A proposed model for microintegration of economic and social data

Paul de Winden[1], Koos Arts[2] and Martin Luppes[1]

Statistics Netherlands
[1]Division of Business Statistics, Statistical Analysis Department, PO BOX 4481, 6401 CZ Heerlen, The Netherlands
[2]Division of Social and Spatial Statistics, Statistical Analysis Department, PO BOX 4000, 2270 JM Voorburg, The Netherlands

Abstract

Globalization affects all aspects of economic and social life. In order to study the effects of an open economy on employment and welfare, combined microdata from business surveys, social surveys and administrative registers are required to make causal inferences. Statistics Netherlands uses two sets of microdata to construct a framework in order to analyze the complex relationships between the dynamics of enterprises and the outcome on employment and welfare. The combined dataset contains hierarchical data on four different statistical units: enterprise groups, enterprises, jobs and individual persons. The backbone of the system contains information from administrative (governmental) registers and therefore covers the total population of enterprises and employees. The required variables for analytical purposes are retrieved from business and social surveys. Depending on the type of surveys and cross sectional slices of units the researcher has to deal with the methodological challenge of sample reweighting and the use of multilevel models. However, these issues are outweighted by the benefits of combining microdata from different sources. In this study, first results are presented on the relationship between job creation, job destruction and economic behaviour of firms.

Introduction

Economical and social effects of globalisation are handled mostly separately in many studies. 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 flows of capital, production, employment and trade. The mere fact that most surveys seems to ´stop at the border´ complicates the analysis of these international flows. This leads to a scattered and incomplete picture of the effects of globalisation.

Effective (government) policies requires profound statistical evaluation of the effects of macroeconomic, microeconomic and employment policy measures. In order to evaluate these outcomes on 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 (Bayard, 2002; ONS Conference 2005; Lane & Stephens, 2006).

In this paper we will present the way Statistics Netherlands matches microdata from administrative registers, business surveys and social surveys. In ‘Sources and methodology’ we describe the basic databases used in order to answering questions on the effects of globalisation. These are the MICRONOME database of microdata from business registers and surveys, and the Social Statistical Database for administrative registers and surveys on persons and households. The next

chapter addresses the basic model of integration of social and business microdata. Next, we will present some preliminary results based on matched microdata and we will outline some future research.

Sources and methodology

The MICRONOME database

The MICRONOME database is composed of different business surveys and administrations. Both data from enterprises as well as enterprise groups are combined within this framework. MICRONOME was generated in 1996 with the start of research on globalization and herewith the importance of large enterprises. These issues require microdata on various variables such as financial structure, employment, international trade, value added, innovation expenditure and so on. MICRONOME is a harmonized micro-database which is intended to be a co-ordination framework for business economical variables and an integration framework for different statistical units, containing business survey and registration data as depicted in the table below.

TABLE 1 Overview of integrated surveys and registers and their corresponding statistical units in the MICRONOME database

Business survey or register	Survey (S) or Register (R)	Statistical unit*	Contents (variables)
Business Register (ABR)	R	ENT, EG	Economic activity, sizeclass, legal form
Survey Finances of (non-financial) Enterprise Groups (SFO)	S	EG	Balance sheet, profit-and-loss account (e.g. turnover)
International Trade Surveys (ITS)	S	EG	International flows of goods and services
Annual Production Surveys (APS)	S	ENT	Turnover, number of employees, operating expenses, provisions
Annual Investment Surveys (AIS)	S	ENT	investments
Survey on Innovation and R&D (CIS)	S	ENT	Innovation expenditure, R&D personel
Survey on Employment and Wages (EWL)	S	ENT	Number of employees, average pay

* ENT = Enterprise; EG = Enterprise Group

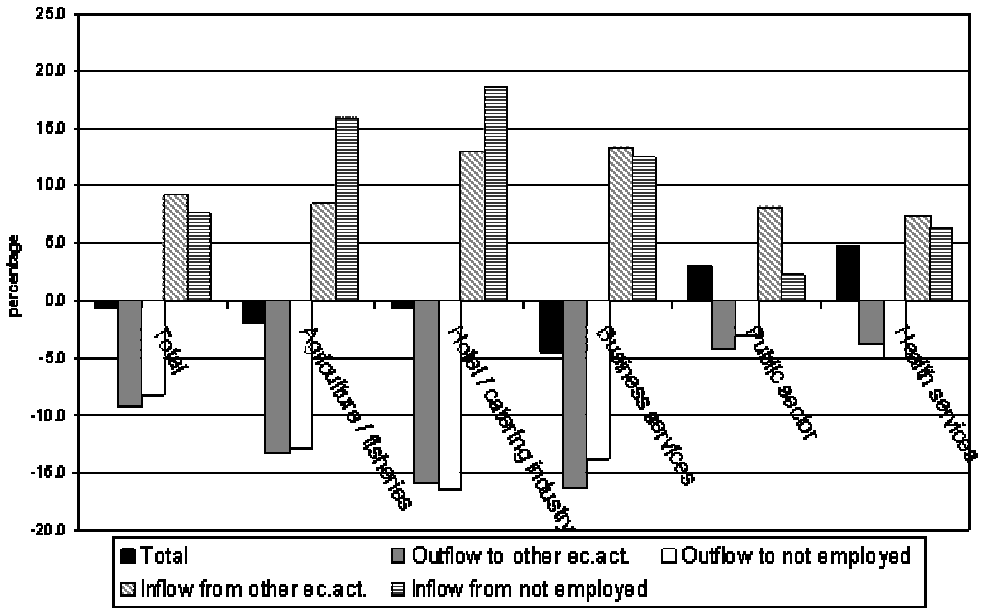
MIRCONOME is a dynamic database from which the researcher can make an indiviual selection of variables depending on the goal and hypothesis of his study. Theoretically, almost one thousand variables can be integrated into this framework, coming from the surveys and registers as depicted in the table above. The various surveys contain a different number of variables and records. For example, the Annual Production Surveys are made up from over 80.000 enterprises containing almost 100 variables. The Survey on Innovation and R&D holds approximately 3.000 enterprises with over 500 variables.
In this paper, applications and possiblities of the MICORONOME framework combined with the Social Statistical Database, in particular for the purpose of globalization, will be elaborated.

Social Statistical database (SSB)

The Social Statistical Database (SSB) of Statistics Netherlands mainly consists of administrative data on persons, households, jobs, benefits and pensions. It covers the entire Dutch population, including people living abroad but working in the Netherlands or receiving a benefit or pension from a Dutch institution. Various sources with data on jobs are integrated for the SSB database. These sources are among others insurance data, tax data and data gathered from the Dutch Survey on Employment and Wages (EWL). In the SSB, several characteristics on persons are available, such as gender, age, ethnic group, region, position in the household and type of household, (household)income, position on the labour market and earned wages. On the microlevel, a direct

relation between employees and enterprises can be established because employees' social security numbers and administrative enterprise numbers are available together in the administrative sources. Subsequently, the administrative units of enterprises, for example tax numbers, can be translated to statistical units of enterprises. For each enterprise data on characteristics of its employees can therefore be compiled. For example the share of males in an enterprise, the share of older employees or the share of persons with relatively low or high wages can be calculated. Data on education are not yet available at an integral level because in the Netherlands, education is not registered for the entire population. For the younger population (approximately until the age of thirty) data are available from registers of different school types. For the rest of the population data are available from the Dutch Labour force Survey (for a period of 10 years) which covers about 10 percent of the Dutch population in the age of 30 years and older, which can be used after reweighing the data for the specific groups under study. Statistics Netherlands has recently compiled this dataset on the educational level.
The SSB is a useful tool in revealing dynamics on the labour market because this framework contains data for all participants on this labour market (employees, self-employed persons, freelancers) over different periods of time. For example, the flows (transitions) of employees for different economic activities can be shown by comparing the employment situation at the end of September of year t and year t+1. This is illustrated in Figure 1 for some industries. The sector of agriculture and fisheries shows a decline in employment of about 2 percent between September 2002 and September 2003. This is the net result of 13 percent outflow of employees to other economic activities, 13 percent outflow out of the labour market (not employed), 8 percent inflow of employees from other economic activities and 16 percent inflow from not employed. This illustrates that a quarter of the population of employees in agriculture and fisheries in the Netherlands is renewed between September 2002 and September 2003. The hotel and catering industry (hotels, restaurants and bars) is the most dynamic industry, followed at short range by business services. The public sector (government) and health services appear to be much less dynamic in terms of employment shifts and therefore exhibit a more stable population of employees.

FIGURE 1 Gross flows of employees per economic sector in the Netherlands, September 2002 – September 2003.



Integration of economic and social microdata

On the playground of current economic developments, the phenomenon of globalization also strongly affects employment and social aspects, which is of major concern to society and

governments. Reliable statistical information is urgently needed on relations between business dynamics and various other economic characteristics on one hand, and data on employment shifts, labour market and underlying characteristics on persons (such as age, gender, ethnicity and education) on the other hand. Without any additional data collection, Statistics Netherlands is able to relate business and social data from various surveys and registers through linking them at the microlevel.
A pilot study was performed using 2002 and 2004 data from the previously described MICRONOME and SSB databases. The linking was performed at the microlevel of enterprises (ENT) and JOBS as depicted in Figure 2. All persons with an official employment contract at a registered enterprise (source: Business Register) are recorded in the SSB database with a job. An aggregation of jobs can subsequently be calculated per enterprise, resulting in a SSB-integration database containing over 400 thousand enterprises. For each enterprise, jobs are added and adjusted for part-time and duration factors, resulting in number of manyears expressed as Full Time Equivalents (FTE). Furthermore, several others variables can be aggregated per enterprise, among others average pay, number of female employees, number of employees older than 50, number of highly payed employees etc. In this way, by linking two major databases, a new framework is established, enabling analyses and output to be generated on the relation between business data and jobs of persons and their social background variables.

FIGURE 2 Relations between economic(left) and social (right)statistical units and corresponding surveys and registers at Statistics Netherlands

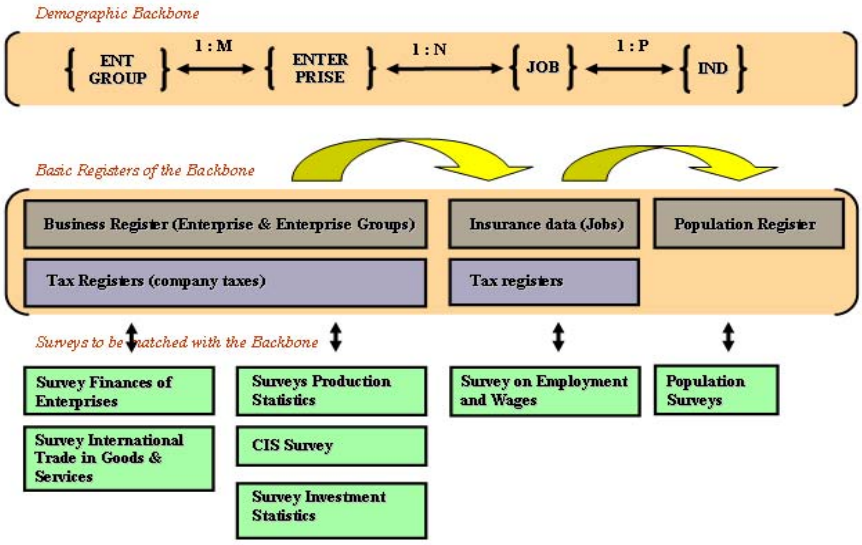


Figure 2 illustrates the relations and linking possibilties between the various statistical units used at Statistics Netherlands. On the left, the two business units are represented, i.e. enterprise group (EG) and enterprise (ENT). The enterprise group is the highest level of organisation for a company where financial decisions are made while the enterprise is the operational unit where the actual production proces is performed. A major company generally comprises one overall enterprise group and several enterprises with more or less different products and production chains. Surveys and registers for business statistics and their operational statistical unit are outlined in Table 1. On the right side of Figure 2, JOBS and individuals (IND) respresent the two statistal units applied in social statistics such as labour force and household surveys, assembled in the previously described SSB.

Preliminary results

A pilot study on linking the MICRONOME and SSB databases was performed using 2002 and 2004 data. The key variable for microintegration was the enterprise with its unique enterprise

identifier as defined by Statistics Netherlands. In this way, records from both databases can be linked one-on-one, in this way combining variables from various sources into one new framework
As outlined in the MICRONOME section, the number of business variables that can potentially be employed is innumerable. It is very important to emphasize that the selection of variables strongly depends on the research question under study. This implies that the intended output defines the composition of the microdata sets. Statistics Netherlands have only just started the microintegration of economic and social data, consequently the composition of microdata sets for its own research as wel as for external users still is in an initial phase.
A selection of four target variables from MICRONOME was selected for the present pilot research, i.e. turnover, value added, innovation expenditure and number of R&D employees. From the SSB-integration database, the number of manyears expressed as Full Time Equivalents (FTE) was picked for initial analysis. After linking the selected databases at enterprise identifier level, a panel of 346 thousand enterprises was formed which all possessed jobs in 2002 as well as 2004. Furthermore, for almost every enterprise, background business characteristics were available from the Business Register, such as economic activity (branch), sizeclass and legal form.
As discussed previously, business variables in MICRONOME unfortunately are not available for all enterprises because most variables are derived from statistical surveys with a limited sample. For approximately 54 thousand enterprises, data are on hand for turnover and value added. In the case of innovation expenditure and number of R&D employees, merely 2 thousand enterprises have been inquired for data. For the latter selection of enterprises, it is important to mention that this, to a large extend, comprises all enterprises that actually innovate.
Table 2 presents a classification of enterprises into economic sectors and firmsize, by this giving a basic understanding of the research panel and the number of enterprises involved in this pilot study. Notably, more than half of the enterprises are small and medium sized firms in the sector of government, social insurance, education, health service and remaining services (Quaternary sector). The industrial firms (Secundary sector) comprise relatively high numbers of large enterprises.

TABLE 2 Number of enterprises in the pilot research panel per economic sector and firmsize

Economic sector	Firmsize unknown	LE*	SME*	Total
Primary sector	152	38	17,480	*17,670*
Secondary sector	629	2,784	48,964	*52,377*
Tertiary sector	363	1,804	49,011	*51,178*
Quaternary sector	2,238	2,488	186,423	*191,149*
Sector unknown	33,992	0	25	*34,017*
Total	*37,374*	*7,114*	*301,903*	**346,391**

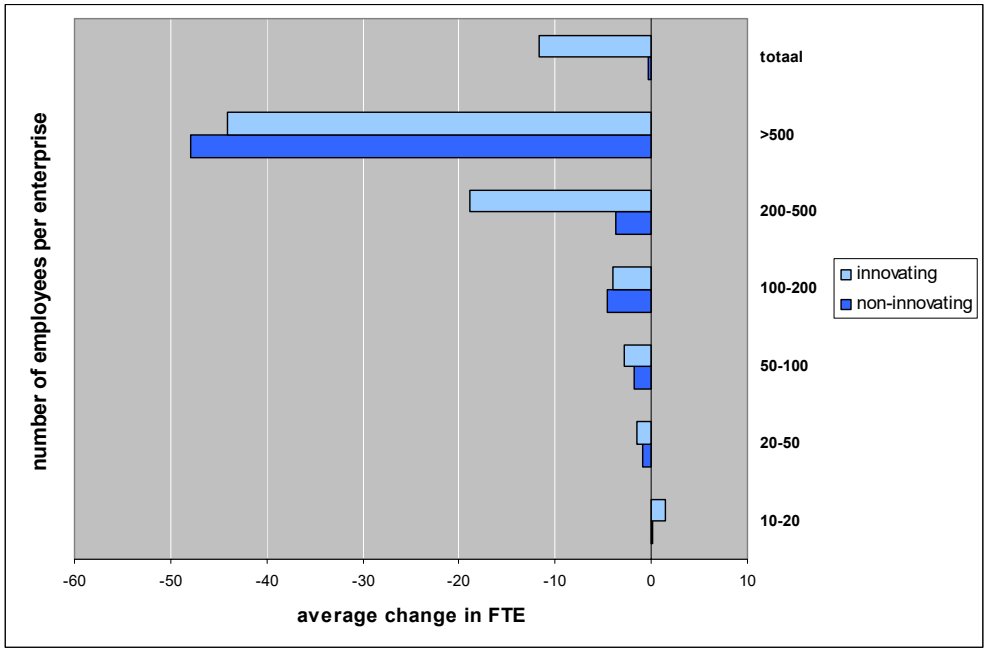
* LE = large enterprises (>100 employees); SME = small- and medium-sized enterprises (<100 employees)

In the present pilot study, some focus was directed at relations between business innovation and R&D on one hand, and changes in manyear (FTE) on the other hand. This illustrates just one out of many interesting perspectives which may emanate from the linking of the MICRONOME and SSB databases. Figure 3 illustrates the average changes in FTE per firm sizeclass (number of employees) with a breakdown for innovating or non-innovating enterprises. These preliminary data already show some interesting results that this type of integrated analyses might produce.

Future research proposals and challenges

Microintegration of economic and social data, as described above, is embedded in the ‘Spearhead International Economic Relations (SIER)’. This research program, started in January 2007, initially focuses on investigating the internal and external need for information on globalisation effects. A major part of this information is requested by national and international governments, as such included in several policy documents as well as in the Lissabon agenda. The latter is an agreement of the European Commission which aims to to enhance the international competiveness of the economy of the European Union.
Briefly, the approach of Statistics Netherlands consists of two seperate tracks. The first contains the publication of regular data in the so called Internationalisation Monitor. This is not further discussed is the present paper. The second track entails the analyses on integrated social and economic microdata as descibed previously. These analyses can be executed by Statistics Netherlands independently or in collaboration with other parties.

FIGURE 3 Average changes in manyear (FTE) for innovating and non-inovating enterprises for various size-classes (number of employees), 2002-2004



In the annual planning of Statistics Netherlands for 2008, some new research initiatives on microintegration are proposed in collaboration with a third party, among others:

- Internal collaboration with the 'Spearhead on Innovation and Productivity'. Within the SIER spearhead, innovation is a crucial theme of globalisation research. An important policy demand is the effect of outsourcing of R&D on innovative performance of firms. These effects can be considered in a national or international perspective (cross border spill-overs).
- Internal collaboration with the 'Spearhead on Social Dynamics'. Within the SIER program, the effects of business dynamics on labour market and employment are also taken into account. On the other hand, business characteristics can be applied as background variables for studying social dynamics.
- Also in 2008, a collaboration will be started with Maastricht University, which will result in one or more dissertations. Within this theme, a so called 'employer/employee' dataset wil be developped to facilitate studies on migration of R&D personel and the effects of reallocation of specific R&D knowledge by firms on its operating results.
- Starting in 2008, a frequently returning publication will appear on a specific theme that is related to globalisation effects. This publication will be brought about in co-operation with the Dutch Ministry of Economic Affairs.

References

1. Arts, C.H. en 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].

2. Bayard et. Al. (2002), The 1990 Decennial Employer-Employee Dataset.

3. 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].

4. Diederen, B., (2001), Handboek Micronoom, interne CBS notitie, Heerlen/Voorburg. [Handbook on Micronoom, internal memorandum Statistics Netherlands, Heerlen/Voorburg].

5. 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.

6. Hamermesh, D.S. (2007), Fun with matched firm-employee data: progress and road maps. Discussion paper No. 2580, January 2007.

7. Luppes, M. and P. de Winden. (2007), Open economy and its relation to employment and welfare (Research agenda of Statistics Netherlands on globalisation). Workshop on International Investment Statistics, OECD, Paris.

8. Lane & Stephens (2006), Integrated Employer-Employee Data: New Resources for Regional Data Analysis (US Census Bureau).

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

10. OECD, (2005a), OECD Handbook on Economic Globalisation Indicators, Paris.

11. OECD, (2005b), OECD Economic Globalisation Indicators (measuring globalisation), Paris.

12. ONS Analysis of Enterprise Microdata Conference (London, 2005).

13. Statistics Netherlands, (2006a), Plan van Aanpak Informatie Ontwikkeling BSH, interne CBS notitie, 4 april 2006. [Plan of approach on information development Statistical Analysis Department, internal CBS memorandum, April 4, 2006].

14. Tiggeloove, N., (2005), Globalisering en het aanpassingsproces van de Nederlandse economie, interne EZ notitie. [Globalisation and adaptability of the Dutch economy, internal memo Ministry of Economic Affairs].

15. Veen, G. van der (2007), Integration of microdata from business surveys and the social statistics database. Paper DGINS September 2007 Budapest, Voorburg/Heerlen.