

The Use of Administrative Sources for Economic Statistics – An Overview

**Steven Vale, Statistical Framework Division,
Office for National Statistics, UK**

1. Introduction

This paper describes the relationship between administrative data sources and economic statistics. It focuses on how administrative data can be used to supplement and enhance statistical outputs, as well as provide key inputs for the business registers that provide the sampling frames for business surveys.

It is based on material from the course “An Introduction to the Use of Administrative Sources for Statistical Purposes”, developed by the author for the “Training for European Statisticians” programme. It is also closely related to Chapter 20 “The Use of Administrative Sources” of the Eurostat Business Registers Recommendation Manual, also by the same author.

The paper first considers the definition of administrative sources, and the types of sources that may be of interest. This is followed by a consideration of the expected benefits and problems relating to their use. The practicalities of using administrative data are covered next, including the specific case of administrative registers that can be linked to a statistical business register, to form what is known as a "satellite register".

2. Definitions and Sources of Information

A traditional definition of administrative sources is that they are files of data collected by government bodies for the purposes of administering taxes or benefits, or monitoring populations. This narrow definition is gradually becoming less relevant as functions previously carried out by the government sector are, in many countries, being transferred partly or wholly to the private sector, and the availability of good quality private sector data sources is increasing. It is therefore necessary to consider a wider definition of administrative sources.

For the purposes of this paper, administrative sources are therefore defined in the widest possible sense:

Administrative sources are sources containing information that is not primarily collected for statistical purposes

This wide definition was proposed by the Eurostat Task Force on Administrative Sources in 1997. It allows the consideration of virtually the whole range of non-survey data, regardless of origin.

In 1998, Eurostat used this definition to collect information on the administrative data sources used for structural business statistics in Member States. The results of this work were published in the "Use of Administrative Sources for Business Statistics Purposes - Handbook of Good Practices"¹. Even though this has not yet been updated, it provides a useful insight into the ways in which different sources can be used.

Eurostat also collects information on the sources used for statistical business registers in their annual business registers questionnaire. This questionnaire is now also used by the UN/ECE so covers a wide range of countries. The most commonly used sources for business registers are taxation systems such as Value Added Tax (VAT), and personal income tax, or compulsory business registration systems, often administered by chambers of commerce. Information is also obtained in some cases from published accounts, private sector business data providers, or utility company records.

In some cases, different sources are used for specific categories of units defined by criteria such as size, economic activity or legal form. On average, European Union countries use around five different administrative sources to maintain their statistical business registers.

Some international organisations and individual countries are also exploring the possibility of automatically extracting data from the internal financial or management accounting systems of businesses. Under the wider definition set out above, this should also be considered to be use of data from an administrative source.

3. Why Use Administrative Sources?

The following paragraphs explore the reasons for using administrative sources, focussing on the advantages of this type of data. The use of administrative sources also raises a number of problems, which are considered, including possible solutions in the next section.

¹ http://forum.europa.eu.int/irc/dsis/bmethods/info/data/new/adso/handbook_en/adso_hbk.htm

The main advantage of administrative sources over survey data is usually cost. Surveys are very expensive, particularly if they are conducted as censuses, or involve the use of personal interviews. Administrative sources are often “free”, particularly if they originate from the public sector. Even if there is a charge, it is often still cheaper to use administrative data than to collect the same information via a survey. Less staff are usually needed to process the data, and there is no need for response chasing. The size and scope of statistical business registers in most countries makes it very difficult and expensive to satisfactorily populate and maintain them solely by survey data.

Using data from administrative sources also helps to reduce the response burden on businesses. Whilst businesses usually understand the reasons for supplying data for registration and taxation purposes, even if they do not like doing so, they may see statistical data requests as an extra, less necessary, burden. If they have already provided details to other government departments, they may become annoyed at receiving similar requests from the national statistical institution. A related advantage is that the use of administrative data may in some cases allow statistics to be produced more frequently, with no extra cost to businesses.

Administrative sources often give complete, or almost complete, coverage of the target population, whereas sample surveys often only directly cover a relatively small proportion. The use of administrative sources therefore eliminates survey errors, removes (or significantly reduces) non-response, and provides more accurate and detailed estimates for various sub-populations, e.g. businesses in small geographic areas. Coverage is often of particular interest from the point of view of statistical business registers, given that the European Union convention is that such registers should cover at least all businesses with a labour input of one person half time or more².

The use of administrative sources may increase the quality of statistical outputs and business registers by allowing access to more up to date information concerning certain variables e.g.

- The opening and closing of units and the dates of these events
- The economic activity code
- Information on unit locations
- Size data (number of employees or turnover).

As well as improving the timeliness of variables in statistical business registers, the use of data from administrative sources can, in some cases, improve the timeliness of statistics that are derived from business register data either directly or via samples. This is because statistical surveys generally take time to plan, to design and pilot forms, to analyse the population and optimise the sample etc.. This is particularly the case for annual or ad-hoc data collections. Therefore access to a suitable administrative source

² Thresholds are often higher in other countries, with many only recording businesses with paid employees.

via the business register can be a more efficient solution. It should be noted, however, that there are also likely to be cases where the use of administrative sources leads to a reduction in timeliness, particularly regarding short-term indicators.

Public opinion relating to the sharing of data, particularly between different government departments, varies considerably from country to country. Where public opinion generally accepts, or is in favour of data sharing, the increased use of existing data sources can help to enhance the prestige of a national statistical institution by making it more efficient and cost-effective.

4. Problems Using Administrative Sources

Although there are many good reasons for using administrative sources, there are also a number of common problems associated with their use. The following paragraphs outline some of these problems and propose methods to solve them, or at least to minimise their impact. One specific problem is that of getting access to administrative sources in the first place. This is treated separately in a companion paper.

The previous section considered how public opinion might favour the sharing of data in some countries. In other countries, however, there may be public unease at the thought of data being shared around government. It is very difficult to reduce such concerns, but possible approaches could include the publication of clear limits and rules regarding the use of data, ensuring that businesses understand that sensitive data used or collected for statistical purposes will not be fed back to other parts of government (particularly tax agencies).

This is in line with the United Nations “Fundamental Principles of Official Statistics”, where Principle 5 (*“Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents”*) encourages the use of administrative data, and taken together with Principle 6 (*“Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes”*), establishes the principle of the one-way flow of data.

The publication of analyses of the costs and benefits, both to government and to businesses, of the use of different sources may also help.

One major problem often encountered when using administrative sources is that the units used in those sources do not correspond directly to the definition of the required statistical units. The process of converting from administrative units (which may often be equivalent to legal units) to statistical units (e.g. enterprises or establishments) is known

as profiling, and typically is a function of the statistical business registers. Eurostat has published guidelines for this process in chapter 19 of their Business Register Recommendations Manual³.

As well as differences in the definitions of units, there are also likely to be differences in the definitions of variables between administrative and statistical systems. The data in administrative sources have generally been collected for a specific administrative purpose, and the needs and priorities relating to that purpose are likely to be different to those of the statistical system. For example, turnover for value added tax (VAT) purposes may not include turnover related to the sales of VAT exempt goods and services, whereas the statistical system is likely to require total turnover.

Similarly, the classification systems used within administrative sources may be different to those used in the statistical world. Even if they are the same, they may be applied differently depending on the primary purpose of the administrative source. Where classification systems are different, it is usually necessary to construct conversion matrices to map the codes in the administrative classification onto those required for statistical purposes (e.g. ISIC or NACE). Such mappings may be one to one, many to one, one to many or many to many. In the latter two cases, some sort of probabilistic allocation is often required. This should result in accurate coding at the aggregate level, but not necessarily at the level of individual units.

Another common problem encountered when using data from administrative sources relates to timeliness. Data may either not be available in time to meet statistical needs, or they may relate to a period which does not coincide with that required for statistical purposes, e.g. a tax year may not coincide with the calendar year required for structural business statistics.

There will generally be some sort of lag between an event happening in the real world, and it being recorded by an administrative source, this is then followed by a further lag before the data are made available to the national statistical institute. Lags relating to births and deaths of enterprises are a major source of business register coverage errors. If these lags can be measured, allowance can be made for them in any statistics based on register data.

Public sector administrative sources are generally set up for the purposes of collecting taxes or monitoring government policies. This means that they are susceptible to political changes. If a policy changes, administrative sources may be affected in terms of coverage, definitions, thresholds etc., or possibly even abolished completely. Such changes may happen suddenly, with little warning, particularly following a change of government. Reliance on a particular source always carries a certain degree of risk. These risks can be managed to some extent by legal or contractual provisions, regular

³ <http://forum.europa.eu.int/irc/dsis/bmethods/info/data/new/embs/registers/chapter19.doc>

contact with those responsible for the administrative source to try to get early warning of possible changes, and by drawing up contingency plans.

If data from several administrative sources are used to supplement survey data or populate a business register, it is likely that the national statistical institute will be faced with problems matching the data. Matching is relatively easy if there is some form of common identification number, but if not, it usually has to be based on variables such as name, address or economic activity code. In such cases it is likely that there will be a certain proportion of false matches and false non-matches, and the need for clerical investigation of possible matches.

Another problem where multiple sources are used concerns consistency between the sources. Data from one source may appear to contradict those from another source. This may be due to different definitions, classifications or differences in timing, or simply to an error in one source. To resolve such conflicts it is necessary to establish priority rules, by deciding which source is most reliable for a particular variable. Once a priority order of sources has been determined for a variable, it should then be possible to ensure that data from a high priority source are not overwritten from a lower priority source. This process is made much easier if source codes and dates are stored alongside variables for which several sources are available.

5. The Quality of Administrative Sources

The issue of quality in relation to administrative sources often calls for some sort of comparative assessment, either against other administrative sources, or against data from statistical surveys. If a statistician is faced with a choice between survey and administrative data, or between two or more administrative sources, how can he or she determine which source has the higher quality?

There is no simple answer to this question, but there are several possible approaches, which may be used in combination. It can be useful to compare sources in terms of coverage and accuracy of the variables. This can be aided by some sort of quality survey to determine the correct values of certain variables.

The closeness of administrative units and variable to the units and variables required for statistical purposes can be an important factor in determining the quality of an administrative source. The fewer transformations required, the lower the risk of error or bias.

Perhaps the best way to assess the quality of an administrative source is, however, to build up a thorough knowledge of that source, including the primary purpose of the source and the way the data are collected and processed. Thorough understanding of a source will allow a more accurate assessment of strengths and weaknesses.

To help develop and document this knowledge, particularly for the benefit of future staff, it is useful to develop some form of template to record key information. Figure 1 shows the high-level template used in the UK for business register source data.

Figure 1 – UK high-level template

Companies House Data	
• Framework:	Contract
• Frequency:	Quarterly updates, continuous on-line access
• Timeliness:	Good
• Quality:	Good
• Delivery:	CD-ROM / Internet
• Key content:	Legal name, company number

More detailed templates are also required to store more information on contacts, units, variables, quality and formats. The information in the templates should be agreed with those responsible for the administrative source, and should be reviewed regularly.

6. Using Administrative Sources in Practice

The following paragraphs consider the processes that are often necessary to transform data from administrative sources into the variables required for statistical business registers or outputs. The main processes are listed and briefly explained, though will need to be adapted to fit the specific requirements of each country and source.

A principle adopted by several national statistical institutes is that the statistical business register should serve as the gateway for administrative microdata. This means that all administrative data at the level of individual businesses or sites should be fed through the business register. In this way the units and variables in the administrative data source can be matched to those in statistical surveys based on the register.

- ***First steps and the transition from administrative to statistical data***

The first steps generally concern processing the administrative data to check the integrity of the data file and the quality and coverage of the information contained.

Preliminary analyses may be useful, or even necessary, to check several points including:

- The file is the expected size, i.e. it contains the expected number of records and variables
- The main variables are in valid formats e.g. dates are in an allowable range, text fields contain only text characters, numeric fields contain only numbers, etc.
- The coverage of the main variables e.g. that identity numbers, addresses, economic activity codes etc. are present for all units.
- The variables are plausible, e.g. numeric variables are in an acceptable range and any codes used are valid.

If the preliminary analyses show that the source is of sufficient quality, and may help to increase the quality of the business register, a second group of procedures are required to translate the data to meet statistical requirements. These procedures refer to the pre-treatment of the administrative data, including the application of statistical definitions of units and variables.

An administrative source is unlikely to use the same definition of units required for statistical purposes, especially in the case of complex businesses. The translation of the administrative data into statistical data may also involve steps such as the creation of algorithms or look-up tables to convert variables and classifications, e.g. tables that convert economic activity codes to ISIC or NACE codes.

- ***Linking and matching***

The next step is to identify the link between units in the administrative source and the corresponding units in the statistical business register. This is relatively easy if there is a common reference number for businesses, as long as there are no errors in this number in either the administrative source or the register.

If there is no common identifier, a probabilistic approach to record linking can be adopted by identifying correspondences or similarities in the name and address, or other characteristics (e.g. legal form, economic activity code, etc.). With this approach it is, however, possible to link units in error (false matches).

Quality checks of the results of the matching should be carried out. These can take the form of checks of auxiliary variables, e.g. economic activity codes, size or legal form. If these variables are consistent it is more likely that the match is correct. If not, further clerical checks may be needed, particularly where larger units are concerned.

It is also useful to periodically check the non-matched units and attempt to establish further links or to determine why they do not match. If the non-matched units are genuine, they may be due to timing or scope differences between the administrative source and the statistical business register.

Matching can be automatic or clerical, and in most cases is a mixture of the two. Automatic matching software is often used to determine definite matches and non-matches, whilst clerical input is usually needed for cases where automatic matching does not give a clear answer. A variety of automatic matching tools are used, some are available as commercial software, whilst others have been developed in-house by national statistical institutes.

- ***Satellite Registers***

One way of using administrative data in practice is to organise these data into specific registers linked to the statistical business register. If these registers meet certain criteria, they can be referred to as "satellite registers". Satellite registers have been defined by Eurostat as registers that are available to the national statistical system (in the National Statistical Institute or elsewhere), contain information about businesses, and fulfil the following conditions:

- They are not an integral part of the statistical business register as referred to in the business registers Regulation, but are capable of being linked to it.
- They are more limited in scope than the statistical business register, e.g. in terms of NACE, but within that scope they may have more extensive coverage of units and/or variables.
- They contain one or more variables that are not found in the statistical business register. Such variables are generally capable of being used for stratification purposes.

Thus satellite registers can be constructed using information from administrative sources, statistical surveys, or a combination of both. In some cases they may add, combine or otherwise transform variables, though in others they may be more or less identical to a particular source. To ensure that satellite registers are sufficiently coherent with statistical business registers, it may be useful to consider additional criteria, e.g. common unit identifiers, common definitions and classifications. The greater the coherence, the more useful a satellite register is likely to be.

The scope of satellite registers can be determined by:

- Economic activity - they may contain businesses with similar activities e.g. retail trade, accommodation, road haulage etc..
- Size - They may contain units with a certain number of employees or turnover over a certain level, e.g. they may contain the subset of "large enterprises"
- Characteristics - They may contain units with a common characteristic, e.g. those that engage in foreign trade (units on the register of intra-Community operators)

Variables specific to a particular sector of economic activity may be made available by using satellite registers, e.g. the variable "sales space" for businesses in the retail trade;

or the variables “category/number of stars” or “number of beds” for hotels. Satellite registers can add value to statistical business registers by adding a wider range of variables both for stratification and analysis purposes.

The use of satellite registers can increase sampling efficiency (i.e. a reduction in the number of units sampled, whilst still retaining the same level of accuracy of survey results) in two ways:

- The statistical business register includes links to additional stratification variables via satellite registers.
- The statistical business register benefits from the improved quality of variables due to the ability to cross-check this register with external sources.

The use of satellite registers can therefore reduce the response burden on businesses either through increased sampling efficiency or by removing the need to conduct a survey altogether. They may also include additional units not present in the statistical business register, thus extending the coverage of the population of interest.

7. Conclusions

In summary, the five key messages in this paper are:

- Administrative sources should be defined in the widest sense, to include any source, public or private, that contains data not primarily collected for statistical purposes.
- There are many benefits in using data from administrative sources for economic statistics and registers that provide the frames for business surveys, but the main benefit is usually reduced costs.
- There are usually a number of problems to overcome when using an administrative source, but these problems can be grouped into categories, for which other countries have usually found solutions.
- Most problems can be reduced by effective planning and management, and, above all, a detailed knowledge and understanding of the source.
- There may be costs involved in overcoming some of the problems with using administrative sources, but in most cases the benefits are still often greater than the costs.

8. Bibliography

- Report of the Eurostat Task Force on Administrative Sources, October 1997 - available on the Internet - <http://forum.europa.eu.int/irc/dsis/bmethods/info/data/new/adso/adso.html>
- Use of Administrative Sources for Business Statistics Purposes - Handbook on Good Practices - Eurostat 1999 - available on the Internet - <http://forum.europa.eu.int/irc/dsis/bmethods/info/data/new/adso/adso.html>
- Matching Records without a Common Identifier - The UK Experience, Steven Vale and Mike Villars, Research in Official Statistics (1/2000), Eurostat, 2001. Also available on the Internet: <http://europa.eu.int/en/comm/eurostat/research/conferences/etk-99/papers/vale.pdf>
- Methods for Automatic Record Matching and Linking and their use in National Statistics, by Leicester Gill, National Statistics Methodology Series No. 25, Office for National Statistics, 2001. ISBN 1 85774 420 9 or http://www.statistics.gov.uk/downloads/theme_other/GSSMethodology_No_25_v2.pdf (Note - 8.72mb PDF file)
- Registers in Official Statistics: A Swedish Perspective, Evert Blom and Folke Carlsson, see - <http://www.unece.org/stats/documents/1999.03.registers.htm> - Paper no. 14 (N.B. Many other papers on this site may be of interest)
- Eurostat Business Registers Recommendations Manual (particularly chapter 20) - see: http://forum.europa.eu.int/irc/dsis/bmethods/info/data/new/embs/registers/embs1_5.html
- Use of Registers and Administrative Data Sources for Statistical Purposes – Best Practices of Statistics Finland. Published by Statistics Finland in 2004