Linking multiple sources to produce economic statistics at Statistics Netherlands

Adelaide Ariel, Arnout van Delden, Ivo Beuken, and Magda Slootbeek-van Laar

Summary

Statistics Netherlands (SN) is working on a redesign programme of the chain of economic statistics. The programme aims to increase the output quality, reduce the administrative burden and reduce staff costs. One of the means to contribute to these goals is to use administrative data as much as possible. Primary data are used only for the larger units and when administrative data are not available – leading to multiple input sources.
SN links the input data at the *micro* level to statistical units to ensure that outcomes cover the target population, to facilitate data editing and to enable output by different classifications. A major challenge is that there is a wide variety of administrative unit types, most of which differ from the statistical ones. Until recently, only a selective and incomplete portion of the input units could be linked to the target statistical units.
The objective of the current paper is to describe our improved methodology to make the transition from input unit types, populations and classifications to the target ones; and the challenges that we are still facing. We give examples for two Dutch tax registers: Value Added Tax (VAT) and Income Tax. Specifically we

- address the new methodology to derive the statistical units. We now use more information on ownership-relations between administrative units;
- explain the linkage of administrative to statistical unit types. Within each enterprise group we search for all the underlying administrative units;
- show how we deal with unequal coverage among the two registers. For example, self-employed individuals are exempted from Income Tax and not from VAT, but there is no straightforward way to identify them;
- show the portion and causes of non-linked units and how we deal with them in the estimation process.

1. Introduction

Statistics Netherlands (SN) is working on a redesign programme of the chain of economic statistics (HEcS, in the Dutch acronym). The primary goal of this programme is to improve the quality of statistical outcomes while at the same time reduce both the administrative burden and cost. Central to this programme is that after the redesign the various processes for economic statistics are not only standardized in terms of the processing steps and their order, but also in terms of the population of business units and their classification. Another important issue in this programme is an extensive use of administrative data. In addition to reducing survey burden, administrative data offer a great range of advantages, such as a possibility to produce new statistics, to assist in data editing and to control plausibility of the data from different sources (Wallgren and Wallgren, 2007). At SN we will use administrative data as much as possible, and use primary data only when administrative data are not available or not suitable. A more comprehensive description about this programme has been detailed in Braaksma (2007) and van Delden and Aelen (2008).

There are two main challenges in the application of register-based statistics. First, in the Netherlands administrative data have a wide variety of unit types, most of which are different from the statistical unit types. Consequently, administrative units need to be transformed into statistical units before they are ready for use. This transformation process is generally known as a linkage process. Data loss and errors due to linking process can occur, resulting in uncertainty in the estimates.

Although not easy, one can probe the cause of not linking by means of, for example, examining the units not linked. When two or more administrative sources are involved, such task is burdensome. Also, the coverage of administrative sources may differ from the target population. Hence, it is important to maximize the number of records that can be linked. Another challenge is related to multiple input sources, namely when both administrative and primary source data are used. In this case, preventing a potential discrepancy in terms of the units covered and the classifications used is desirable.

In this paper we will share our experiences in using multiple sources to produce economic statistics. Two important Dutch registers of the tax office, value added tax (VAT) and Income Tax, will be used as an example on how we combine data from different administrative data sources. To illustrate the application of our methodology, we will provide a case from a project called Direct Estimation of the Totals (DRT), which is a part of the HEcS programme. The DRT project aims to use VAT data as the main source for monthly, quarterly and yearly turnover estimates. These estimates have to be consistent and released on time, which puts pressure on the data editing and estimation process. In addition to the VAT data, the DRT project plans to use data of the income level from the Income Tax register to support the detection of errors in turnover values of the VAT data. Therefore, data from Income Tax need to be somehow linked with the VAT data. In addition, SN examines the use of Income Tax data for employment statistics.

The remaining of this paper is outlined as follows. Section 2 describes the VAT and income tax registers and units. Section 3 address the new methodology to derive the statistical units based on legal units and how we link administrative units into statistical units. Section 4 illustrates the method. In addition we explain how we deal with unequal coverage and with non-linked units in the estimation process.

2. VAT and Income tax registers and units

The Dutch tax authority demands legal units in the Netherlands to report their VAT on a monthly, quarterly or yearly basis, depending in most cases on their yearly VAT levy. Depending on the kind of economic activity, for small and medium enterprises (SME, < 50 employees) monthly reporters may contribute up to 70% of their total turnover (de Wolf and van Bemmel 2007). Legal units can also join into so called “fiscal units” (within restrictions) in order to declare tax. VAT units have an n:m relationship to the statistical units (enterprises). Before the redesign, only 60% of the VAT-units within the SME can be linked to enterprises.

To simplify the income tax declaration, the Dutch government announced a new regulation, which requires legal units to report their income tax directly to the tax office. In the past, Income tax was declared to another governmental organisation. Now it seems easier for SN to combine both VAT and Income tax data for statistical purposes. In practice, the tax authority does not oblige legal units to

use the same administrative identification for their VAT and income tax. A company can use a legal unit to declare income tax that is completely separate from the units that declare VAT. That makes it complicated for SN to correctly link Income and VAT data to statistical units.

3. Business register for multiple sources

To use data from different registers, it is necessary to have a business register that describes the underlying relationships among legal units that are owned by the same enterprise group unit. Currently, SN redesigns its business register, and an important project called Unit Base (EHB in the Dutch acronym) has been set up for this purpose. The basic idea is to identify any existing legal units that belong to the same enterprise group unit and describe the hierarchical links (ownership relationships) between them. In addition to this, *groups of legal units* that declare VAT together, the 'fiscal units' should belong to the same enterprise group, because those fiscal units have > 50% ownership relations between them. In the EHB project, we use a source from the tax office to trace these ownership relationships. EHB leads to an improvement of the operationalisation of the enterprise group unit compared to the current situation where we only use 100% ownership relations. EHB also tries to include all legal units within an enterprise group that declare income tax. A detailed description is given in Vassen-Otten and Beuken (2009).

Figure 1. Future General Business Register contains all administrative identifications (id) per enterprise group (EG) unit.

Vat register

Turn over	Period	Id_VAT unit
100	June 2009	1
200	June 2009	2
205	Apr-June 2009	4
500	Apr-June 2009	5
70	June 2009	6
490	Apr-June 2009	7

linkage ⟷

General Business Register

Administrative part			Statistical Part		
Id_VAT unit	Id_Income unit	Id_legal unit	Id_EG unit	Id-Enterprise unit	Top2000 unit
1	-	L900	EG1	E100	Yes
1	-	L902	EG1	E100	Yes
2	-	L903	EG1	E101	Yes
-	3	L903	EG1	E102	Yes
4	4	L907	EG2	E103	No
5		L908	EG2	E103	No
5		L910	EG2	E103	No
5		L911	EG2	E103	No
6	6	L912	EG3	E104	No
7	7	L913	EG3	E104	No

EHB improves the linkage of administrative data to *enterprise groups*. However, statistics dealing with production (turnover, production revenue and costs etc.) need to be classified by industry (kind of economic activity) based on *enterprise* unit. An enterprise group may consist of one or more enterprises. The enterprise group is the actor of the financial processes; the enterprise is a smaller entity which is the actor of the production process. In practice only the larger enterprise groups consist of two or more enterprises; for 'the bakery around the corner' enterprise group coincides with one

enterprise. At SN approximately the largest 2000 enterprise groups are divided into two or more enterprises. For the rest of enterprises, we will *approximate* the enterprise (as a concept) by the enterprise group, so in the business register those smaller enterprise groups have only one enterprise. In order to achieve a co-ordinated classification by industry and size class between administrative and survey units we classify each enterprise and link both administrative and survey units to them.

The new design brings a significant improvement, as almost all of register units can be identified and linked to the statistiscal units in our business register. Previously, for only about 60% VAT units of SME could be linked to our enterprises (van Delden and Aelen, 2008). Using the new methodology, only about 5% of the units cannot be linked; most of them are due to outdated information from the tax office. Moreover, adding and combining data from different registers becomes more straightforward. We illustrate in the next paragraph how we combine VAT and Income data using this new system.

4. Linking multiple sources: an example

We start the linking process by choosing the target population of enterprises. In this example, we select all (SME) EG of the industries that are relevant for the short-term statistics (STS). Next, data from the VAT and the Income Tax register are added to these enterprises by linking their administrative identification code (see Figure 1). Here, in addition to this code, the data from the Income Tax register also have an enterprise unit code since they have been linked to (former) General Business Register and corrected by Employment Statistics Department at SN.

Unequal coverage. As mentioned previously, each register may have a different coverage. Any EG that belongs to the STS population is obligated to report their VAT, thus almost all of them can be found in VAT register. Income Tax register, on the other hand, is meant generally only for units with employees, and one of the purposes of this tax is to provide social security for employees. Linking with Income Tax will divide the EGs into EGs found and EGs *not* found in this register. We expect self-employed to be mostly in the last group. Since there is no straightforward way to identify whether a self-employed individual does have employees (other than himself), we choose to isolate all of EGs having only one legal unit and coded as self-employed. Self-employed found in the Income Tax register turn out to have a much larger turnover, compared to self-employed not found in this register. Furthermore, we also identify EGs not coded as self-employed but not found in Income Tax. For them, we reapproach the linking with Income Tax via their enterprise unit code. In Table 1 we summarize the average value of VAT and Income Tax per EG for the second quarter of 2007 as an example; similar patterns also appear in the other quarters. This table shows that the average quarterly turnover of enterprises that are not found in Income Tax register is structurally smaller than of those for which income tax data were found, irrespective of whether they are self-employed individuals or not.

Table 1. The average value per enterprise of quarterly turnover (VAT) and amount of income (Income tax) in the second quarter of 2007

Enterprises with VAT	%	Quarterly Turn over (EUR)	Quarterly Income (EUR)
1. Found in Income Tax			
(a) via fiscalenumber	26.0	707 359	140 715
(b) via enterprise unit-code	0.1	237 605	78 649
(c) self-employed	9.1	85 611	12 465
2. Not found in Income Tax			
(a) self-employed	37.2	17 515	
(b) others	27.6	77 805	

Estimation of totals. Turnover data from the VAT is used in the DRT project to estimate levels and changes of turnover levels classified by industry to produce STS output. Aggregating only over the *linked* turnover of all enterprises of the target population would underestimate the turnover sum. For the DRT project the income tax data serve only as reference material for editing turnover data. That means that non-linked income data are just not available as reference material. However, it is important to discover the reasons for EGs not linked with Income Tax data, particularly for those not coded as self-employed. As shown in Table 1, although not many, there is a number of EGs with a significantly large turnover that can be identified in Income Tax only by using their enterprise unit-code. We intend to list them and try to resolve the linkage manually, e.g. by using information from the tax office or from the legal units.

References

Braaksma, B. (2007). Redesign of the chain of economic statistics in the Netherlands. Seminar on registers in statistics – methodology and quality. May 21-23, 2007, Helsinki.

De Wolf, P.P. and Van Bemmel, K.J.H. (2007). Methodology for STS statistics based on VAT data. Report of Statistics Netherlands. *(in Dutch*)

Van Delden, A. and Aelen, F. (2008). Redesigning the chain of economic statistics at Statistics Netherlands: STS-statistics as an example. IAOS conference 'Reshaping official statistics' October 14-16, 2008, Shanghai.

Vassen-Otten, A. and Beuken, I. (2009). New enterprise group delineation using fiscal information. Concept paper.

Wallgren, A., and Wallgren, B. (2007). Register-based Statistics. Administrative data for statistical purposes. John Wiley and Sons, Ltd.