** Partnership; **Sc=** Limited liability company). The adopted rules are:

When comparing J=(**I**, **Sp**) or J=(**Sp**, **Sp**) $\Rightarrow$ Surname and Name are compared.

If J=(**Sp,Sc**) or J=(**Sc,Sc**) $\Rightarrow$ Significative Sub-components (words) are compared.

If J=(**I**, **I**) or J=(I, **Sc**) $\Rightarrow$ We have a disagreement by default.

Agreement rules for enterprise address

We have an outcome for each component of address. The possible outcomes on the toponymic (**T**) component are **E**qual or **D**iffer or **M**issing. To compare street name (**sNa**), a simple string comparator is used. This string comparator divides each sNa in consecutive sub-strings of 3 characters, finds the number of common sub-strings and computes the percentage of common sub-strings on the length of the shorter sNa. Finally the possible outcomes on street number (**sNb**) component are **E**qual or **D**iffer or **M**issing. Enterprise address matching rules are delineated according to different combinations of components outcomes. For example when you compare two addresses of two different enterprises we have an agreement if toponymic components are equal or missing for both addresses, the percentage of matched address name sub-strings is high (>=80%) and the street numbers are equal. We have a disagreement when the toponymic components are different.

Agreement rules for enterprise economic activity

NACE Rev.2 codes at 4 digits are compared to produce outcomes. We have an agreement if NACE codes of two enterprises are equal, we have a partial agreement if NACE codes are compatible and we have a disagreement if NACE codes are different.

Outcomes defined by each *variable* rule constitute the components of the comparison vector γ.

2.3.3.6 Determination of thresholds

For Business Demography purpose, ad hoc matching decision rule is applied to divide the pairs in link and non-link. In presence of a comparison vector having two over three agreements the pair is a match (according to the continuity). There is an exception: if the comparison vector γ is composed by a disagreement on the enterprise name, an agreement on the enterprise address, an agreement on

enterprise economic activity (γ=(D,A,A)) and the NACE code belong to a list of economic activity "at risk" (construction, hotel..) then the pair is declared as non-matched.

Once you identify the two sets of pairs M and U it is possible to determine the set of Entries due to the continuity rules application (E_{cont}).

$$RB_t = E_t - E_r - E_{ev} - E_{cont}$$

where:

E_t =Entries in a reference period *t*

E_r =reactivations: enterprises active both in year (*t*-2) and *t*, but not active in (*t*-1)

E_{ev} = entries due to events of structural changes

E_{cont} =entries due to the continuity rules application (Record-Linkage).

For the sake of consistency, and in line with user needs, the method of comparing populations of active enterprises used for the production of data for enterprise births should also be followed for enterprise deaths. This will also help to gain from synergies in processing.

2.4 Ceased Enterprises (Deaths)

2.4.1 Concepts

According to the Commission Regulation No 2700/98, we define a Ceased Enterprise (Death) as:

A death amounts to the dissolution of a combination of production factors with the restriction that no other enterprises are involved in the event. Deaths do not include exits from the population due to mergers, take-overs, break-ups and restructuring of a set of enterprises. It does not include exits from a sub-population resulting only from a change of activity.

2.4.2 The identification process

Like the real enterprise births, the identification of real enterprise deaths is based on the same procedure that is built up on a set of automated and manual steps aiming to identify the "non-real" components from the set of exit units into the BR. The cessations in year *t* are a subset of the population of active enterprises in year *t*, which have ceased their economic activity between 01.01 and 31.12. They can be identified by comparing the population of active enterprises in year *t* with the population of active enterprises in year (*t*+1). Exits are identified as enterprises that are only present in year *t*. Like enterprise births, exits should be checked for reactivation in the following two calendar years, because enterprises dormant for less than two years are considered reactivations and therefore not deaths. An enterprise death occurs only if the unit has been inactive for at least two years. In order to find the events that were not real enterprise deaths, but rather cessations due to events like break-ups, mergers or take-overs, a matching criteria (as for enterprise births) should be carried out. Finally the continuity rule is applied to identify the cases where another unit is involved in the cessation of the enterprise. As for enterprise births, the matching should consider name, location and economic activity.

According to this procedure for the identification of real enterprise deaths in year *t*, populations in (*t*+1) and (*t*+2) are needed.

In practice, at the time *t*, it is possible identify only provisional real deaths(*t*-1) (due to the unknown reactivations at year *t*+1).

For the above reasons, the following procedure for the identification of enterprise real deaths(*t*-1) determines a one year lag difference compared to the real births(*t)*. To improve timeliness of real deaths, it is necessary to develop further estimation techniques.

2.4.3 Case study. Real Deaths estimation method: the Italian methodology

To estimate Real Deaths at year *t*, a time series of enterprise deaths rates of the previous five years is analysed. In addition, nationally available administrative source, the Social Security data SS with information at (*t*+1) is used to identify employer enterprise deaths.

The estimation method consists in the building of strata made up by three variables: Economic activity, Legal form and Size class. In total we have about 5,600 strata. For each *j*-th stratum, death rate (*t*) is given by a average over the period [*t*-5, *t*-1]. Then it is reweighted using the ratio=number of SS enterprises/ number of BR enterprises if strata presents employees.

Formally:

$$death\_rate(t)_j = \left[\frac{1}{5}\sum_{1=1}^{5} death\_rate(t-i)_j\right] w_j \qquad \text{for } j=1,\ldots,n \text{ and } i=1,\ldots,5.$$

where:

$$w_j = \begin{cases} 1 & \text{if } j \text{ is with employees } =0 \\ \left(1-\left(n°enterprises SS\right)_{(t+1)}/\left(n°enterprises BR\right)_{(t)}\right) & \text{if } j \text{ is with size classes}>0^4 \end{cases}$$

If the number of SS enterprises in (*t*+1) is greater than the number of BR enterprises in year (*t*) then w_j <0 and consequently the number of deaths in year *t* are 0.

The number of deaths year *t* in stratum *j* is given by the weighted average rate over period multiplied by the number of active units in year *t*. This technique allows to estimate only provisional real deaths in year (*t*) (due to the unknown reactivations at year *t*+2).

Actually we are studying a new methodology that takes into account the timeless information about the number of deregistrations in the Chamber of Commerce (administrative data). These types of information is useful to forecast sudden change in death rate due to social or economic events.

2.5 Surviving Enterprises

2.5.1 Concepts

According to the Commission Regulation No 2700/98, the definition of a Surviving Enterprise is:

[4] If the number of SS enterprises in year (*t*+1) is equal to the number of BR enterprises in the previous year (*t*) then w_j =0 and consequently the number of deaths in year *t* are 0.

an enterprise born in year (*t*-1) or having survived to year (*t*-1) from a previous year is considered to have survived in year *t* if it is active in terms of turnover and/or employment in any part of year *t* (= survival without changes).

If the enterprise is not active in year *t* it has survived if its activity is taken over by a new enterprise set up specifically to take over the factors of production of that enterprise in year *t* (= survival by take-over).

This definition of survival excludes cases where enterprises merge, or are taken over by an existing enterprise in year (*t*-1). In these cases the continuation of the enterprise involves an enterprise established before year *t* and therefore the enterprise is not considered to have survived.

To ensure consistency between data on births and survivals, it is important that the identification of cases where an enterprise is taken over by a new enterprise is based on the use of the same information as when evaluating whether a new enterprise is a birth or not. Therefore an enterprise is only counted as survived, if the enterprise that takes over the factors of production is a new enterprise.

2.5.2 The identification process

The production of statistics on survival can be based on three populations, which are all part of the production of the statistics on births: a) Births in year (*t*-1), or enterprises having survived to t from a previous year (RB(*t*-1)); b) Active enterprises in year *t* (A(*t*)); c) Enterprises that have commenced activity in year *t* with the purpose of taking over the factors of production of an enterprise that commenced activity before *t* (TO(*t*)). By matching these populations by identification codes (like the BR code, or a fiscal code) it is possible to have the following outcomes:

- RB(*t*-1) is present neither in A(*t*) nor in TO(*t*). RB(*t*-1) is not survived in year *t*.
- RB(*t*-1) is present only in A(*t*). RB(*t*-1) is survived in year *t* without any change.
- RB(*t*-1) is present only in TO(*t*). RB(*t*-1) is survived in year *t* with change (take-over took place in *t*-1).

2.6 Business Demography Indicators

The business demography data will be used to produce additional indicators related to enterprise births, deaths and survival such as the following:

- Births/Deaths as a percentage of the population of active enterprises (birth/death rates).
- Births/Deaths by size class.

Additional indicators will be produced to demonstrate the impact of the newly born/death enterprises on the economy:

- Persons employed in newly born/death enterprises in year *t* as a proportion of the total number of persons employed in the population of active enterprises in year *t* (both in head counts).
- Employees in newly born/death enterprises in year *t* as a proportion of number of persons employed in newly born/death enterprises in year *t* (both in head counts).

The first of these indicators reflects the employment creation/destruction potential of newly born/death enterprises. The second reflects the potential employment creation/destruction going beyond the entrepreneurs themselves. Other possible indicators are:

- Gross Job Turnover (**GJT**)= the sum of the number of jobs by creation and destruction during a year, it is a measure of job reallocation.
- Net Job Turnover (**NJT**) = the difference between the number of jobs by creation and the number of job of destruction.

Possible indicators for surviving:

- Survival rate in years *t*, as ratio between the number of enterprises born in year (*t-i)* (i=1 to n) and survived in year *t* and the number of enterprise births in year (*t-i*).
- The number of persons employed in surviving enterprises in their *i*-th year of activity divided by the numbers of persons employed in real births in the initial year
- The number of persons employed in *t* divided by the number of survival enterprises in *t*.
- Although it is not an indicator, the study of the enterprises cohorts surviving from the year of birth (*t-i*) (usually *i*=1 to 5) to year *t* is very interesting. The analysis of the characteristics of such enterprises could describe the behaviour of younger enterprises in terms of employment growth or in terms of economic variables growth such as turnover and value-added over time. Further analysis focuses on the study of efficiency and productivity of such enterprises with respect to the older enterprises.

2.7 Background on Entrepreneurship

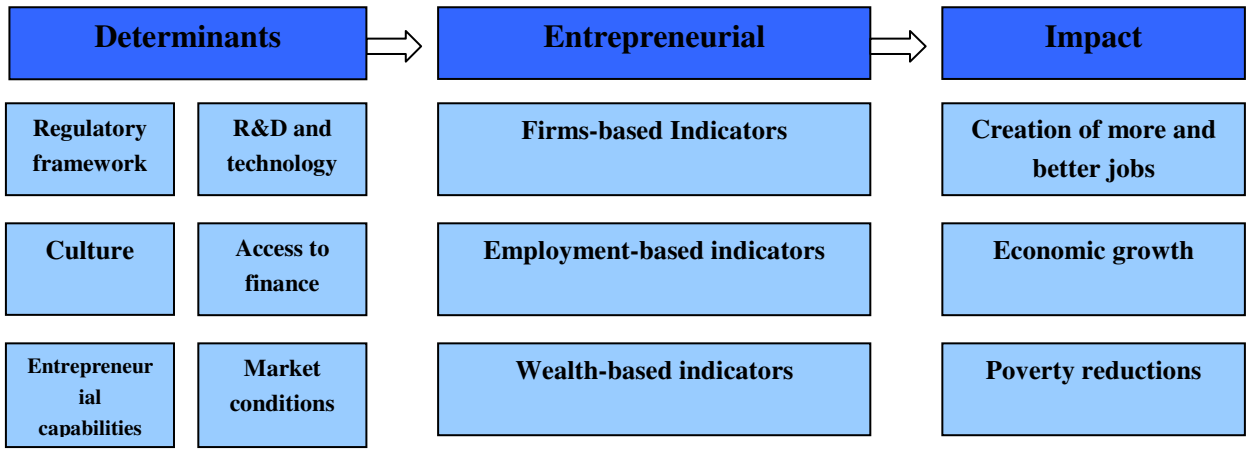
In recent years, the political and academic interest in entrepreneurship and its determinants has grown. Policy makers give more importance to the development of high growth enterprises and to the conditions that foster this growth. The Entrepreneurship Indicators Programme (EIP), launched by OECD in September 2006, has as main goal to build internationally comparable statistics on entrepreneurship and its determinants. In 2007, Eurostat joined forces with the OECD to create a joint OECD-Eurostat EIP, and work began with the development of standard definitions and concepts as a basis for the collection of empirical data.

The OECD-Eurostat approach has tried to combine the more conceptual definitions of entrepreneurship with (available) empirical indicators. The following definitions were established:

- ENTREPRENEURS are those persons (business owners) who seek to generate value through the creation or expansion of economic activity, by identifying and exploiting new products, processes or markets.
- ENTREPRENEURIAL ACTIVITY is enterprising human action in pursuit of the generation of value through the creation or expansion of economic activity, by identifying and exploiting new products, processes or markets.
- ENTREPRENEURSHIP is the phenomenon associated with entrepreneurial activity.

Given the multifaceted nature of entrepreneurship and the myriad factors that may affect, a simple entrepreneurship model was proposed as a first step towards establishing a framework for the

development of empirical indicators that are both relevant and available. These indicators are grouped together into three themes: *Determinant* (factors that impede or motivate entrepreneurship); *Entrepreneurial performance* (measures that provide information of the state of entrepreneurship); and, *Impacts* (outcomes of that performance on the economy as a whole).



Focusing the analysis on entrepreneurial performance, it is possible to identify three sets of indicators: the first set is relating to firms-based indicators such as employer firm birth rate, employer firm death rate; the second and third set is relating to the entrepreneurship effects in terms of employments and wealth such as high-growth firm rate by employment, gazelle rate by employment, high-growth firm rate by turnover, gazelle rate by turnover.

2.8 Employer Enterprise Birth

2.8.1 Concepts

With reference to entrepreneurship we are interested in the subpopulation of enterprises with one or more employees. While the "standard" BD on enterprise birth covers all units (without any threshold concerning very small units) the employer enterprise birth focuses on the enterprises with at least one employee. Therefore this new definition of employer enterprise births (EEB) is added to complement the enterprise birth.

By definition there are two conditions which qualify an enterprise as an employer birth: it was an enterprise birth in year *t* (real birth), and had at least one employee in the year of birth, or it existed before year *t*, was not an employer for the two previous years and had at least one employee in year *t* (entry by growth). Results on take-overs should be available from the methodology used to identify enterprise deaths. Where possible, the information on units that took over other units (which ceased to exist but were not deaths) should be used to identify enterprises that reached the one employee threshold by taking over another one. These should be removed from the population of births by growth.

2.8.2 The identification process

To identify the Employer Enterprise Births it is necessary to have the following sets of population:

N_t population of active enterprises (with zero and >zero employees) in year *t*

$N(1)_t$ population of active enterprises (with >zero employees) in year *t*

$N(0)_{t-1}$ population of active enterprises (with zero employees) in year *t*-1

$N(1)_{t-1}$ population of active enterprises (with >zero employees) in year *t*-1

$N(0)_{t-2}$ population of active enterprises (with zero employees) in year *t*-2

$N(1)_{t-2}$ population of active enterprises (with >zero employees) in year *t*-2

$RB(1)_t$ real births (with >zero employees) in year *t*

A merge by identification code of the three years populations determines the following patterns:

***t*-2**	***t*-1**	***t***	***Output***
$N(0)_{t-2}$	$N(0)_{t-1}$	$N(1)_t$	**Births by Growth**
$N(0)_{t-2}$	$N(1)_{t-1}$	$N(1)_t$	-
$N(0)_{t-2}$	missing	$N(1)_t$	**Births by Growth**
$N(1)_{t-2}$	$N(0)_{t-1}$	$N(1)_t$	-
$N(1)_{t-2}$	$N(1)_{t-1}$	$N(1)_t$	-
$N(1)_{t-2}$	missing	$N(1)_t$	-
missing	$N(0)_{t-1}$	$N(1)_t$	**Births by Growth**
missing	$N(1)_{t-1}$	$N(1)_t$	**-**
missing	missing	$RB(1)_t$	**Real Births with at least one employee**

In summary, the Employer Enterprise Births in year *t* (EEB_t) are the Real Births with at least one employee $RB(1)_t$ and the active enterprises in year *t* with at least one employee ($N(1)_t$) which are in population $N(0)_{t-2}$ or $N(0)_{t-1}$ or both and which are neither in population $N(1)_{t-2}$ nor in $N(1)_{t-1}$ (Births by Growth). In order to remove from births by growth some active units that grow because events of takeover the following links for *t* have been identified: a) Employer Enterprise Births (EEB_t) linked to $Exits_{t-1}$ that cease for events; b) Employer Enterprise Births (EEB_t) linked to Active units ($A_{t-1,t}$) that shrink for events; c) Employer Enterprise Births (EEB_t) linked by continuity rules to $Exits_{t-1}$.

2.9 Employer Enterprise Death

2.9.1 Concepts

Like employer enterprise births there are two conditions which qualify an enterprise as an employer death (EED): it was an enterprise death in year *t* (real death), and had at least one employee in the year of death, or it had at least one employee in year *t* and continued to exist in years *t*+1 and *t*+2 without employees (death by decline). Results on split-offs should be available from the methodology used to identify enterprise births. Where possible, the information on new enterprises that were split-offs should be used to identify original enterprises that moved below the one employee threshold because a new unit emerged from a split-off. These original enterprises should be removed from the population of deaths by decline.

2.9.2 The identification process

To identify the Employer Enterprise Deaths in a reference year *t*, it is necessary to have the following sets of population:

N_t population of active enterprises (with zero and >zero employees) in year *t*

$N(1)_t$ population of active enterprises (with >zero employees) in year *t*

$N(0)_{t+1}$ population of active enterprises (with zero employees) in year *t*+1

$N(1)_{t+1}$ population of active enterprises (with >zero employees) in year *t*+1

$N(0)_{t+2}$ population of active enterprises (with zero employees) in year *t*+2

$N(1)_{t+2}$ population of active enterprises (with >zero employees) in year *t*+2

A merge by identification code of the three years populations determines the following patterns:

t	***t*+1**	***t*+2**	***Output***
$N(1)_t$	$N(0)_{t+1}$	$N(0)_{t+2}$	**Deaths by Decline**
$N(1)_t$	$N(1)_{t+1}$	$N(0)_{t+2}$	-
$N(1)_t$	$N(0)_{t+1}$	$N(1)_{t+2}$	-
$N(1)_t$	$N(0)_{t+1}$	missing	**Deaths by Decline**
$RD(1)_t$	missing	missing	**Real Deaths with at least one employee**

In summary, the Employer Enterprise Deaths in year *t* (EED_t) are the Real Deaths with at least one employee $RD(1)_t$ and the active enterprises in year *t* with at least one employee ($N(1)_t$) which are in population $N(0)_{t+1}$ or $N(0)_{t+2}$ or both and which are neither in population $N(1)_{t+1}$ nor in $N(1)_{t+2}$ (Deaths by Decline). Results on split-offs should be available from the methodology used to identify enterprise births. Where possible, the information on new enterprises that were split-offs should be

used to identify original enterprises that moved below the one employee threshold because a new unit emerged from a split-off. These original enterprises should be removed from the population of Deaths by decline.

2.10 Employer Surviving Enterprise Definition

An employer enterprise born in year (*t*-1) or having survived to year (*t*-1) from a previous year is considered to have survived in year *t* if it is active in terms of turnover and/or employment in any part of year *t* and if it has at least one employee in year *t* (= survival without changes).

If the enterprise is not active in year *t* it has survived if its activity is taken over by a new enterprise (with at least one employee) set up specifically to take over the factors of production of that enterprise in year *t* (= survival by take-over).

2.11 High Growth Enterprise Definition

A variety of approaches can be considered as proving the basis for defining high-growth enterprises. According to Eurostat-OECD Manual on Business Demography, the enterprise's growth should be measured both in terms of employment (number of employees) and in terms of turnover.

2.11.1 Concepts

The definition of high-growth enterprises recommended is as follows:

All enterprises with average annualised growth in employees (or in turnover) greater than 20% per annum, over three year period and with 10 or more employees at the beginning of the observation period should be considered as high-growth enterprises.

Because the percentage of growth can be too high and to avoid excluding too many enterprises, another definition of Medium-Growth enterprises is proposed:

All enterprises with average annualised growth in employees (or in turnover) between 10% and 20% per annum, over three year period and with 10 or more employees at the beginning of the observation period should be considered as medium-growth enterprises.

From the High-Growth enterprises (HG) and the Medium-Growth enterprises (MG) are excluded the enterprises whose growth was due to demographic events such as mergers, take-overs and break-ups.

2.11.2 The identification process

When trying to identify high-growth enterprises, previously it is necessary to define the potential population of high-growth as all enterprises that were active in three consecutive years, excluding the enterprises born at the beginning of the observation period. It is necessary because the newly born enterprises with at least one employees in the year (*t*-3) could born at different periods in time during the year (*t*-3). Consequently their average turnover in the birth year is significantly lower than in following years simply because of the shorter average period of activity in the birth year. The same problem would not occur if only employment are measured, because it is measured as an annual average over the operating period and does not accumulate over the year. But, because high-growth enterprise are always identified from the same population, the real births in the year (*t*-3) are removed both to identify the high-growth measured in terms of employment and the high-growth measured in terms of turnover.

In practice:

Let N_t be the population of active enterprises in year *t*:

Step1: a merge by identification code of the population N_t and N_{t-1} to N_{t-3}

Step2: we exclude the real births in (*t-3*) from the Potential High-growth:

$$Potential\_HG_t = \left(N_{t-3} \cap N_{t-2} \cap N_{t-1} \cap N_t\right) \setminus RB_{t-3}$$

Step3: Size threshold of 10 or more employees at the beginning of the period (*t*-3).

$$Employees_{t-3} \geq 10$$

To identify the HG_t (or MG_t) enterprises only the enterprises with 10 or more employees are taken into account. This threshold of 10 employees is a convention and it is applied to avoid the introduction of biases that overstress the importance of small enterprises.

Step4: Growth threshold : 20% per annum for the HG

Growth threshold 10%-20% per annum for the MG

For example the HG are obtained applying to the population of reference (step 3) the following rules to employees or to turnover:

$$\sqrt[3]{\frac{employees_t}{employees_{t-3}}} - 1 \geq 0.2$$

or

$$\sqrt[3]{\frac{turnover_t}{turnover_{t-3}}} - 1 \geq 0.2$$

Step5: From HG_t (or MG_t) should be exclude the enterprises that grow because events of mergers or takeovers from units that cease or from units that transfer activity. By excluding such events we obtain the “pure” High Growth or "pure" Medium Growth enterprises.

2.12 Gazelle Enterprise Definition

2.12.1 Concepts

Gazelles are the subset of high-growth enterprises that are born (real births) at most five years ago.

We define Gazelles as “*all enterprises up to 5 years old with average annualised growth greater than 20 percent per annum, over a three year period.*

2.12.2 The identification process

In a given reference year *t*, gazelles may be in different cohorts of newly born enterprises RB_{t-3}, RB_{t-4} and RB_{t-5}, i.e., enterprises in their third, fourth or fifth year of survival. But, to be consistent with the definition of High-Growths, survivals from population RB_{t-3} are not considered.

In practice:

Let :

HG_t the High-Growth enterprises in the year *t*;

RB_{t-4} the Real Births in the year (*t*-4*)*

RB_{t-5} the Real Births in the year (*t*-5*)*

Merging by identification code the HG_t with RB_{t-4} and HG_t with RB_{t-5}, we obtain the Gazelles in year *t* ($Gazelles_t$).

$$Gazelles_t = (HG_t \cap RB_{t-4}) \cup (HG_t \cap RB_{t-5})$$

2.13 Entrepreneurship Indicators

Data on High-Growth (Medium-Growth) enterprises and Gazelles can be used to build indicators of entrepreneurship performance such as:

Rate of HG (MG): number of HG (MG) as percentage of total potential population of HG with at least 10 employees.

Rate of Gazelles among newly born enterprises: number of gazelles as a percentage of potential population of HG with 10 employees were born 4 or 5 years ago.

3. Design issues

4. Available software tools

5. Decision tree of methods

6. Glossary

For definitions of terms used in this module, please refer to the separate "Glossary" provided as part of the handbook.

7. References

Commission Regulation (EC) No 2700/98 of 17 December 1998 concerning the definitions of characteristics for structural business statistics.

Council Regulation (EEC) No 2186/93 establishing a common framework for business registers for statistical purposes and repealing.

EUROSTAT-OECD *Manual on Business Demography Statistics*. Eurostat Methodologies and working papers, 2007 Edition.

Garofalo, G. and Viviano, C. (1999), Continuity rules: re-delineation in the Italian context. In: *13th Internation Roundtable on Business Survey Frames*, Paris.

Garofalo, G. and Viviano, C. (2000), The problem of links between legal units: statistical techniques for the enterprise identification and the analysis of continuity. In: *Quaderni di Ricerca*, Istat N° 1/2000.

Garofalo, G., Paggiaro, A., Torelli, N., and Viviano, C. (2001), A record linkage procedure for the management and the analysis of the Italian statistical business register. In: *ICES-II, Proceedings of the Second International Conference on Establishment Surveys (Survey Methods for Businesses, Farms, and Institutions, June 17-21, 2000, Buffalo, New York)*, American Statistical Association, Alexandria, Virginia, 1612–1617.

OECD. *Entrepreneurship at a Glance 2012*. OECD Publishing http://dx.doi.org.

Interconnections with other modules

8. Related themes described in other modules

1. Statistical Registers and Frames – The Statistical Units and the Business Register
2. Micro-Fusion – Data Fusion at Micro Level
3. Micro-Fusion – Object Matching (Record Linkage)

9. Methods explicitly referred to in this module

1.

10. Mathematical techniques explicitly referred to in this module

1.

11. GSBPM phases explicitly referred to in this module

1.

12. Tools explicitly referred to in this module

1.

13. Process steps explicitly referred to in this module

1.

Administrative section

14. Module code

Dynamics of the Business Population-T-Business Demography

15. Version history

Version	Date	Description of changes	Author	Institute
0.1	28-02-2013	first version	Patrizia Cella	Istat
0.2	29-05-2013	revision	Patrizia Cella	Istat
0.3	06-08-2013	revision	Patrizia Cella	Istat
0.4	01-10-2013	revision	Patrizia Cella	Istat
0.4.1	17-10-2013	preliminary release		
1.0	26-03-2014	final version within the Memobust project		

16. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	21-3-2014 17:41