Catalogue no. 11-522-XIE

Statistics Canada International Symposium Series - Proceedings

Symposium 2005 : Methodological Challenges for Future Information needs

2005

Statistics Canada **Statistique Canada**

Canada

Proceedings of Statistics Canada Symposium 2005
Methodological Challenges for Future Information Needs

USE OF ADMINISTRATIVE DATA FOR STRUCTURAL BUSINESS STATISTICS

K.J.H. van Bemmel[1]

ABSTRACT

The subject of this paper is the use of administrative data like tax data and social security data for structural business statistics. At Statistics Netherlands value added tax and incorporated tax data will be used for small-sized businesses for many economical sectors in addition to survey data. In this paper also the newly developed statistics on general practitioners is discussed. For that statistics non-incorporated tax data is also used and the output is based on administrative data only. Apart from the advantages also the limitations of using administrative data are discussed. The overall conclusion is that the advantages outweigh the disadvantages.

KEY WORDS: Administrative data, tax data, social security data, structural business statistics

1. INTRODUCTION

Many statistical offices try to reduce the response burden by using data gathered for administrative purposes rather than statistical purposes. These administrative data are going to replace data collected by traditional surveys more and more. In the Netherlands there is a new law that states that Statistics Netherlands can only maintain a survey if the data can not be obtained via administrative sources from governmental or government related organisations. This forces Statistics Netherlands to obtain as much administrative data as possible.

Apart form reducing the response burden there are also other reasons to make more use of administrative data. First of all the traditional surveys often have low response rates and also the quality of the response is sometimes problematic, especially for small-sized businesses. Secondly, the administrative data often contain interesting information that was not asked for in the surveys.

The topic of this paper is the use of administrative data for structural business statistics. Tax data obtained from the Inland Revenue Service is used to determine the revenues and expenses of businesses, while social security sources are used to obtain information on jobs in the businesses. Apart from the advantages also the limitations of using administrative data will be discussed.

In the past years value added tax has already been used at Statistics Netherlands for structural business statistics while incorporated tax data has just started to be an addition to surveys for small-sized businesses on a large scale. For a newly developed statistics on general practitioners non-incorporated tax data and social security sources are used.

The outline of the paper is as follows. Section 2 and 3 consist of a general description of the main administrative data that are used for structural business statistics, namely tax data and data from social security sources. The use of value added tax and incorporated tax data at Statistics Netherlands will be discussed in Section 2. Section 3 describes the data on jobs available at Statistics Netherlands. In Section 4 the method of the newly developed statistics on general practitioners is explained. In that method the administrative data are not used in conjunction with survey data, but the output is based on administrative data only. An overall conclusion on using administrative data for structural business statistics is presented in Section 5.

[1] K.J.H. van Bemmel, Statistics Netherlands, P.O.Box 4000, 2270 JM Voorburg, The Netherlands

2. TAX DATA

2.1 Value added tax

Value added tax has already been used for several years for structural business statistics at Statistics Netherlands. In the past it was not used as a direct replacement of the turnover in surveys, but was used in the determination of weights to get aggregated results. Just recently Statistics Netherlands has started to use value added tax as a direct replacement for survey turnover for small-sized businesses. The structural business statistics of 2004 will be the first ones for which this new method will be applied. The replacements will only be adopted for those economical sectors for which there was a good agreement between survey and value added tax turnover in previous years.

2.2 Incorporated tax data

Statistics Netherlands also disposes of the tax data of all incorporated businesses. Besides value added tax also these tax data are used for the small-sized businesses in the new method for structural business statistics for some economical sectors. The most relevant tax data are the revenues and expenses listed in the profit and loss accounts.

Not all variables that are asked for in the traditional surveys are contained in the tax data. Therefore surveys are still needed if the same output has to be generated. However, by using relationships between tax data and survey data a significant reduction of the number of businesses that are surveyed is still achieved.

The method is as follows. Some key variables X like total expenses and purchase value are contained in the incorporated tax data. Because the amount of usable incorporated tax data is much less than the amount of usable value added tax data, the aggregated results for these variables X are not determined directly by incorporated tax data only. Instead a relationship between these variables X and the value added tax turnover O^{VAT} is determined on basis of the available tax data: $X_t = R_t \bullet O_t^{VAT}$. The aggregated results for X are then calculated using this relationship R and the aggregated value added tax turnover. If the relationship R_t in year t is not available, the relationship R_{t-1} in year *t*-1 can be used.

Other variables Y are not contained in the incorporated tax data. For these variables a relationship with the key variables X is used: $Y_t = S_t \bullet X_t$ or $Y_t = S_{t-1} \bullet X_t$, with S based on survey data. Because of correlation between Y and X, the number of surveyed units can be lower compared to a method with surveys only. Moreover the relationship between Y and X is often reasonably constant over time, which makes it possible not to survey the businesses every year. That is why sometimes S_{t-1} is used instead of S_t.

Some variables like the number of stores are not included in the approach. This is because the correlation with the key variables X is too small. In these cases the variables still have to be obtained in the old fashion, or have to be estimated by other means, for example by using the data of the surveyed medium-sized businesses.

The method is restricted to small-sized businesses. Larger businesses often have more complex structures with many administrative entities making the use of administrative sources more problematic. In total the number of surveyed businesses is reduced by approximately 20 %. For some economical sectors the percentages are much higher. The number of surveyed businesses for retail for example is reduced by approximately 50 % and by approximately 70 % if only small-sized businesses are concerned.

In the new method only tax data of incorporated businesses is used. Therefore one should correct R_t for the fact that part of the population consists of non-incorporated businesses. This correction is determined on the basis of old survey data of incorporated and non-incorporated businesses.

2.3 Non-incorporated tax data

The new method only uses incorporated tax data. In principle the non-incorporated tax data could be used in the same manner. The problem is that the non-incorporated tax data available at Statistics Netherlands are only a

small sample of all non-incorporated tax data. Furthermore the non-incorporated tax data are not as recent as the incorporated tax data. The reason for these problems is that up to a year ago the non-incorporated tax data, unlike the incorporated tax data, were not standardised and electronic. However starting with the tax forms of 2004, also the non-incorporated tax data are electronic and standardised. Statistics Netherlands will also get these tax data, making an extension of the method to non-incorporated tax data possible.

3. DATA ON JOBS

3.1 Job file

The tax data are not the only administrative data that can be used for structural business statistics. Also the data from social security sources are useful. For all employees part of the wage is handed over to social security institutions. To facilitate this, information on jobs is stored by the social security institutions, like the starting date and final date of the jobs and the wage.

Other data on jobs is obtained from the Inland Revenue Service and from a survey. The survey is the only source with information on the (contractual) amount of working hours. In the near future this information will also be contained in the administrative data of the social security and the survey will be abolished.

At Statistics Netherlands the data on jobs from social security, Inland Revenue Service and the survey are integrated into one file. This file is called the job file. In the job file the key consists of an identification number of the employee and an identification number of the employer. Statistical information like age and sex can be added via the Population Register. Many variables are available for all jobs like wage, starting date and final date of the job, and the number of days worked during the year. Since some employees only work a couple of hours during a day, the number of days worked during a year does not provide directly the number of hours worked during a year. However, for a sample of jobs information on the contractual amount of hours worked is obtained by the survey. For small-sized businesses this sample is quite small, but the precision of statistical results on the number of hours worked can be improved by making use of regression with integrally available variables like wage and number of days worked.

3.2 Job file compared to survey

A difference with traditional surveys is that in the job file there is only information on employees that are on the payroll of the businesses. Information on hired personnel and agency staff is not present in the job file, but is asked for in the traditional surveys. One can conclude that for the most important part of the personnel, the employees on the payroll, the job file provides much more information than the surveys. However, for a small part, the non-payroll employees, the job file has no information.

The lack of timeliness of the job file is a problem for obtaining timely output. Instead the number of employees can be estimated using the aggregated results of the survey on jobs only. In the near future the administrative social security sources will have fewer problems with timeliness, because the administrative processes in the social security are being simplified.

4. GENERAL PRACTITIONERS

4.1 Introduction

In the new method for some structural business statistics that was shortly described in chapters 2 and 3, administrative sources are used in conjunction with a survey. In the newly developed statistics on general practitioners, administrative sources are not used in conjunction with a survey, but the output is based on administrative data only. The method is described for general practitioners, but is applicable to many self-employed professions. In the method non-incorporated tax data and the job file are used. The lack of timeliness of these data is accepted for the moment, since in the near future the data will be more recent.

4.2 Characteristics

Most general practitioners are self-employed and are organized in small-sized businesses. The far majority of the businesses are non-incorporated businesses, mainly one-man businesses and partnerships. Furthermore

general practitioner services are exempted from value added tax. A surgeon apothecary has to hand over value added tax only for the sale of medicine.

4.3 Approach

Until last year Statistics Netherlands had no structural business statistics on general practitioners. Because of the demand of reducing the response burden, a traditional survey has not been an option. Therefore the administrative data are the only sources. Since there is no value added tax for general practitioner services and since only a very small part of the businesses is incorporated, the revenues and expenses have to be determined via non-incorporated tax data. Because of the limited amount of available non-incorporated tax data at Statistics Netherlands, extra tax data for a sample of practitioners was asked for at the Inland Revenue Service. Information on jobs is obtained via the job file.

4.4 Sources and linking

Figure 1 gives an overview of the data that are used.

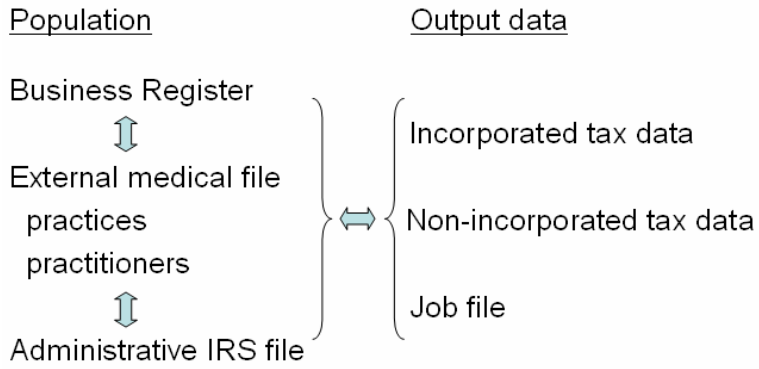

Figure 1: Overview of the data files that are used

In the job file, jobs of employees are ascribed to units in the Business Register. Also the incorporated tax data are linked to units in the Business Register, but for general practitioners only a very small part of the population is covered by incorporated tax data. The appropriate non-incorporated tax data are obtained via the so-called social-fiscal numbers of the practitioners. These numbers are obtained by linking an administrative file of the Inland Revenue Service (IRS) with an external medical file with information on all practices and all practitioners in the Netherlands. The external medical file was obtained from a research institute in the field of health care, and gives useful additional information on the practices (one-man or group practices, practices with or without a surgeon apothecary).

In the near future a redesigned Business Register will be in production, in which also the Inland Revenue Service will be used as one of the main sources. Also social-fiscal numbers of self-employed persons will be linked to the Business Register units. These developments will make the use of non-incorporated tax data much easier and will also solve the problem of incompleteness of the Business Register concerning the businesses of self-employed persons. External files, like the described medical file, might still be useful for the determination of the correct activity codes.

4.5 Structure of practices

Using tax data of practitioners that have only one business is quite straightforward. General practitioners that have more than one business make things more complicated. In many cases, a group of general practitioners work together in a partnership. Shared expenses, such as expenses for housing and labour, show up in the profit and loss account of the partnership, but more personal business expenses, like car expenses and depreciation of goodwill, are found on the profit and loss account of one-man businesses of the individual practitioners. Turnover may show up on either profit and loss account. In order to get the correct total amounts for this group of practitioners, the several profit and loss accounts have to be aggregated in the correct manner. Double counting should of course be prevented.

This combining of different tax data complicates the method of making use of administrative data. But somehow it also shows an advantage over the traditional surveys. Using survey data only, the structure of the practice is

often not visible and one just has to assume that the response provides the correct total amounts in which all relevant revenues and expenses are taken into account.

4.6 Output on revenues and expenses

By using tax data the following output is obtained: total turnover and "other revenues", the purchase value of the turnover (mainly purchase of medicines), labour expenses, depreciation and "other expenses". Because of the described linkage with the external medical file the output could be stratified to different kind of practices like one-man and group practices.

In principle many tax forms also contain information on the expenses for housing, inventory, interest, etcetera. Since the profit and loss accounts of non-incorporated businesses were not standardised, these variables were not filled in at every tax declaration, in which cases the concerning expenses are probably part of "other expenses". Because of this, output for these variables was not generated. In the future, when integrally available standardised non-incorporated tax data is available, the output can be more specified.

There are also variables that are asked in many surveys, but will never be obtained via administrative sources only, also not via the future standardised tax data. An example is a specification of turnover, although this is not an important issue for general practitioners. Another difference is that in many surveys labour costs for non-payroll employees (hired personnel, agency staff) is asked, while in the tax data these expenses are often not recognizable as such.

One should furthermore be aware of the fact that in surveys economical data is asked which might differ from the fiscal equivalents. For example, the economical and fiscal values for depreciation often differ. This theoretical difference might not be there in practice, since in many cases self-employed persons will just provide the same data for a survey, as they provide to the Inland Revenue Service, without bothering about the economical values.

4.7 Output on jobs

The output on jobs consists of the number of jobs of general practitioners and other personnel on the payroll. Again information on non-payroll employees cannot be obtained. Apart from the number of jobs also information on the jobs is given, like wage and the number of days worked during the year. In many traditional surveys information on the occupation of the employees is asked, but this information is not contained in the job file.

Since the job file is linked to the Population Register also the numbers of employees per age classes and sex are derived. In principle many more figures could have been derived. Ethnicity is also contained in the Population Register. Output is also possible for variables that are contained in other data files that are linked to the Population Register. Data on sickness absence is an example.

4.8 Editing problems

By using administrative data a huge amount of data can be obtained with a limited amount of time and effort. However, to generate correct output from these data, a lot of effort is often still needed. Almost every administrative file has some limitations (incomplete, not up-to-date) and the more sources are used, the more inconsistencies will be found. Also the linking of different data sources is often more time-consuming than expected. Furthermore there are fewer opportunities for checks since businesses cannot be contacted to discuss tax data.

5. CONCLUSION

The overall conclusion is that the use of administrative data for structural business statistics has many advantages especially for small-sized businesses. There will be less response burden, and figures for main variables can be obtained as demonstrated in the newly developed statistics on general practitioners. More specified output is possible in the near future, when Statistics Netherlands will also have the disposal of the standardised integrally available non-incorporated tax data.

Not all the desired output can be made by using administrative data only, like some specifications of expenses and turnover and information on non-payroll employees. This does not mean that the administrative data are not

useful for these variables. One can work with a reduced survey with a limited number of surveyed businesses if there is correlation between these variables and the variables that are contained in the administrative sources. This strategy is already used for small-sized incorporated businesses for many economical sectors. In this strategy administrative sources are exploited as much as possible.

REFERENCES

Linder, F. (2004), "The use of administrative registers and sample surveys in the Dutch Census of 2001", in E. Nordholt et al. (eds.) *The Dutch Virtual Census of 2001; Analysis and Methodology*, Statistics Netherlands, pp. 243-260.

Matthews, S. (2005), "Use of Administrative Data in Statistics Canada's Annual Survey of Manufactures", paper presented at the Work Session on Statistical Data Editing, Ottawa, Canada.

Statistics Finland (2004), "Use of Registers and Administrative Data Sources for Statistical Purposes; Best Practices of Statistics Finland", Helsinki, Finland.

Statistics Sweden (2001), "The Future Development of the Swedish Register System", Stockholm, Sweden.