Integration of microdata from business surveys and the social statistics database

Gosse van der Veen
Director General Statistics Netherlands, Netherlands

Statistics Netherlands

Voorburg/Heerlen, December 2007

Explanation of symbols

.	= data not available
*	= provisional figure
x	= publication prohibited (confidential figure)
?	= nil or less than half of unit concerned
?	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2005-2006	= 2005 to 2006 inclusive
2005/2006	= average of 2005 up to and including 2006
2005/'06	= crop year, financial year, school year etc. beginning in 2005 and ending in 2006
2003/'04-2005/'06	= crop year, financial year, etc. 2003/'04 to 2005/'06 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Abstract

Effects of economic and social globalisation should not be treated separately in our statistical systems (at either national or international level). Integration of microdata from business surveys, social surveys and government administrations is inevitable for the compilation of microdatasets for research on causal effects of globalisation on welfare and employment.

1. Introduction

In national and international political debate the issue of globalisation is very prominent. Although there seems to be a difference in attention to this theme from several political and scientific perspectives as well as over time periods, it is nevertheless seen as an important factor which affects every day life in an economic and sociological sense ([12], [13], and [23]). The rejection in some EU countries of a European Constitution seems to reflect the perceived disadvantages of open borders. But on the other hand, for the Dutch situation at least, opening the borders has resulted in an additional purchasing power of an average 300 euro per citizen [8].

Many studies lack causal inferences on the effects of globalisation. The complicated context of several variables influencing the everyday behaviour of companies, people, and governments makes it very difficult to determine the causal effects of (inter)national policies. Furthermore, the economic and social effects of globalisation are usually treated mostly separately in many studies. In addition, the mere fact that national statistics ´stop at the border´ complicates the analysis of the international flows of goods, capital and persons. All these issues lead to a scattered and incomplete picture of the effects of globalisation. In other words, there is an urgent need for information about the effects of globalisation on the distribution of welfare and employment.

In this paper I shall argue that the effects of economic and social globalisation should not be treated separately in our statistical systems (at either national or international level), and that a balanced focus is required on the positive as well as the negative effects. In the following chapters I shall outline the approach Statistics Netherlands has developed in order to provide society with timely, accurate and consistent information on globalisation issues.

2. The demand for information on globalisation

2.1 International demand

From the international perspective, the OECD in particular has conducted outstanding work with its *Handbook on Economic Globalisation Indicators* [19]. This handbook, which in great part reflects data availability within participating countries, describes a set of indicators which outline the extent and intensity of the economic globalisation process. Although the ambition of the Handbook is to explain effects, it is clearly delimited to *descriptive* indicators.

As stated in the introduction, the core question addressed to statisticians is to map *coherent* information on the effects of globalisation, especially what it means in terms of growth, welfare and employment. From an international perspective mainly aggregated information at country level is required to compare countries and groups of countries. This type of information plays an important role in the definition of international standards for balancing economic power between countries and groups of countries. As long as we agree on the underlying conceptual and methodological issues, these criteria are very useful. But given the fact that most of these criteria are based on different administrative data and surveys, the correlation between these aggregates is, when not based on the same underlying microdata, artificial (spurious correlations).

2.2   The national demand for information

The national demand for information predominantly focuses on the effects of an open economy on distribution of welfare and employment ([6], [17], and [18]). The targets from the Lisbon Agenda in particular require that within the national territories information is available on the effects of policy measures aiming at reinforcement of productivity and participation. Neither question, how to improve productivity and how to increase participation in the labour market, can be solved by looking just at the economic dimension or just at the social dimension. This type of information requires data that enable us to describe the (causal) relation between features of an open economy (economic globalisation) and the distribution of individual welfare and employment (social globalisation).

Effective government policy requires profound statistical evaluation of the effects of macroeconomic, microeconomic and employment policy measures. In order to evaluate the policy outcomes at macro, meso and micro level, information about the effects of endogenous and exogenous factors on economic growth and employment are a *conditio sine qua non*. For this purpose, we need more than the standard information on trading flows, investment flows, outsourcing and outplacement of business activities. Basically, microdata on the relation between economic activity ('business') and its effects on individual employment and welfare ('people') are the answer.

The demand for information on effects of globalisation is twofold. First, much emphasis is given to the effects of globalisation on employment dynamics, especially the dynamics of job creation and job destruction. The determination of vulnerable groups on the labour market in particular would help policymakers to address more effectively the effects of outsourcing and offshoring. Also information on the effects of internationalisation of R&D, especially the role and position of the SME (small and medium-sized enterprises), is not directly available in existing data sources. Secondly, many studies look at the negative effects of globalisation. But globalisation also generates positive effects, like (incoming) investments in R&D, and growing levels of income [8].

The challenge is to find solutions for these questions within the domains of existing (micro) data. Solutions have to be found both in integration of existing microdata and the use of sophisticated statistical methods and models. It is policy at Statistics Netherlands not to impose additional administrative burden when data collection is needed.

3.   Integration of microdata

Statistics Netherlands has (exclusive) access to a wide range of government administrations and carries out a wide variety of business and social surveys. We therefore hold a unique position within the Netherlands, being the only institution able to provide consistent information and data on globalisation issues. Globalisation has a major impact on the sections of the economy described by Statistics Netherlands, such as international trade, traffic and transport, industry and services, construction, finances and production of multinational enterprise groups and (inter)national business demography, employment and distribution of welfare. The phenomenon is characterised by a high degree of complexity and numerous interactions betweens the various aspects as observed in the separate social and business statistics.

To a certain extent, regular statistics already provide information on globalisation, based on various European regulations such as the Structural Business Statistics, Short Term Statistics, European Business Register Regulation, Foreign Affiliate Trade Statistics, Intrastat, Extrastat, etc. At the national level, regular statistics hardly seem to provide relevant information. This limitation is mainly a consequence of the missing link between economic and social statistics. Furthermore, the *often perceived* absence of adequate detail and data on new phenomena related to globalisation also results in limited use of existing microdata.

As stated above, integration of microdata is inevitable for an accurate analysis of globalisation effects ([7], [9], [13], [14]). This approach is reinforced by the political decision in the Netherlands that new information requirements must not lead to additional administrative or response burden. Two projects currently underway at Statistics

Netherlands aim at integration of microdata within a controlled and standardised environment. One project, the Social Statistics Database, aims at social registers and surveys and therefore has individual persons, households, jobs, benefits and pensions as statistical entities ([1], [3], [21]). The other project, the Micronome database, matches business registers and surveys and therefore has the enterprise and enterprise groups as statistical units ([10], [11]). In order to match microdata from different sources we use both databases to compile customised datasets which are appropriate for analysis with standard software like SPSS. This microdata infrastructure is regarded as a good starting point for research on globalisation issues within the Dutch statistical programme.

Figure 1 gives a simplified model of the combined Micronome and Social Statistics databases. The heart of the model is based on three central registers: firstly, the business register which includes the total population of enterprise groups and enterprises; secondly the integrated database of data on jobs from the insurance administration, data on jobs from the tax administration and data from Statistics Netherlands' survey on jobs and wages [2]. This integrated database, the Jobs Register. is compiled yearly and contains information at the level of jobs, especially job characteristics. The third register, the Population Register, can then be used to add information on characteristics of the persons occupying these jobs. The population register also contains information on persons working in the Netherlands or receiving a Dutch benefit or pension while resident abroad.

For every job, there is a key in the business register with the enterprise, and a key identifier for the job or persons occupying that job (a number of the Industrial Insurance Board and the Income Tax number respectively). Subsequently, within the domain of business registers and surveys there is additional information on the structure of an enterprise (group), as there is additional information within the domain of social registers (for example social security, pensions, self-employed) and surveys on persons and households.
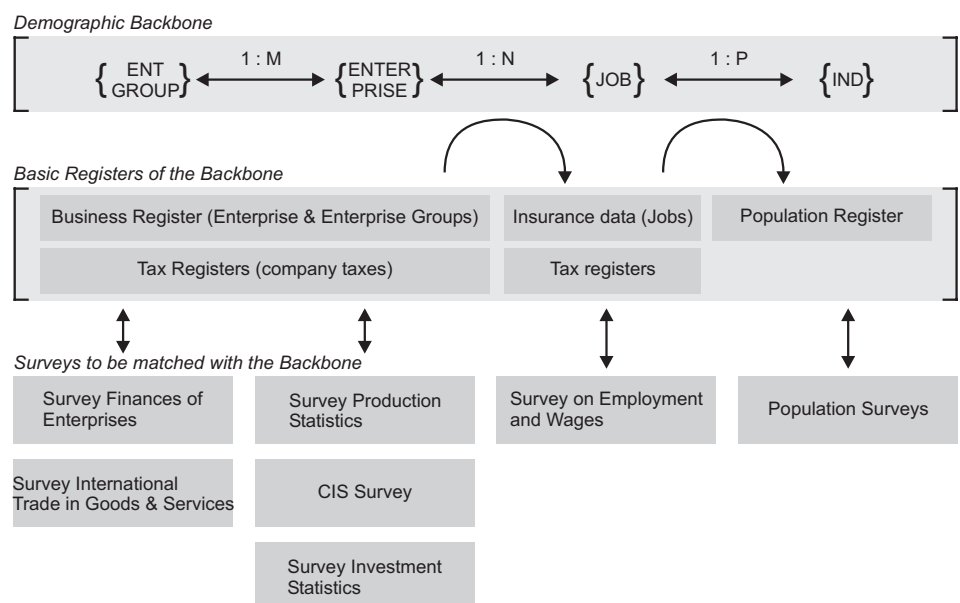
The core of the model is the insurance register. It plays a pivotal role between the business register and the population register. As stated above, in this way four types of statistical units are matched at micro level within these databases. These four statistical units form the demographic backbone of the databases. Consequently, different types of analysis are possible, using these statistical units. Basically, one can define causal models on how economic dynamics affects job dynamics, but also the other way around, i.e. how population dynamics are related to job dynamics. Furthermore, introducing job dynamics and population dynamics in business demography could improve its information value.

In practice only those units and variables which are necessary to conduct a specific analysis are selected. Components of the database are:

1. the datasets at the *level of enterprise groups* contain information on balance sheets and on international trade;
2. the datasets at the *level of enterprises* contain information on production, turnover, innovation expenditures;
3. the datasets at the *level of jobs* contain information on salary, wages, insurance, etc.;
4. the datasets at the *level of persons* contain information on individuals and households.

The coverage of available variables depends on whether the information is collected for administrative purposes or statistical purposes. In most cases information originally collected for administrative purposes has a high coverage, whereas information for statistical purposes mostly has a low coverage due to sampling (compared with administrative registers).

As stated above, the integration of units and variables depends on the research question under study. This implies that the output is leading in the composition of the microdata sets. We have just started defining basic sets which can be used for further research.

4. Our present position

Based on reported gaps in available information about globalisation and its effects on society, Statistics Netherlands has developed two lines of publications. The first line follows the conceptual framework of the OECD Globalisation Indicators, translated into more detailed indicators and extended with some specific Dutch indicators. This publication is called the Internationalisation Monitor and will be published on a regular basis on the website of Statistics Netherlands. The first publications were released in July at http://www.cbs.nl.

The second line aims at publications based on the integrated microdata from administrative registers, and business and social surveys. Selections from microdata themselves will be available for external research through the Statistics Netherlands' Centre of Policy Related Statistics, which facilitates research with microdata.

Both lines of publications have been developed within the spearhead research programme International Economic Relations. With respect to the microdata, the programme will produce a specific description of information available and information needed, at the level of variables, entities and relations betweens variables (hypotheses). This review will also document the availability and accessibility of information from private and public sources not yet included in the Statistics Netherlands' work programme. Private agencies such as Dun & Bradstreet and Bureau van Dijk hold valuable information for the purpose of studying multinational enterprise groups. Usability of these data sources is also part of the research projects on outward FATS and the ´European Business Register´ (Statistics Netherlands is in charge of the latter project). Further projects address issues like effects of outsourcing and offshoring on employment, innovative performance, and labour migration. For more details see [16].

5. Challenges

The approach of matching microdata from registers, business surveys and social surveys is not new. The first studies on so-called employer-employee microdata were carried out as early as the late 1980s [4]. Many studies were carried out on the basis of sample surveys among employees and subsequently matching these data with company data ([5], [15], [20], and [22]). Today we have more comprehensive data, better access to large administrations, and the appropriate IT technology to handle large datasets. Furthermore we have benefited from the advances of ongoing standardisation and harmonisation of government registrations. On the other hand, there is a strong and growing demand for consistent information about *effects* of globalisation, with very short time-to-market cycles. This marks an important change in the way we make our

statistics. It is not merely a matter of large datasets and fast software; it also challenges us to change our way of thinking. Several issues have to be addressed and to be solved:

1. *Conceptual issues*; the availability of integrated microdata on enterprises and persons is a ´mèr a boire´ for research on effects of globalisation; the major advantage being the possibility to make causal inferences on aspects of economic behaviour of companies and the outcome at individual level, especially on welfare and employment, based on comprehensive models on economic and social dynamics.

2. *Methodological issues*; any selection of units and variables requires recalculation of weighting factors, depending on the unit of analysis and the corresponding population frames. Another methodological aspect is the application of hierarchical data models. Traditionally, this type of analysis is not part of the day-to-day statistical work, which is mostly concerned with non-hierarchical datasets and models. So efforts will have to be undertaken to train staff in this type of analysis.

3. *Organisational issues*; in order to access and maintain the vast amount of microdata, dispersed over a wide variety of files and metafiles, a systematic method of processing and compiling microdata-sets for specific analysis is required. Statistics Netherlands is carrying out important redesigns of its statistical processes (the Masterplan and the Redesign of Economic Statistical Processes) in order to cope with the new demand for consistent and timely microdata.

4. *Dissemination issues*; the demand for matched microdata of business and social surveys, and the demand from research institutions, universities and ministries will increase rapidly. However, time-to-market periods to respond to this demand have to be short, and the appropriate expertise is not always at hand. In order to match the demand we shall have to cooperate more closely with certified research institutions and universities so that we can use the capital in our microdata more effectively, and speed up the development of new expertise in new areas of research. For this purpose Statistics Netherlands has a system of Spearhead Research programmes, uses long-term framework contracts with universities, and - with its Centre for Policy Related Statistics – has an extended facility for using microdata, on site or via remote access. Generally, the role of the NSI as an 'intermediate' in public information services will intensify, while at the same time it will retain a more or less constant position as the provider of ready-made tables and standardised statistics for the general public.

5. References

[1]   Al, P. and Bart F.M. Bakker (eds.), 2000, Re-engineering Social Statistics by micro-integration of different sources. Netherlands Official Statistics, jrg. 15, nr. (summer)

[2]   Arts, C.H. and Hoogteijling, E.M.J. (2002), Het Sociaal Statistisch Bestand 1998 en 1999, Sociaal-economische maandstatistiek, december 2002, pp. 13–21. [The social statistical framework 1998 and 1999, Social-economic monthly statistics, December 2002, pp 13-21].

[3]   Bakker, B.F.M., 2002. Statistics Netherlands' Approach to Social Statistics: The Social Statistical Dataset, In: OECD Statistics Newsletter, vol. 2002, nr. 11, blz. 4-6

[4]   Bayard, K et. al. (2002), The 1990 Decennial Employer-Employee Dataset, Research Paper NBER, Washington.

[5]   Bender, S. et al (2007), The Analysis of Firms and Employees: Quantitative and Qualitative Approaches, Paper presented at the Conference on Firm and Employees, September 29-30, 2006, Nurnberg, Germany [updated version]

[6]   Brinkhorst, L.J. and De Geus, A.J., (2004), Kiezen voor groei, Min. EZ en SZW, Den Haag, brief van de ministers EZ en SZW aan de 2e Kamer). [Choosing growth, Ministry of Economic Affairs and Ministry of Social Affairs and Employment, The Hague].

[7]   CEIES, (2003), Statistik und die wirtschaftliche Globalisierung, Report of the 22nd CEIES Seminar, Copenhagen, June 2-3, 2003. [Statistics and the economic globalisation].

[8]   CPB (2006), China and the Dutch Economy; stylised facts and prospects, Research Document 127, Scheveningen.

[9] Denis, C., K. McMorrow and W. Röger, (2006), Globalisation, Trends, Issues and Macro Implications for the EU, European Commission, Directorate-General for Economic and Financial Affairs, Rapport Nr. 254.

[10] Diederen, B., (2001), Handboek Micronoom, interne CBS notitie, Heerlen/Voorburg. [Handbook on Micronoom, research report Statistics Netherlands, Heerlen/Voorburg].

[11] Goedegebuure, R.V., and M.J.G. Luppes (2003), Studies on internationalisation using integrated micro-data (*The MICRONOME database at Statistics Netherlands),* Paper CAED conference 2003, London.

[12] Gurria, A. (2007), OECD, a relevant hub for global issues (Opening statement made by the Secretary-General to the OECD Ministerial Council Meeting 2007).

[13] Hamermesh, D. (2007), Fun with matched Firm-Employee Data: Progress and Road Maps, Forschungsinstitut zur Zukunft der Arbeit (IZA), Discussion Paper IZA No. 2580.

[14] Hatzichronoglou, Th. (2006), The Impact of Offshoring on Employment: Measurement Issues and Implications (preliminary version), Paris.

[15] Lane, J., and B. Stephens (2006), Integrated Employer-Employee Data: New Resources for Regional Data Analysis (US Census Bureau)

[16] Luppes, M.J.G. & P. De Winden (2007), Open Economy and its Relation to Employment and Welfare (Research Agenda of Statistics Netherlands on Globalisation), OECD Paper DAF/INV/STAT/RD(2007)4, Paris.

[17] Ministry of Economic Affairs, (2005), Nationaal Hervormingsprogramma 2005-2008 in het kader van de Lissabonstrategie, (rapport tbv de Europese Commissie), Ministerie EZ (eindredactie). Publicatienummer 05EP16 [National Reform Program 2005-2008 within the scope of the Lissabon strategy, (report on behalf of the European Commission), Ministry of Economic Affairs (editorial board). Publication number 05EP16].

[18] Ministry of General Affairs (2007), Samen werken, samen leven, Beleidsprogramma Kabinet Balkenende IV 2007-2011, Den Haag. [Working together, living together; Policy Program Cabinet Balkenende IV 2007-2011, The Hague]

[19] OECD (2005), OECD Handbook on Economic Globalisation Indicators, Paris.

[20] Office of National Statistics UK (2005), Abstracts Analysis of Enterprise Microdata Conference, London.

[21] Schulte Nordholt E., M. Hartgers, R. Gircour (eds.), 2004, The Dutch Virtual Census of 2001, Analysis and Methodology, (Voorburg, Statistics Netherlands)

[22] Spittal, M (2002) The New Zealand Conference on Database Integration and Linked Employer-employee Data, Social Policy Journal of New Zealand, Issue 18, June 2002.

[23] Van Bergeijk, P.A.G. (2007), Globaliseren zonder angst, Special Economisch Statistische Berichten over globalisering. [Globalising without fear, Special Economic Statistical Bulletin on globalization]