ORGANISATION DE COOPÉRATION ET DE DÉVELOPPEMENT ÉCONOMIQUES

OECD OCDE

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Meeting document 7

STATISTICS DIRECTORATE

SBS Expert Meeting "Towards better Structural Business and SME Statistics"
OECD, Statistics Directorate
3-4 November 2005
La Muette Room 4

Agenda Item 6 :

Business registers and SME's

Item 6(c):

The Coverage of Micro-Enterprises in Business Registers

Mr. S. Vale, OECD

The Coverage of Micro-Enterprises in Business Registers

Steven Vale, OECD / UK Office for National Statistics

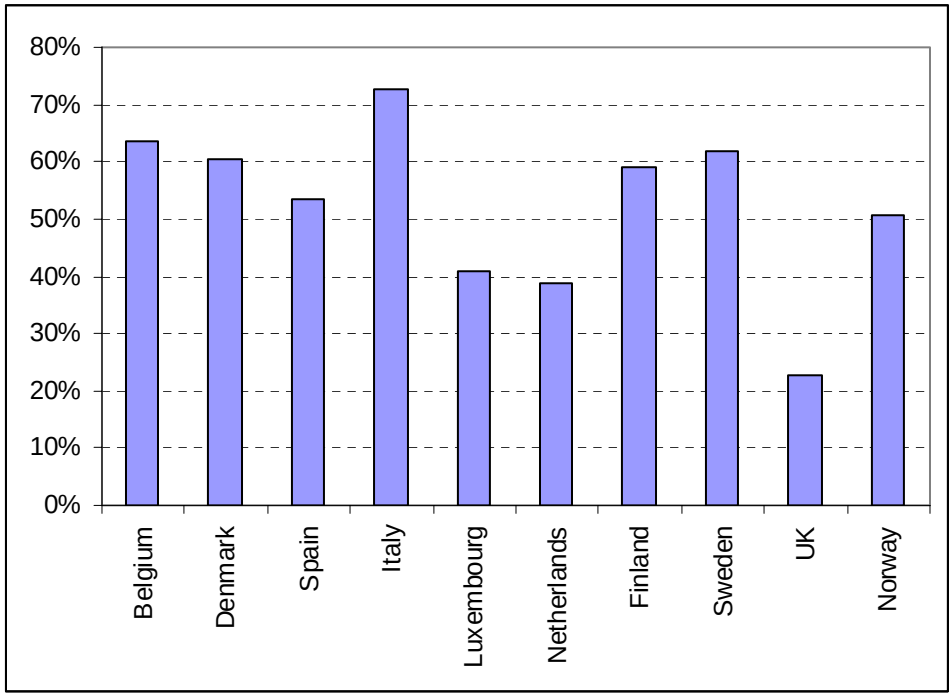
Introduction

For many countries, the main source of statistics on small and medium-sized enterprises (SMEs) is the statistical business register. This paper explores some of the issues relating to the use of register data for this purpose, with specific reference to the coverage and quality of data on very small businesses, often referred to as micro-enterprises. It provides some thoughts on how to improve data comparability, with the aim of stimulating further discussion on this topic.

The main issue affecting the comparability of data is the definition of what actually constitutes a SME. Many discussions have focussed on how big a business needs to be before it stops being included in statistics on SMEs, this is to some extent affected by the relative sizes of national economies. An enterprise considered large in a small country may be considered medium or even small in a larger country, thus it could be argued that a single fixed definition in terms of the upper size limits of the small and medium categories is not the optimal solution.

From a business register point of view, however, the main coverage issues lie mainly at the other end of the size spectrum. Here there seems to be rather more scope for defining internationally applicable definitions and thresholds. In terms of pure numbers of businesses, this is also the area where harmonisation could have the greatest impact, as the extent to which micro-enterprises are included in SME statistics can have a huge impact on data comparability. This is clearly demonstrated in the table below.

Table 1. Proportion of businesses with no employees



Source: Taken from “Business Demography in Europe, Results for 10 Member States and Norway”, Table 2.5, published by Eurostat in 2004. Data for 2001 (except Belgium – 2000)

These data are based on statistical business registers, where there is a requirement within the European Union to include all businesses with a labour input of at least one person half-time, and a recommendation to cover smaller businesses if possible. The variation in the above table is rather too large to be attributable to real differences between countries, and is at least partly due to the difference between required and recommended register coverage. Evidence to support this comes from other data sources, compiled on different methodological bases, which indicate figures such as 72.8% for the United Kingdom (2004)[1] and 75.0% for the United States (2001)[2]. For some Japanese data sources, the figure would be close to zero as only corporate businesses, or establishments with four or more persons employed are included.

Statistical business registers typically have a life-span of around ten to fifteen years before technological platforms and elements of the underlying models become obsolete. At present a number of countries are planning or developing the next versions of their registers. Thus, if the requirements for SME statistics can be more clearly defined, particularly those regarding coverage, there will be a better chance of feeding them in to the design of this new wave of registers.

Sources and Thresholds

As seen above, the coverage of very small units in SME statistics is usually determined by the coverage of such units in statistical business registers. These registers provide the source of basic SME data, and the mechanism for integrating survey and administrative data to supply additional variables.

In turn, the coverage of statistical business registers is very much dependant on the administrative sources that supply them with data. In many cases, administrative sources supplying data from sales or value added tax registrations are used. The thresholds for such sources vary from zero up to GBP 60,000 in the case of the United Kingdom. Where higher thresholds exist, data are often supplemented from other sources to mitigate the impact. Typically sources relating to payments to employees are also used, often making it impossible to define the actual threshold applying to a statistical business register, or data derived from it, in terms of a single variable.

Thresholds related purely to sales or value added can cause problems, as it is quite possible in certain economic activities, e.g. software development, for a business to have employees but no sales for a year or more, while it is developing products. Financial thresholds are also rather difficult to compare meaningfully across countries that have very different costs of living. Finally, even if it is possible to establish international definitions based on financial thresholds, it is very difficult to update them due to different rates of inflation in each country.

Thresholds relating to labour input are often more appropriate, but again it is important to know how that labour input is measured, e.g. in terms of wage-related costs, head counts of employees, or full-time equivalents, as this could also have a impact on comparability. For businesses with employees, head-count data are usually available, or can be estimated. For those without employees, a measure of labour input in terms of hours worked would probably be most appropriate. This would need to be modelled or estimated in some way, as the necessary information is unlikely to be directly available other than through a survey, and even then, only for a small proportion of such businesses.

[1] UK Department of Trade and Industry, Small Business Service - http://www.sbs.gov.uk/SBS_Gov_files/researchandstats/SMEStats2004.pdf
[2] US Small Business Administration - http://www.sba.gov/advo/research/data_uspdf.xls

Some countries require all businesses to be registered regardless of size, but even these are unlikely to record very low levels of business activity. For example, individuals who occasionally sell second-hand or surplus goods to neighbours, via markets, or through internet auction sites will often not be registered. Thus regardless of whether it is explicit or not, or even whether it can be defined, every statistical business register operates a threshold for the inclusion of small businesses, a threshold that is different in every country.

Data Quality Issues

One particular problem that often affects the smallest businesses in a statistical business register is data quality. Register maintenance resources are typically focussed on the largest and most complex enterprises, as these have the highest impact for most business statistics. This means that the smaller enterprises are maintained mainly, or in some cases (e.g. Australia) solely by administrative sources. This can cause problems with the timeliness of data, particularly where a business has ceased activity, but still has outstanding legal or tax obligations.

Turnover and employment are typically used as size indicators in business registers. Often at least one of these is unavailable for the smallest businesses, particularly recent start-ups, and it is rare for such businesses to have their size indicators updated more frequently than once per year. These time lags, and the fact that one variable is often imputed from the other, mean that size indicators for very small enterprises are often not as reliable as those for larger enterprises.

These problems can be reduced by introducing time-lags into the production of SME statistics, i.e. waiting for one or two years before producing data for a given reference year to allow the business register to be updated. Unfortunately, this can prove unpopular with users, who often want data to be more rather than less timely.

An alternative approach is to define a series of size-classes or thresholds based on data quality, so that data known to be of good quality are not compromised by the inclusion of categories of units with known problems. For example, a break between businesses with employees and those without would be better than defining a size class than included both. Thus a size-band of zero to nine employees would be split into two parts, one for one to nine employees, and the other for zero employees, with most of the coverage and comparability issues confined to the latter.

The concept of employees is proposed here as the basis for size-bands, because it is often a relatively easy variable to obtain from administrative sources, and the distinction between employer and non-employer businesses is already used by a number of countries. An alternative measure referred to as “persons engaged” or “persons employed”, which includes working proprietors and any unpaid workers, could also be considered, but this is often more difficult to measure in practice. The use of employee-based size bands also fits better with trends in business demography data, where it is seen as useful for policy purposes to separate out the creation of paid employment from the involvement of unpaid family workers.

Minimum size criteria for enterprises?

The discussion of thresholds and the diminishing quality of data on micro-enterprises lead to questions about the minimum size below which a business ceases to be of interest for statistical purposes. The answer will vary to some extent depending on the type of statistics

being produced. Businesses with no employees will be of little interest for statistics on wages, whereas those with low sales will have a negligible impact on most financial data. Statistics on SMEs and business demography, which focus more on numbers of businesses, are likely to require the widest coverage.

It is clear that a business with a hundred employees, producing goods or services for sale on market terms is of interest for virtually all business statistics purposes, whereas someone who services a neighbour's car once a year, in return for a token payment, is not. Indeed the latter is very unlikely to be measurable even if it was of interest. Thus the answer will lie somewhere between these two cases.

Before going any further in terms of the answer, it is useful to consider what is being measured, and what the data will be used for. SME statistics are, by definition, concerned with enterprises. Enterprises can be defined in various ways, but typically they have to produce goods or services for sale on market terms, whilst operating as an autonomous unit.

The definitions do not appear to offer any specific thresholds, but the autonomy and market orientation aspects are either explicit or implicit in most versions. If these aspects are expanded on, they can be seen as ruling out at least some business activities undertaken on a hobby basis, where profit maximisation is often a secondary consideration. They also provide a route to exclude "pseudo-enterprises", sometimes also referred to as "false self-employed" where a person acts as an employee of an enterprise, but for legal or tax purposes is technically self-employed.

SME and business demography statistics are often used as indicators of job creation. If an employee decides to supplement his or her income by starting a part-time, self -employed business activity, can this really be seen as the creation of a job? Arguably a job is only created at the time that the person decides to make this business activity their main source of income, so again, business activities that occupy only a few hours per work are of little relevance.

This line of reasoning suggests a threshold in terms of percentage of income might be appropriate. For example, to be considered an enterprise, a business activity must account for over 50% of a person's income. This approach might cause problems in terms of stability over time, and consistency. A business activity of someone with a high total income might not be counted, even though it generated more income than a similar activity that would be counted simply because it was carried out by a person with a low total income.

An alternative approach would be to return to the concept of hours worked, referred to in the section on thresholds above. If an assumption could be made that people split their time logically between different activities to maximise their income, a threshold in terms of hours worked could be appropriate. Unfortunately measurement difficulties tend to rule out the use of hours worked as a variable for very small businesses.

To get around the measurement problem, it is necessary to look at the variables that tend to be more readily available. One solution is to determine the average annual turnover per head for a control group of small businesses[3], for example those with between one and nine employees, and a specific category of economic activity (ISIC or NACE 4-digit code?). This can then be used to define a threshold equivalent to half a person, and thus to exclude any

[3] Note, the calculation of average turnover per head ratios can be greatly affected by atypical values, so in practice it is often necessary to use trimmed means, or a technique such as winsorisation to reduce the impact of such outliers.

non-employer business with an annual turnover equivalent to less than half the average turnover per head value for that activity.

There are also potential drawbacks with this approach, particularly for non-employer businesses with irregular turnover patterns, or for seasonal activities, which may occupy a person on a full-time basis for part of a year. It should also be noted that such a threshold would take a different value for each combination of country and economic activity, however, instead of being considered a threshold in its own right, it should be seen as a proxy for the more comparable threshold of half a person labour input. The big advantage of this approach, however, is that it provides a threshold that would be relatively easy to apply in practice, at a relatively low cost, in most countries.

If this sort of approach is adopted for SME statistics, it could have knock-on impacts for data on business demography and business start-ups. If businesses with a labour input of less than half a person are excluded from SME data on the basis that they are not really considered to be enterprises, it could be seen as logical to exclude them from business demography data as well. This would, in turn, help to increase the international comparability of business demography data, and to ensure consistency with SME data.

The problem of dealing with the “false self-employed” remains. As this phenomenon varies according to the tax and legal frameworks in place in each country, it would clearly be preferable to exclude these units from SME statistics. Unfortunately it is rarely easy to identify them in a consistent way using just administrative data, so labour force type surveys seem to offer the best chance of determining the proportion of small businesses accounted for by this group. As this phenomenon also tends to be more prevalent in certain economic activities than others, this dimension would need to be considered when collecting and processing survey data.

Metadata

Finally, it is important not to forget the role of metadata in informing those making comparisons of SME data between countries. Descriptive metadata are useful to give background information on specific issues, but some sort of simple reliability or comparability indicator for different size-bands could be a useful addition to existing data. This would also draw attention to the lower reliability of data in the smaller size-bands, increasing the pressure for further harmonisation.

Conclusions

To increase international comparability of SME data, it is necessary to focus on minimum as well as maximum size thresholds. Data quality issues, the nature of data sources, potential enhancements to those sources, and the role of metadata also need to be taken into consideration. Based on the text above, it is possible to put forward a number of proposals for further discussion:

- A first step towards improved comparability would be the definition of a range of size bands that take into account differences in quality and national data availability. This would help to resolve issues at both the upper and the lower end of the SME size range, whilst improving comparability in the short-term for relatively little effort.
- For the smallest units this would mean a break between those with one or more employees, and those with no employees. This should allow more meaningful comparisons between countries down to the one employee level.

- The zero employee category can then be further broken down into those that have a labour input of half a person or more, and so are more likely to fulfil the requirements of being an enterprise, and those with a lower labour input, which may not be real enterprises.
- The amount of labour input should be determined in terms of the national average labour input for small businesses with that economic activity, using turnover per head ratios.
- The use of survey data to determine the proportion of the “false self-employed” by economic activity, should be further explored.
- Consideration should be given to the use of reliability or comparability indicators as a form of summary metadata, to better inform users about data comparability.
- Any thresholds adopted for SME statistics should also be considered for data on business demography to ensure consistency.
- The requirements of SME statistics in terms of business register coverage and comparability are similar to those for business demography, and should be clearly articulated and considered as key user needs when statistical business registers are being re-designed.

By refining and adopting a combination of the above proposals, it should be possible to make some fairly major advances in the international comparability of SME data sets for relatively little cost. This has the potential to offer a number of “quick-wins” as far as the users of the data are concerned.