**21th International Roundtable on Business Survey Frames**

**Paris – November 24 – 27, 2008**

**Session Nº 6A**
**Paper N° 1**
**H.J. Konen, Statistics Netherlands**

# Methodology to continue the Dutch Harmonised Data Collection

**Abstract**

In the Harmonised Data Collection (HDC) Business Demography Statistics are compiled. The methodology for this data collection is described in the Eurostat-OECD manual on Business Demography Statistics. In the manual it is recommended to compare two consecutive populations of active enterprises. The enterprise-units in these populations are subsets from the corresponding Business Registers. Since July 2006 Statistics Netherlands is using a new Business Register (BR) to survey the production process of enterprises. In this new Business Register 'new' data-sources have been used for legal units and some of their characteristics. In addition, the algorithm to derive enterprise units has also been changed. By this conversion, Dutch birth- and death rates will be distorted whenever these indicators are derived without restrictions. To solve this distortion, Statistics Netherlands developed a method so the time series of the Dutch Business Demography indicators can be continued. The transition of the BR however will still cause distortions at some level in the time series of the Dutch birth- and death rates.

## 1. Introduction

The Dutch Business Register (BR) is an internal source with information available on several important characteristics of an enterprise: economic activity, legal form, name, address and links between enterprise- and legal units. Its records make up the population of about 1,000,000 statistical enterprise units each year. The main objective of the BR is to serve as a coordinated sample frame to observe the production process of enterprises. The quality of this register depends strongly on used legal sources and register maintenance procedures. Within the Business Register Department of Statistics Netherlands algorithms are developed to determine enterprise units and several procedures are carried out to update the characteristics of enterprise units. Each month the Dutch BR is updated in order to obtain a monthly sample frame to observe production factors in the Short Term Statistics (STS). For the selection of the legal units, each month the most recent information on employment and turnover available is used to update the business register. This information however often is not up to date or even not available at the time yet. Hence it is practically inevitable that some dormant enterprises are still registered in the BR of that month. In order to obtain a population of active enterprises for one complete reference year, first the sample frames of the 12 months separately are joined. Secondly each enterprise unit is updated with information about employment and turnover that refers to the reference year to be sure the corresponding enterprise was active. For the employment, the number of employees for each unit is based on the Statistics on Labour and Wages. Turnover-data are originated from the Tax-office.

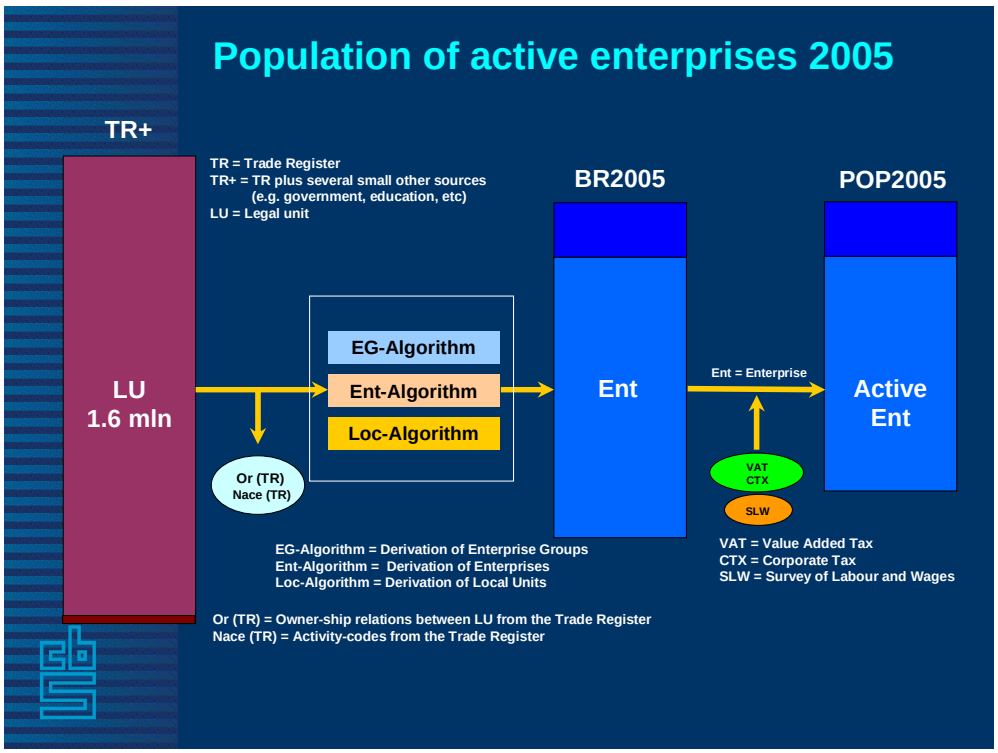
Since July 2006 Statistics Netherlands is using a new Business Register (BR) to observe production factors of enterprises. The main source of legal units of the old BR - approximately 96% - was originated from the Trade Register of the Chamber of Commerce. The remaining 4% of the legal units was originated from the administration of social security. Legal entities which were not obliged to register in the TR, were often only registered in the tax administration of the tax-office. This means that some companies were not covered in the old BR. From 2010 onwards a new Trade Register (TR) will be used by Statistics Netherland as the main source for legal entities in the Dutch BR. In this new TR all legal units have to register for legal reasons. In advance, the Ministry of Finance and the Chamber of Commerce joined forces and created a new register for legal units (together with Statistics Netherlands). This is the Single Business Register (SBR). One main advantage of this cooperation is that also the links between legal units and fiscal identifiers of the tax office are improved.

Statistics Netherlands decided to develop a new BR that is based on the Single Business Register. This means that organisations that in former days were only known at the tax office nowadays are covered in the new BR. With the construction of the new Business Register, more information from the tax office is used to determine the activity status and relations between units. As a consequence of this many new enterprise-id's are created and a lot of old enterprise-id's have been dropped from the BR.

## 2. Distortion in Dutch Business Demography Statistics

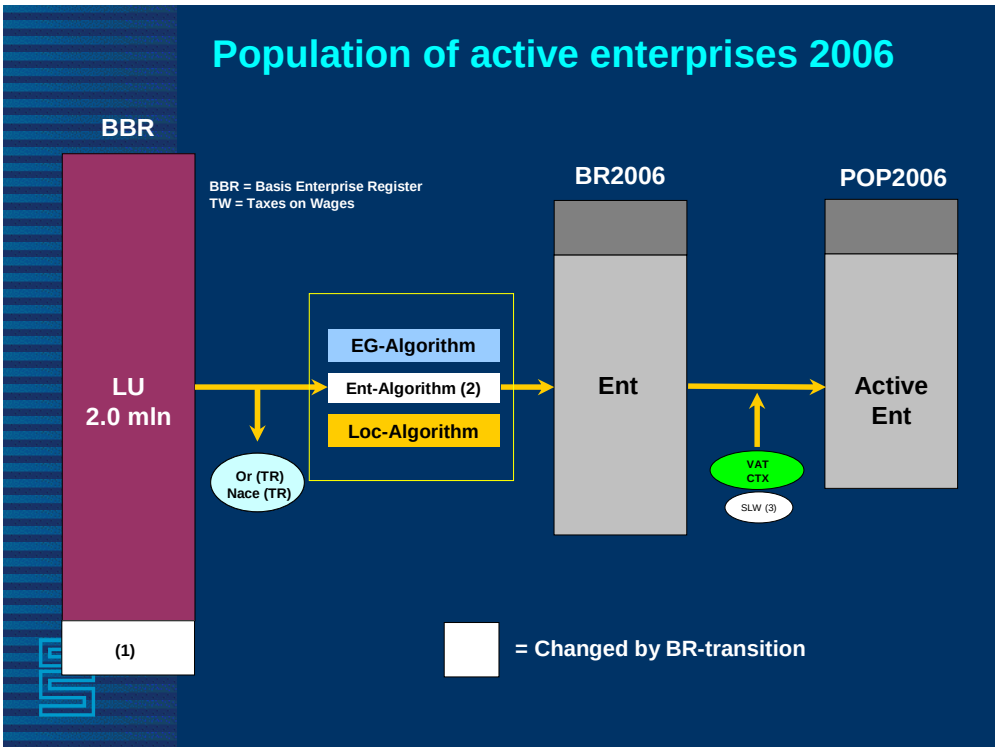
In the Harmonised Data Collection 2008 (HDC) the populations of active enterprises in reference years 2005 and 2006 are compared. First both populations are described before zooming in on the most important changes that cause the distortions in the indicators produced in the HDC.

*Population 2005*
The most important source of legal units in the BR 2005 is the Trade Register of the Chambers of Commerce (TR). In this TR only legal units that are obliged to register are updated. Several legal units are not linked to fiscal units of the tax-office. This means that not all available information about employment or turnover from the tax-office could be used to determine the population of active enterprises.
The size class variable in the population is based on the number of employees in the Survey of Labour and Wages (SLW). The base of the SLW is a sample from the BR from approximately 70.000 enterprises out of approximately 380.000 employer-enterprises (according the BR). Note that all enterprises with 10 or more employees are included in the survey. After weighting the response of the SLW-sample the number of jobs is estimated. In a second step a mathematical model is used to estimate the number of employees for the other 310.000 units with less than 10 employees.




*Population 2006*
One of the reasons to develop a new BR was to reduce the statistical administrative burden for small and medium-sized enterprises. Therefore administrative sources from the government will be used more to produce Business Statistics. Related to this matter is also the use of a new Trade Register from 2010 onwards, as mentioned before. Before this period however, the Chamber of Commerce, the Ministry of Finance and Statistics Netherlands cooperate in the Single Business Register (SBR). In the BR of 2006 until 2010 the SBR is used to register new enterprises, to remove enterprises that have terminated economic activities or to update any changes in the characteristics of an enterprise. The

introduction of a new BR to compile Business Statistics means also that all SBS-statistics are more or less based on the enterprise-units in the new BR.




To compile birth- and death rates in the HDC of 2008 the active populations of 2005 and 2006 are compared. The number of newly entered enterprises is corrected by all 'non-birth' events in order to obtain the number of real births in the population of 2006. The number of removed enterprises is corrected by all 'non-death' events in order to obtain the number of real deaths in the population of 2005. Because many new enterprise-id's are created and a lot of old enterprise-id's have been removed from the BR, the entry rate and exit rate are disturbed. In case we just subtract the consecutive active populations of 2005 and 2006 we would obtain an enterprise entry rate of 18% in 2006 and an enterprise exit rate of 17% from 2005 (for enterprises in Nace sections C,D,E,F,G,H,J,K). These rates are normally about 14-13%. We assume that there are three major changes that cause these distortions.

- *Transition in primary source TR - SBR*
  The use of the SBR has improved the links between LU and Fiscal ID's. Hence more information on the activity status of units can be used from the tax office.
  The use of the Single Business Register (SBR) instead of the Trade-register (TR) as a primary source for Legal Units (LU) introduces new entries of enterprise-units in the new BR. Many of these enterprises entered in
  Nace division 01 Agriculture
  Nace division 65 Financial intermediation
  Nace division 74 Other business activities
  Nace division 85 Health and social work
  Nace division 92 Recreational, cultural and sporting activities

- *Introduction of the 'Joining Algorithm'*
  In the new BR the automatic algorithm to determine enterprises has changed, because of the introduction of the so called 'Joining-Algorithm'. The main principle of this algorithm is to join LU's together as one enterprise (within one Enterprise Group). The idea is to join LU's in case of
  - Similar economic activity or
  - Group registration for VAT

Note that the Nace code and the Legal form are determined on enterprise level. Those characteristics are similar to those of the main legal unit. In case an enterprise consists of more than one legal unit, that legal unit which has the largest contribution in terms of employment within the enterprise-unit, is chosen as main legal unit. This is one of the reasons why the transition can only be a part of the solution and will still cause distortions at some level in the time series of the Dutch birth- and death rates.

- *Changes in the Survey of Labour and Wages*
  In the populations of reference years before 2006, employment data for employers with less than 10 employees were estimated by weighting sampled data from the Survey of Labour and Wages. Since reference year 2006, employment data in the not based on estimates anymore, but on full social security registration at the tax office on a monthly base. To determine the number of employees in an individual enterprise in the active population of 2006, we used the annual average of the employment over the operating period (i.e. periods with employees) and we rounded this to the nearest whole number.

## 3. Method to compile the HDC

If we want to apply the rules set out in the Eurostat-OECD manual to compile the HDC, then from the last paragraph it is clear that in summary, we have to make strong assumptions in order to develop implementation-methods to:
- estimate the number of new 'Tax-only'-enterprises in 2006;
- describe the relation between old and new enterprise-id's;
- make a correction for non-active enterprises in the population of 2005.

- *Method to estimate the number of new 'Tax-only'-enterprises in 2006*
  In the BR 2006 a many new enterprise units have entered the population, since the SBR also contains enterprises, that were absent in the old BR, because they were only registered as a company at the tax office but not in the TR. It's known that only a small part of them are actually a new born enterprise. Most of those enterprises also existed in the year 2005, but were not known as such in our old BR. In the BR of 2006 we made a strong assumption to distinguish between new and existing so called 'Tax-only'-enterprises. We used the percentage of newly registered 'Tax-only'-entries in 2007 as an approximation for the part of real new 'Tax-only'-enterprises in the active population of reference year 2006.

- *Method to describe the relation between old and new enterprise-id's*
  During the conversion from the old BR to the new BR, the identification number of the enterprise was kept unchanged for enterprises with a similar legal structure. A new enterprise-id was introduced when a new legal structure for an enterprise was created. Other difficulties were introduced by a different spelling of company names, altered location codes or changes in activity-codes even for identical enterprise-id's in both BR's. This causes complications when matching steps are about to be applied to determine the 'one by one takeovers' (which are not defined as real births). In order to apply the methodology as described in the Eurostat-OECD manual, we create a shadow-sub-population 2006 for this part.

  Because of the introduction of the 'Joining-Algorithm', many former enterprise units in the old BR up to June 2006 have been merged or occasionally splitted in the new BR of 2006. Since the identification number of a legal unit did not change during the transition-process in July 2006, we could link the main legal unit of an enterprise in the old BR up to 2006 to exact the same legal unit of an enterprise in the new BR. In this way it was possible to obtain a transition-matrix between the old and the new identification numbers of the enterprises in both populations.

  What still is missing in the transition-matrix described above, are the real new enterprise units that should have been entered the old BR of 2006. Those units namely entered the new BR of 2006 and could not enter the old BR up to 2006 because the conversion took place in the middle of the year. For this we assume that all unlinked enterprises are new entries and would have been linked one on one if the conversion would have been at the end of the year.

  Hence it is possible to transform each sub-population of the new BR into a sub-population of the old BR and vice versa. Using this principle, it is possible to create a of 'Shadow'-subpopulation of

active enterprises in 2006 which can be compared with the population of active enterprises of 2005, in order to obtain entries and exits and carry out matching steps as mentioned in the manual.

- *Method to make a correction for non-active enterprises in the population of 2005*
  In the 2005-population, some enterprises were labelled as "being active" based on estimates (by weighting sampled employment data of the SLW). Nowadays however, 2005-data on employers is also available based on the registration from the tax office. We compared both 2005-employer-enterprise populations and were able to distinguish those enterprises that were falsely declared active in the population of enterprises of 2005. Corrections were made by subtracting this so called 'falsely' sub-population from the set of enterprises that was removed from the population in 2005.

## 4. Results and conclusion

In the table below the entry- and exit rate for the total of enterprises in Nace sections C,D,E,F,G,H,J,K are listed before and after the application of the suggested method. It is obvious that some rates could derogate more for detailed Nace-, Size class- or Legal Form-aggregates. Note that the entry rate (exit rate) is not the same as birth rate (death rate), since that indicator should be corrected first for the number of enterprises that fulfil the conditions of non birth events (non-death events). In the end we conclude that we can expect that the suggested method will improve also the time series for birth and death rates of the Dutch Business Demography Statistics.

| **Nace sections C,D,E,F,G,H,I,J,K** | **2004** | **2005** | **2006** |
|---|---|---|---|
| **Before** | | | |
| Entry rate | 12,8 % | 13,7 % | **17,7 %** |
| Exit rate | 9,8 % | **16,6 %** | **-** |
| **After** | | | |
| Entry rate | 12,8 % | 13,7 % | **14,7 %** |
| Exit rate | 9,8 % | **11,5 %** | **-** |

In the annex a list of formulas is stipulated to determine the size of several base populations determined in the HDC 2008. The number of enterprises in the derived populations forms the numirators and the denominators for Business Demography indicators.

**Annex: Formulas to compile the Harmonised Data collection**

In the HDC 2008 particularly the number of enterprises in several sets of enterprises should be estimated:
- Number of enterprises in the population of active enterprises;
- Number of enterprises in the population of enterprise deaths;
- Number of enterprises in the population of enterprise births;
- Number of enterprises in the populations of enterprise survivals (1,2,3,4 and 5 year survivals).

We note in advance that in the definitions below, a subscripted h (as in $N_h(y)$) always refers to a sub-population of the new BR and a subscripted o (as in $N_o(y)$) to a sub population of the old BR. The character (y) within brackets always refers to the reference year.

| | |
|---|---|
| $N_h(y)$ | Set of active enterprises as a subset of the new BR |
| $N_o(y)$ | Set of active enterprises as a subset of the old BR |
| $R_h(y)$ | Set of newly born enterprises as a subset of the new BR |
| $R_o(y)$ | Set of newly born enterprises as a subset of the old BR |
| $R_h^{Tax}(y)$ | Set of newly born enterprises as a subset of the new BR, which contains of enterprises only registered by the tax-office (New Tax-only-enterprises). |
| $F_o(y)$ | Set of falsely declared active enterprises as a subset of the old BR |
| $M_R(y)$ | Set of enterprises that fulfil the conditions of a non birth event as a subset of all newly listed enterprises in the BR |
| $M_D(y)$ | Set of enterprise that fulfil the conditions of a non death event as a subset of all removed enterprises in the BR |
| $T_o(X_h(y))$ | Set of enterprises listed in the old BR created by applying the transition matrix $T_o$ on a subset of enterprises $X_h(y) \subset N_h(y)$ |
| $T_h(Y_o(y))$ | Set of enterprises listed in the new BR created by applying the transition matrix $T_h$ on a subset $Y_o(y) \subset N_o(y)$ |

**Step-by-step plan for each population:**

Note that the number of enterprises in a population X is quoted as $|X|$

a. Number of active enterprises $N_{06}$ ( SBS variable 11910)

$$N_{06} = |N_h(2006)|$$

b. Number of deaths $D_{05}$ (SBS variable 11930)

(1) $N_o(2006) = T_o(N_h(2006))$

(2) $D_{05} = \left|\left(N_o(2005) \setminus N_o(2006)\right) \setminus F_o(2005)\right| - \left|M_D(2005)\right|$

c. Number of real births $R_{06}$ (SBS variable 11920)

(1) $N_o(2006) = T_o(N_h(2006))$

(2) $R_o(2006) = N_o(2006) \setminus N_o(2005)$

(3) $R_{06} = \left|T_h(R_o(2006)) \cup R_h^{Tax}(2006)\right| - \left|M_R(2006)\right|$

d. Number of 1, 2, 3, 4, and 5 year survivals:
$S_1(R_{05},N_{06})$, $S_2(R_{04},N_{06})$, $S_3(R_{03},N_{06})$, $S_4(R_{02},N_{06})$, $S_5(R_{01},N_{06})$
(SBS variables 11941, 11942, 11943, 11944, 11945)

$S_1(R_{05},N_{06}) = \left|R_o(2005) \cap N_o(2006)\right|$

$S_2(R_{04},N_{06}) = \left|R_o(2004) \cap N_o(2005) \cap N_o(2006)\right|$

$S_3(R_{03},N_{06}) = \left|R_o(2003) \cap N_o(2004) \cap N_o(2005) \cap N_o(2006)\right|$

$S_4(R_{02},N_{06}) = \left|R_o(2002) \cap N_o(2003) \cap N_o(2004) \cap N_o(2005) \cap N_o(2006)\right|$

$S_5(R_{01},N_{06}) = \left|R_o(2001) \cap N_o(2002) \cap N_o(2003) \cap N_o(2004) \cap N_o(2005) \cap N_o(2006)\right|$

e. Example to determine the number of survivals $S_2(R_{05},N_{07})$ in the HDC 2009

$S_2(R_{05},N_{07}) = \left|U_1^o(2005,2006) \cap T_o\left(U_2^h(2006,2007)\right)\right|$

where,

(1) $U_1^o(2005,2006) = R_o(2005) \cap N_o(2006)$

(2) $U_2^h(2006,2007) = T_h\left(U_1^o(2005,2006)\right) \cap N_h(2007)$