

OECD Science, Technology and Industry Working Papers
2019/03

DynEmp and MultiProd: Metadata

Isabelle Desnoyers-James,

Sara Calligaris,

Flavio Calvino

https://dx.doi.org/10.1787/3dcde184-en

OECD

OECD Science, Technology and Industry Working Papers

The release of this working paper has been authorised by Andrew Wyckoff, OECD Director for Science, Technology and Innovation, and by Laurence Boone, OECD Chief Economist.

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the authors. Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed, and may be sent to Directorate for Science, Technology and Innovation, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

Note to Delegations:

This document is also available on O.N.E under the reference code:

DSTI/CIIE/WPIA(2018)5/FINAL

This document, as well as any data and any map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 811181. The European Commission is not responsible for any use that may be made of the information produced by the project.

© OECD 2019

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for commercial use and translation rights should be submitted to rights@oecd.org.

DYNEMP AND MULTIPROD: METADATA

Isabelle Desnoyers-James, Sara Calligaris and Flavio Calvino[1]

Abstract

This report presents a new effort to collect comprehensive metadata for DynEmp and MultiProd, two OECD distributed microdata projects that collect information to analyse employment dynamics and productivity in a harmonised way across countries. It gives an overall description of both projects, presents the methodologies used for the data collection, summarises the main features of the data by country and provides further country-specific information.

Table of Contents

1. The distributed microdata approach 5

2. The MultiProd project 7

2.1. Overall description 7
2.2. Methodology 7
2.3. MultiProd metadata 12

3. The DynEmp project 30

3.1. Overall description 30
3.2. Methodology 31
3.3. DynEmp metadata 33

Appendix 46

Disclaimer for New Zealand data 46

Endnotes 48

Tables

Table 2.1. Productivity measures used in MultiProd 9
Table 2.2. National data sources 13
Table 2.3. Sectoral aggregation 14
Table 2.4. Sectoral classifications and coverage 15
Table 2.5. Labour input variables 16
Table 2.6. Recording units for employment and capital, and availability of intermediate inputs 17
Table 2.7. Thresholds in the production surveys 18
Table 3.1. National data sources 34
Table 3.2. Time coverage 35
Table 3.3. Sectoral classifications and coverage 36
Table 3.4. Entry and age information 36
Table 3.5. Employment data characteristics 37
Table 3.6. Timing of recording for employment data 38
Annex Table 1. Contributors to MultiProd and DynEmp 47

Figures

Figure 1.1. Distributed microdata analysis 6

1. The distributed microdata approach

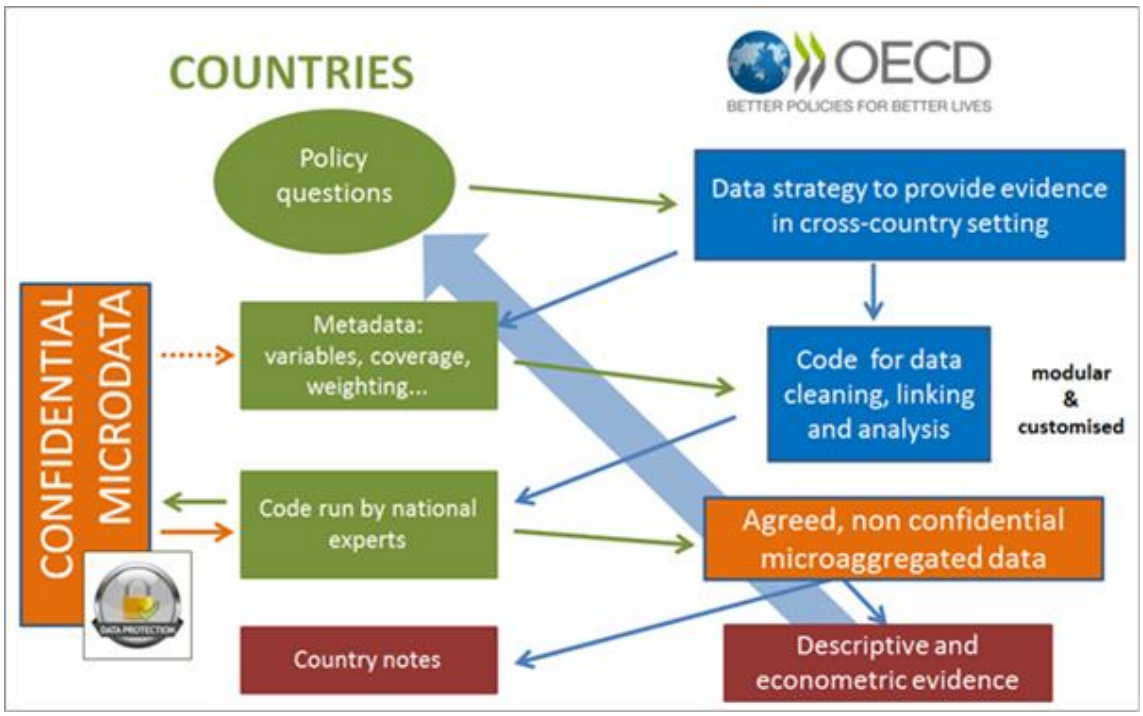
Given the substantial within-industry heterogeneity of productivity across firms, productivity analyses ought to go beyond the representative firm paradigm. Indeed the same average productivity may arise from different underlying productivity distributions, thus different economic patterns. For example, low average productivity may arise either from weak market selection (so there would be too many firms at the bottom of the distribution), or from a lack of innovation (in which case there would be too few firms at the top). Although characterised by a similar average productivity, these two situations have very different policy implications. In order to uncover the possible micro drivers of productivity growth and design the appropriate policies, it is therefore necessary to study the entire productivity distribution as well as investigate the heterogeneity of firms across multiple dimensions.

Moreover, the patterns of employment dynamics display substantial differences depending on firm characteristics and demographics. Net employment flows hide the concomitant creation and destruction of jobs by firms of different age and size, and they do not disentangle the contributions of entry, exit and post-entry growth. Yet aggregate employment growth typically depends on the contribution of a small number of firms with specific characteristics. Assessing the dynamism of the business environment therefore requires identifying the different channels of employment change for different types of firms.

Only microdata allows for an analysis of productivity patterns and employment dynamics at such fine level of granularity. Researchers and policy analysts have increasingly recognised the importance of microdata for understanding the complexity of economic dynamics and the heterogeneity in economic outcomes across firms. Yet confidentiality concerns and other administrative issues constitute cumbersome obstacles to the transnational access to official microdata.

A possible way to circumvent these obstacles is the so-called “distributed microdata approach”. Such decentralised method consists in collecting counts, totals and other statistical moments of the distribution of firm characteristics (employment, output, productivity, wages, age, etc.) by a centrally written, but locally executed, routine that is flexible and automated enough to run across different data sources in different countries. The advantages of this novel data collection methodology are manifold. It puts a lower burden on national statistical agencies and limits running costs for such endeavours. Importantly, it also overcomes the confidentiality constraints of directly using national micro-level statistical databases, while at the same time achieving a high degree of harmonisation and comparability across countries, sectors and over time (Figure 1.1). The OECD pioneered this method at the beginning of the 2000s, in collaboration with representatives in national statistical agencies or experts in public institutions who have access to the confidential microdata.

Figure 1.1. Distributed microdata analysis



This report relies on two distributed microdata projects carried out by the OECD Directorate for Science, Technology and Innovation: MultiProd and DynEmp[2]. The MultiProd project provides a comprehensive picture of productivity patterns across a large set of countries over the past two decades (Berlingieri et al., 2017[1]). The DynEmp project presents a complete picture of the dynamics of employment across firms with different characteristics such as size and age, as well as with respect to the entry, exit or incumbent status of firms (Criscuolo, Gal and Menon, 2014[2]). These projects use the distributed microdata approach to collect counts, totals and other statistical moments of the distribution of interest variables from national microdata sources. The resulting harmonised databases allow for the comparative analysis of productivity and employment patterns across countries and over time.

The paper proceeds as follows: Section 1 describes the MultiProd metadata, followed by a presentation of DynEmp metadata in Section 2. For each project, a first part gives an overall description of the project and its methodology, in particular regarding the basic requirements of the data, the procedures implemented by the program, and the output generated. A second part, focused on metadata, contains a number of tables summarising the main features of the data by country, and presents country-specific information.

Information is based on questionnaires, direct interaction with participants and actual features of the data directly included in the output databases.

2. The MultiProd project

2.1. Overall description

Productivity growth plays a central role in shaping the welfare of societies. The MultiProd project studies firm-level productivity patterns across countries, sectors and time. It contributes to better economic policies by studying how productivity at the firm level translates into aggregate productivity.

Increasingly available microdata unveil wide differences in firm-level productivity, even in narrowly defined industries. Comparable averages and other central tendency measures hide the potentially different nature of observed productivity patterns. For example, low average productivity may be due either to a lack of highly productive firms or to the survival of too many inefficient businesses, with contrasted policy implications.

The MultiProd project aims at overcoming this limitation by creating a harmonised cross-country micro-aggregated database on productivity patterns from confidential micro-level sources. Therefore, it contributes to the analysis of productivity by offering new evidence based on firm-level data. This enriches the policy debate on productivity by extending the analysis beyond aggregate industry performance to the important underlying dynamics and developments within industries. Thanks to information at the level of businesses, this new data source offers, for example, new insights on productivity dispersion within industries, its evolution over time and its structural and policy drivers. Policy makers can gain further insights from these measures of "inequality in firm performance" by means of an international comparison owing to the collection of harmonised, and therefore comparable, data. Each country can be then compared to a cross-country benchmark, which sheds light on the strengths and weaknesses of the economy in terms of productivity.

The first full version of MultiProd was released in September 2015 (MultiProd v1.0). For most countries, the document reflects the current version of the code, MultiProd v1.1, which collects additional variables and include updated deflators for a greater number of countries and years. The project is still at its first data collection round, but a new full version of the code will be released in 2019 followed by a new data collection.

2.2. Methodology

2.2.1. Basic requirements of the MultiProd input data

The data collected in MultiProd are computed by running a standardised STATA® routine on firm-level data. The program produces a set of statistics based on micro-level longitudinal information on output, inputs (labour and capital), labour costs, sector of activity, and possibly age and ownership of the firm. The information is used to calculate firm-level labour and multifactor productivity (MFP), which are then aggregated to the 2-digit sector level, separately for each year. The program also gathers statistics of the joint distribution of the latter variables with age, size, ownership (domestic and foreign), demographics (entrant, exiting, incumbent), productivity quantiles and gross output quantiles.

The MultiProd project relies on two main data sources at the country level: (i) production surveys, which exhaustively report firm-level usage of factors and intermediated inputs, as well as balance sheet information, but may only cover a sample of the population of firms;

and (ii) business registers, which typically contain less information, but cover the entire population. When necessary, MultiProd re-weights production surveys based on the population structure from business registers. For further details on the methodology, see (Berlingieri et al., 2017[1]).

The input data respect a number of basic requirements:

1. The data source must be longitudinal. When multiple data sources are available, the preferred one is the one that provides the best trade-off between coverage of units, time span and quality of information on inputs (employment, capital stock and intermediate inputs), output (gross output and value added) and labour costs.
2. Panel dimension: individual units need to be identified by a unique longitudinal identifier (id) that has to be constant over time. If the unit exits the production survey, its identifier must not be re-assigned to another firm.
3. Sectoral coverage: the aim is to cover the full non-financial market sector, especially manufacturing and non-financial market services. If available, the code collects information at the 2-digit level for all sectors of the economy.

The minimum set of required variables is:

- A unique firm identifier.
- A calendar year to which the time-varying variables refer to.
- A 3 or 4 digit sector identifying the main economic activity of the unit, following the ISIC (NACE) classification.
- Birth year of the unit.
- Firm employment in headcounts or full-time equivalents.
- Total labour costs of the firm.
- Gross output of the firm (can be sales if gross output is unavailable).
- Value added of the firm preferably at basic prices if available.
- Investment or capital stock (otherwise MFP cannot be calculated).

2.2.2. Measures of productivity

The MultiProd program computes a series of productivity measures, summarised in Table 2.1.

Table 2.1. Productivity measures used in MultiProd

Productivity measures		Definitions[i]
Labour productivity *(LP)*	Based on gross output: *LP_GO*	$\frac{\text{Real } gross\ output\ (or\ sales, GO)}{Employment\ (L)}$
	Based on value added: *LP_VA*	$\frac{Real\ value\ added\ (VA)}{Employment\ (\text{L})}$
Multi factor productivity (MFP): index-numbers[ii]	Solow-residual *MFP_S* with country and industry specific labour/intermediate shares (β_l^S, β_{ii}^S: median of shares across each year by country and industry)	$va - \beta_l^S l - (1-\beta_l^S)k$ or $go - \beta_l^S l - \beta_{ii}^S ii - (1-\beta_l^S - \beta_{ii}^S)k$
	Solow-residual *MFP_SW* with industry specific labour/intermediate shares ($\beta_l^{Sw}, \beta_{ii}^{Sw}$: median across countries and years by each industry)	$va - \beta_l^{Sw} l - (1-\beta_l^{Sw})k$ or $go - \beta_l^{Sw} l - \beta_{ii}^{Sw} ii - (1-\beta_l^{Sw} - \beta_{ii}^{Sw})k$
	Superlative index *MFP_SI*, following Griffith et al. (2009) [iii]	$\widetilde{va} - \beta_l^{SI}\tilde{l} - (1-\beta_l^{SI})\tilde{k}$ or $\widetilde{go} - \beta_l^{SI}\tilde{l} - \beta_{ii}^{SI}\tilde{ii} - (1-\beta_l^{SI} - \beta_{ii}^{SI})\tilde{k}$
	Superlative index MFP growth: *MFP_GR* [iv]	$\Delta va - \beta_l^{GR}\Delta l - (1-\beta_l^{GR})\Delta k$ or $\Delta go - \beta_l^{GR}\Delta l - \beta_{ii}^{GR}\Delta ii - (1-\beta_l^{GR} - \beta_{ii}^{GR})\Delta k$
Multi factor productivity (MFP): production function based[ii]	Residuals from Ordinary Least Squares estimates *MFP_OLS*	$va - \beta_l^{OLS}\, l - \beta_k^{OLS} k$ or $go - \beta_l^{OLS} l - \beta_{ii}^{OLS} ii - \beta_k^{OLS} k$
	Residuals from Ordinary Least Squares with firm fixed effect estimates *MFP_FE*	$va - \beta_l^{FE}\, l - \beta_k^{FE} k$ or $go - \beta_l^{FE} l - \beta_{ii}^{FE} ii - \beta_k^{FE} k$
	Residuals from Wooldridge (2009) type estimates *MFP_W*	$va - \beta_l^{W}\, l - \beta_k^{W} k$

Notes: All output and input variables are firm and year specific, but for simpler exposition, firm and time indices are omitted. Small letters (va, l, k, ii) denote natural logarithms of, respectively, Value added, Labour, Capital and Intermediates. For simpler exposition, MFP measures are defined in natural logarithms. Variables with a tilde on top ($\widetilde{va}, \tilde{l}, \tilde{k}$) indicate log-differences from the industry and year mean value. $\beta_l^{SI}, \beta_{ii}^{SI}$ denote the average of the labour and intermediate share in the reference firm (taken to be the geometric average across firms in the same country*industry) and the current firm. Variables with a Δ indicate log-growth. $\beta_l^{GR}, \beta_{ii}^{GR}$ denote the two-period average $\left(\text{e.g.: } \beta_l^{GR} = \frac{\beta_{l,t} + \beta_{l,t-1}}{2}\right)$ of the firm-level labour and intermediate share.

The gross output (GO) or total revenue based labour productivity (LP) aims at capturing the amount of output produced by a firm for a given amount of labour input. It is the crudest measure but potentially offers the highest coverage among all productivity variables. Note that even though gross output differs from total revenues in case of changes in inventories, both measures are admitted in MultiProd since their correlation tends to be very high at annual frequency. Of course, one of the main problems with a gross output based LP measure is that it does not control for intermediate input usage. As a result, a company with significant reselling activity will probably rank very high in this measure.

Value added (VA) based LP alleviates this problem, as value added itself is the difference between output (or sales) and intermediate inputs (including resold goods, typical in retail trade). Of course, neither type of LP controls for differences in capital intensity across firms

or differences in other inputs. Multifactor productivity (MFP) measures control for a wider set of inputs, using various methodologies. In MultiProd, two broad types of MFP measures are considered: index number approaches and the methods based on production function estimation.

MFP measures can take gross output or value added as the output variable, and capture both labour (measured by the number of employees or persons engaged), capital (measured by the total depreciated capital stock and/or by the book value of capital), and intermediates (when gross output is used, measured as total cost of intermediates) as inputs. The MultiProd program is flexible to produce both. Introducing more sophisticated measures either on the labour side (e.g. using the number of hours, workers with different skills) or on the capital side (e.g. using different types of assets, relying on capital services) would have put additional constraints in terms of country, sector and time coverage and comparability, as their availability is more limited. The choice of simpler but cruder input measures has been driven by the aim of collecting comparable statistics across countries while achieving the broadest possible coverage.

Although a number of MFP measures are implemented in the code, the current version of MultiProd is actually set to compute only two measures of MFP: an index-based measure (MFP_SW: Solow residuals using cross-country industry-specific labour and intermediate shares) and the semi-parametric measure of Wooldridge (MFP_W).

2.2.3. Data harmonisation procedures

Weighting

Since the data sources may cover only a subset of the population, MultiProd adopts a weighting strategy for the firm level samples coming from production surveys. This approach makes the results representative for the entire population of firms and comparable across countries. The strategy entails the following two main steps:

- Preparation of the population structure from the Business Register.
- Calculation of actual weights based on the population information contained in the business register. The weights are computed (at the 4-digit industry and eight size-classes level) as the number of firms in the population divided by the number of firms in the survey, after data cleaning.

At the time of writing, re-weighting is implemented only for Ireland, Italy and the Netherlands, where production surveys, rather than administrative data, have been used to run MultiProd. However, thankfully for these countries, business registers are available. Therefore, it has been possible to re-weight the sample to be as much as possible representative of the whole population of firms.

Deflation and PPP

MultiProd uses the OECD STructural ANalysis (STAN) database for obtaining deflators, depreciation rates, capital-labour ratios, as well as LP levels and growth rates at the country, industry year level. [3] The deflators for gross output, value added, intermediates, and gross fixed capital formation are available at the sectoral level, while we use the Consumer Price Index at the country-level for wages. All variables are deflated using their own deflator with the exception of capital and investment, which are deflated using the same deflator.

Every nominal monetary variable is transformed into real 2005 US dollars. To ensure international comparability, exchange rates and PPPs are therefore applied in that year. Since most of the manufacturing sector is usually traded internationally, the series for manufacturing are simply adjusted using the nominal exchange rate (as average over 2005), taken from Eurostat. On the other hand, all other sectors are adjusted using country-level PPPs for 2005 from the Eurostat-OECD PPP Programme.

Data cleaning

The program carries out a series of basic consistency checks and cleaning of the data:

- *Duplicates*. The program checks for duplicates in terms of firm identifier and year. In case duplicates are found, they are not treated and the program simply stops and warns the user.
- *Units of measurement*. A pop-up question at the very beginning of the code asks whether all monetary values are in local currency and single unit (not thousands).
- *Variables availability*. If key variables are not available in the production survey, the program checks whether they are in the business register; if so it takes them from there. The code also checks the availability of the variables used for MFP calculation and verify whether the robustness measures of MFP (using deflated capital stock) can be calculated. If neither capital nor investment is detected, the program warns the user because MFP cannot be calculated; if the user decides to proceed, the output of the program will be based on LP only.
- *Industry variables*. The program checks for the consistency of the industry variables. If the industry variable is available only in the old classification system, it asks whether it is NACE Rev.1.1 or ISIC Rev.3.1, because the external transformation file works only for those two systems. If the code detects that the industry classification is at two digits, it explicitly asks the user whether a more detailed industry classification is available.[4]
- *Birth*. In case the birth year of the firm is missing, the variable is set to be equal to the first year in which the firm reports positive employment (unless it is the first year in the database). The code also corrects for cases in which the reported birth year is more recent than the first year in which the firm reports positive employment. Finally, the code allows for left censoring, and the birth is set to missing in case it coincides with the year of censoring.
- *Implausible jumps*. The code drops observations with implausible jumps (consecutive large one-off increases and large one-off decreases or vice versa) in the main input variables of interest (employment, value added, total revenues, wage bill).
- *Missing or zeros*. The code drops any observation with missing, zero or negative employment. Moreover it sets FTE employment, wage bill, total revenue, and the capital stock (if available) to missing whenever they take zero or negative values.
- *Outlier filtering of final computed variables*. The code has also a routine to remove outliers from the main variables of interest. This applies only to internally computed variables (e.g. productivity growth) and is aimed at eliminating any outlier that might affect the results. The focus is on growth rates, hence on the within-firm rather than the cross-sectional variation. This is to avoid throwing out

correct and potentially useful information, for instance a very large firm would be dropped if one simply looked at the cross sectional variation in employment. In the baseline case the value is not corrected and the variable is set to missing in that year. For the specific case of productivity variables, the outlier filtering is also applied to the (log-) levels in order to eliminate outliers in the cross-section. This is done mainly to avoid outliers to be included in the group of firms at the frontier.

2.2.4. The MultiProd output

The MultiProd output is a collection of files that contain statistics for different variables computed yearly at a detailed sectoral level. The program is flexible to allow for multiple levels of aggregation at which the output is produced and, for each of those levels, the types of aggregated data to be included.

The statistics are never collected at the level of the individual firm. Instead, the program splits firms along various dimensions, into cells, and for each cell collects aggregate yearly data. In addition, these statistics are collected in both level and growth. The dimensions used for the split are the following, which can be specified at different levels of aggregation:

1. Sector (SNA A7 or A38 aggregation level).[5]
2. Firm-level productivity distribution: splitting firms into productivity quantiles.
3. Sales distribution: splitting firms into quantiles based on gross output.
4. Size class: splitting firms into groups based on employment levels.
5. Age of the firm: splitting firms into groups based on age.
6. Ownership: independent firm vs. affiliate of a business group, and nationality of the group.
7. Demographics: entrants, exiting, incumbents, etc.

The program further combines these dimensions to group firms in joint categories (e.g. the interaction of age and size classes). However, note that, although possible, the code never combines more than three dimensions at the same time for confidentiality reasons.

2.3. MultiProd metadata

2.3.1. Overview of the dataset

As of June 2019, 20 countries have been successfully included in the MultiProd database: Australia, Austria, Belgium, Canada, Chile, Denmark, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Sweden and Switzerland. Only partial results have been provided by Luxembourg and New Zealand. Other countries are currently running the code and will be included later on.

The data for each of those countries are collected at the firm-level (typically the legal unit), except Japan for data coming from the Census of Manufactures, where the source data are at the establishment-level but have been aggregated at the firm level.

Table 2.2 provides information on the national data sources underlying the final database, as well as the temporal coverage by country.

Table 2.2. National data sources

Country	Years	National data sources
Australia	2002-2012	Expanded Analytical Business Longitudinal Database (EABLD)
Austria	2008-2015	Structural Business Statistics (*Leistungs- und Strukturstatistik*)
Belgium	1998-2014	Central Balance Sheet Office, National Social Security Office and VAT data
Canada	2001-2012	Statistics Canada's National Accounts Longitudinal Microdata File (NALMF)
Chile	2005-2014	Chilean Tax office database (*Servicio de Impuestos Internos de Chile*)
Denmark	2000-2012	Accounts Statistics for Non-Agricultural Private Sector, General Enterprise Statistics
Finland	1995-2013	Structural business statistics data (surveys, corporate tax records, and Statistics Finland's Business Register)
France	1995-2015	FICUS/FARE, DADS, LIFI, SIRUS (*Contours des entreprises profilées*)
Germany	2003-2013	Four production surveys (JBU, KSE, IEU and AFiD-Panel Structure survey in the services sector)
Hungary	1998-2012	Corporate Income Tax data (CIT) of National Tax and Custom Administration
Ireland	2006-2014	CIP ("Census for Industrial production") for manufacturing, ASI ("Annual Services Inquiry") for services, Business Register
Italy	2001-2015	*Bilanci società di capitali con dipendenti,* ASIA (Business Register), *Indagine sulle grandi imprese* (SCI), Database *Commercio Estero* (COE)
Japan1	1994-2014	Basic Survey of Japanese Business Structure and Activities (firm-level data)
Japan2	1994-2014	Census of Manufactures (establishment-level data aggregated at firm level)
Luxembourg	2003-2012	*Répertoire des entreprises* (Business Register) and Structural Business Statistics (SBS) data
Netherlands	2000-2015	*Productiestatistieken* (PS) and *Algemeen Bedrijvenregister* (BR)
New Zealand	2000-2011	Longitudinal Business Database (LBD) and Integrated Data Infrastructure (IDI)
Norway	1995-2012	Accounts statistics (incorporated firms) and Establishment population register
Portugal	2004-2012	Integrated Business Accounts System (IBAS)
Sweden	2002-2012	SBS administrative data (tax data)
Switzerland	2009-2012	Production survey (*Wertschöpfungsstatistik*)

Notes: In Australia, fiscal years are from the 1st of July to the end of June the following year. Hence, for instance, the fiscal year 2002 indicates the 2001–02 period. In New Zealand, fiscal years are industry specific; however, for most firms, the data reported for 2001 refers to the period from 1 April 2001 to 31 March 2002. In Japan, fiscal years are years ending on March 31st, e.g. the fiscal year 2012 refers to the calendar year April 1, 2012 to March 31, 2013.

Access to French data used in this work benefited from the use of Centre d'Accès Sécurisé aux Données (CASD), which is part of the "*Investissements d'Avenir*" programme (reference: ANR-10-EQPX-17) and supported by a public grant overseen by the French National Research Agency (ANR).

Access to the anonymised data for New Zealand was provided by Statistics New Zealand in accordance with security and confidentiality provisions of the Statistics Act 1975 and secrecy provisions of the Tax Administration Act 1994. The results in this report have been confidentialised to protect individual businesses from identification. See Appendix for the full disclaimer.

Japan has two different data sources. The Basic Survey of Japanese Business Structure and Activities covers most industrial sectors but only refers to firms with 50 or more employees. The Census of Manufactures has a broader population coverage in terms of employees, but refers to the manufacturing sector only. See the country-specific information for details.

With respect to sectoral aggregation, the list of STAN A7 (macro-sectors) and A38 industries used as a baseline is given in Table 2.3. MultiProd typically covers the whole economy with the exceptions of the financial sector (ISIC Rev. 4 64-66), the public sector (ISIC 84), and households and other activities (ISIC 97-99). However, the analysis conducted so far focuses on manufacturing and non-financial market services in order to enhance cross-country comparability.

Table 2.3. Sectoral aggregation

SNA A38 aggregation based on ISIC Rev.4 classification	SNA A7 aggregation
01 to 03 AGRICULTURE, FORESTRY AND FISHING	Agriculture
05 to 09 Mining and quarrying	Mining
10 to 12 Food products, beverages and tobacco	Manufacturing
13 to 15 Textiles, wearing apparel, leather and related products	
16 to 18 Wood and paper products, and printing	
19 Coke and refined petroleum products	
20 Chemicals and chemical products	
21 Basic pharmaceutical products and pharmaceutical preparations	
22 to 23 Rubber and plastics products, and other non-metallic mineral products	
24 to 25 Basic metals and fabricated metal products, except machinery and equipment	
26 Computer, electronic and optical products	
27 Electrical equipment	
28 Machinery and equipment n.e.c.	
29 to 30 Transport equipment	
31 to 33 Furniture; other manufacturing; repair and installation of machinery and equipment	
35 Electricity, gas, steam and air conditioning supply	Utilities
36 to 39 Water supply; sewerage, waste management and remediation activities	
41 to 43 CONSTRUCTION	Construction
45 to 47 Wholesale and retail trade, repair of motor vehicles and motorcycles	Market services
49 to 53 Transportation and storage	
55 to 56 Accommodation and food service activities	
58 to 60 Publishing, audiovisual and broadcasting activities	
61 Telecommunications	
62 to 63 IT and other information services	
64 to 66 FINANCIAL AND INSURANCE ACTIVITIES	*Excluded*
68 REAL ESTATE ACTIVITIES	Market services (continued)
69 to 71 Legal and accounting activities; activities of head offices; management consultancy activities; architecture and engineering activities; technical testing and analysis	
72 Scientific research and development	
73 to 75 Advertising and market research; other professional, scientific and technical activities; veterinary activities	
77 to 82 Administrative and support service activities	
84 Public administration and defence	*Excluded*
85 Education	Non-market services
86 to 88 Human health and social work activities	
90 to 93 Arts, entertainment and recreation	
94 to 96 Other service activities	

Table 2.4 provides details on sectoral classifications and coverage of the source data in each country, with a few other country-specific exceptions in terms of sectoral coverage. The very few excluded industries typically do not belong to manufacturing or non-financial market services, the only sectors taken into account so far.

Table 2.4. Sectoral classifications and coverage

Country	Classification followed	ISIC Rev.4 industries covered[6]
Australia	ANZSIC2006 from 2007	All
Austria	(Ö)NACE 2008 (NACE Rev.2)	All except 1, 85-90
Belgium	NACE-BEL 2008	All
Canada	North American Industry Classification System (NAICS)	All
Chile	CIIU4.CL 2012 (based on ISIC)	All
Denmark	NACE Rev.2	All except 35 for years 2005-06 and 2008-12, 85 to 90 for years 2008-12
Finland	NACE Rev.1.1 and 2	All
France	Nomenclature d'Activités Française (NAF) Rev.1 and 2	All
Germany	WZ03 and WZ08 (NACE Rev.1.1 and 2)	Sections B, C, H, J, M, N and division S95
Hungary	TEÁOR 03 and 08 (NACE Rev.1.1 and 2)	All except 12, 19
Ireland	NACE Rev.2	All except 1-3, 5-9, 86-88, 94
Italy	Ateco Rev.1.1 and Rev.2	All except 1
Japan1	Japan Standard Industrial Classification (JSIC), Rev.10 to 13	All except 1-3, 36-39, 41-43, 49-53, 61, 68, 85-88 from 2007; less industries covered before 2007
Japan2	Japan Standard Industrial Classification (JSIC), Rev.10 to 13	Manufacturing industries
Luxembourg	NACE Rev.1.1 and 2	All except 1-3, 85,, 86-88, 90 to 93, 94 and 96
Netherlands	NACE Rev.1.1 and 2	All except 1, 5
New Zealand	ANZSIC2006	All
Norway	NACE Rev.1.1 and 2	All
Portugal	NACE Rev.2	All
Sweden	NACE Rev.1.1 and 2	All
Switzerland	NOGA 2008 (NACE Rev.2)	All except 1-3

Table 2.5 indicates if the variable used as labour input covers the number of persons engaged or only salaried employees. “Employees” are defined as all persons who, by agreement, work for another resident institutional unit and receive remuneration. “Persons engaged” are instead defined as the total number of persons who work in or for the establishment, including working proprietors, active business partners and unpaid family workers, as well as persons working outside the establishment when paid by and under the control of the establishment. As shown in the table, for the majority of countries we have information on employees or both, the only exception being Finland (person engaged expressed in full time equivalent, where firms with less than one person engaged have been dropped from the sample), Germany and the Netherlands. The distinction between the variables is relevant because it matters, for instance, when computing MFP and wages: when computing MFP ideally we would like keep into account all persons involved in the production process, i.e., “persons engaged”; when computing wages as the ratio between cost of labour and employment, instead, we would like to keep only salaried workers, i.e., “employees”. Therefore, whenever available, both measures are exploited for different purposes.

Table 2.5. Labour input variables

Country	Persons engaged	Employees
Australia		✓
Austria	✓	✓
Belgium		✓
Canada		✓
Chile		✓
Denmark		✓
Finland	✓	
France		✓
Germany	✓	
Hungary		✓
Ireland	✓	✓
Italy	✓	✓
Japan1		✓
Japan2	✓	
Luxembourg		✓
Netherlands	✓	
New Zealand		✓
Norway		✓
Portugal	✓	
Sweden		✓
Switzerland		✓

Table 2.6 focuses on inputs, detailing their availability. If information on investment is available, the capital input is measured through the Perpetual Inventory Method (PIM); otherwise, the book value of capital is used. Moreover, the labour input can be expressed in head counts (HC) or full-time equivalent (FTE).

Table 2.6. Recording units for employment and capital, and availability of intermediate inputs

Country	Employment		Capital		Intermediates
	HC	FTE	PIM	Book value	
Australia		✓	✓	✓	✓
Austria	✓	✓	✓		✓
Belgium	✓	✓	✓		✓
Canada	✓			✓	✓
Chile	✓		✓		✓
Denmark	✓	✓	✓		✓
Finland		✓	✓		✓
France	✓	✓	✓	✓	✓
Germany	✓		✓		✓
Hungary	✓			✓	✓
Ireland	✓	✓	✓		✓
Italy	✓			✓	✓
Japan1	✓		✓	✓	✓
Japan2	✓		✓	✓	✓
Luxembourg	✓		✓		✓
Netherlands		✓		✓	✓
New Zealand	✓			✓	✓
Norway	✓			✓	✓
Portugal	✓		✓		✓
Sweden		✓	✓		✓
Switzerland	✓	✓			✓

Note: For Austria, FTE data are available for employees only. For Ireland, employment data are expressed in head counts up to 2007 and in full-time equivalents from 2008 onwards. For the Netherlands, depreciation cost is used rather than book value.

As a baseline, MultiProd covers businesses with at least one employee. There may be also a revenue threshold for some countries. Table 2.7 summarises the thresholds used in each country for inclusion in the sampling frame (in case the data comes from a production survey). The second column details the threshold above which the entire population of firms is included in the database.

Table 2.7. Thresholds in the production surveys

Country	Threshold used for inclusion in the sampling frame	Census threshold	Presence of business register
Australia	1 employment unit	To be confirmed	Yes
Austria	1 employee or 300k in revenues	10 or 20 employees depending on the sector. Higher size (or more flexible) threshold for firms to answer the survey from reference year 2013.	Yes
Belgium	1 employment unit	1 employment unit	Yes
Canada	1 employment unit	1 employment unit	Not needed
Chile	1 employment unit	1 employment unit	Not needed
Denmark	1 employment unit	All firms with more than 50 employees are included	Not needed
Finland	1 employment unit	1 employment unit	Yes
France	1 employment unit	1 employment unit	Not needed
Germany	Manufacturing: 20 employees Services: EUR 17 500	Manufacturing: 20 employees for the annual report of companies and the investment survey; 500 employees for the cost structure survey	No
Hungary	Since 2004, firms reporting at least HUF 50 million turnover in two consecutive years are included.	Since 2004, firms reporting at least HUF 50 million turnover in two consecutive years are included.	Not needed
Ireland	CIP: 3 employees up to 2007, no threshold from 2008 onwards ASI: 2 persons engaged	CIP: all firms with at least 10 employees are surveyed up to 2007, at least 3 persons engaged from 2008 onwards ASI: all firms with more than 20 employees are surveyed	Yes
Italy	1 employee (limited companies)	1 employee (limited companies)	Yes
Japan1	50 employees and paid-up capital or invested funds of at least JPY 30 million Firm-level data	50 employees and paid-up capital or invested funds of at least JPY 30 million Firm-level data	No
Japan2	All establishments in the manufacturing sector with 4 or more employees; up to 2008, the universe of establishments in the manufacturing sector in years ending with 0, 3, 5 and 8 were surveyed. Data then aggregated at firm level.	All establishments in the manufacturing sector with 4 or more employees; up to 2008, the universe of establishments in the manufacturing sector in years ending with 0, 3, 5 and 8 were surveyed. Data then aggregated at firm level.	No
Luxembourg	Only active market enterprises are included (units with non-zero value of turnover or non-zero employees, units active according to the SBS but not properly recorded in the administrative sources).	45 employees or EUR 7 million annual turnover are exhaustively selected. A sample of smaller enterprises is drawn randomly by strata.	Yes
Netherlands	10 employees	50 persons employed up to 2009; from 2010, there is a limited number of larger activities codes for which only firms with more than 100 persons employed are census.	Yes
New Zealand	Mandatory tax filing threshold of NZD 40 000 (rising to NZD 60 000 at end of period), or 1 employee	1 employment unit	Not needed
Norway	1 employment unit for enterprises (limited dependent companies)	1 employment unit for enterprises (limited dependent companies)	Not needed
Portugal	1 employment unit	1 employment unit	Not needed
Sweden	1 employment unit	1 employment unit	Not needed
Switzerland	3 persons including the owner	50 employees	No

Note: The census threshold is the threshold above which the universe of firms is reported in the production survey. Information on employment unit can be found in Table 2.6.

2.3.2. Country-specific information

Australia

The data source is the Business Longitudinal Analysis Data Environment (BLADE)[7] from the Australian Bureau of Statistics (ABS). It integrates financial and business characteristics data for all active businesses in the Australian economy.

The database is a combination of business register data and administrative records businesses supply to the Australian Taxation Office (ATO): Business Income Tax (BIT) data, Business Activity Statement (BAS), and Pay-As-You-Go (PAYG) data, i.e. Individual Payment Summaries. Business Register information is used to reconcile the reporting structures of the ATO and the ABS. A series of surveys conducted by the ABS, mainly the Business Characteristic Survey and the Economic Activity Survey, are also available as supplements to the core body.

Type of Activity Units (TAUs) are the official ABS statistical unit. For simple units - typically small to medium sized firms - the Australian Business Number (ABN) is equivalent to a TAU. For complex units – typically large enterprises – TAUs are created in line with the ABS units model and often align with company reporting lines. Thus, a TAU is similar to a firm that is operating under an Enterprise Group. TAUs can operate in different geographical areas, and state of operation is based on where the majority of employment occurs. Industry codes – such as ISIC – and institutional codes that distinguish among basic sector of the economy – for example, households, government, non-financial corporations etc. – are assigned to TAUs and form the basis for ABS estimates of production, etc. by industry and sector.

Employment that does not result in the creation of an annual payment summary is not included. Statutory head count measures are available that simply sum each scrambled tax file number receiving wage payments. An estimate measure of full-time equivalents using these records is also available.

Austria

The data source is the Structural Business Statistics (*Leistungs- und Strukturstatistik*) from Statistics Austria. The population covers enterprises (legal units) with main economic activity in NACE sections B to N and division S95 and that are economically active in the respective reference year in the Business Register, i.e. that have a yearly turnover of more than EUR 10 000 and/or one person employed.

The SBS survey is mandatory for firms above certain industry-specific revenue and size thresholds, ranging from EUR 300 000 to EUR 3 million and from 10 to 20 employed persons. All firms above these thresholds are included in the survey, which amounts to approximately 35 000 firms per reporting year and corresponds to 10.5% of the population in 2013, covering 87% of turnover and 75% of employees. Although some firms below these thresholds are included in the survey using administrative data and model-based estimation, they are not included in MultiProd since these firms are not randomly sampled, but only included if they exceed either the revenue or the size threshold. This implies for example that the only small firms (i.e. with less than ten employees) that are included in the data are on average more productive since they have revenues larger than the revenue threshold. Re-weighting the sample would therefore not make it representative of the whole population.

The implementation of an amended national SBS Regulation resulted in an increase of the turnover thresholds and an alternative employment threshold for the service sector (from the reporting year 2013). Flexible reporting thresholds were generally introduced from the reporting year 2014.

The Business Register is used for the variables birth year and foreign, all other variables come from the SBS survey. Data for employees primary come from the social security data; self-employed from a mixture of sources (again social security, but also liberal professions); finally, turnover and gross wages are originated mainly from primary survey. Other data necessary for value added, such as purchases of goods/services, are surveyed as well.

The book value of capital is unavailable and, hence, the variable Capital is computed through the PIM method, using as starting values statistics coming from the OECD STAN database. However, since the period considered is relatively short, this methodology could lead to biased estimations of capital.

Belgium

The data sources for Belgium consist of:

- Balance sheet data from the Central Balance Sheet Office of the National Bank of Belgium (NBB);
- VAT data from the Federal Public Service Finance for turnover, intermediate inputs and investment of small firms not belonging to a VAT unit;
- The register of the National Social Security Office (NSSO) for employment data.

The balance sheet data cover all Belgian companies with at least one employee, as they are required to file annual accounts with the NBB. There are two types of annual accounts: full and abbreviated. Firms have to file full annual accounts when they exceed at least two of the three following thresholds: (i) 50 employees; (ii) EUR 7.3 million in turnover; and (iii) EUR 3.65 million in total assets. Firms that are not required to file full accounts do not have to report turnover, investments and intermediate inputs; therefore, these data items come from VAT declarations. However, if those firms belong to a VAT unit, the information on turnover, investments and intermediate inputs is consolidated at the unit level and missing at the firm level.

All employers in Belgium are required to report staff movement to the NSSO, so all social security agencies receive an electronic notification every time an employee is hired by or leaves an employer. In addition to data on changes in employment, the NSSO also records other information, such as wage data.

Employment data refer to the number of employees. Temporary workers are not included, nor non-employer businesses.

Canada

The data source is the National Accounts Longitudinal Microdata File (NALMF) from Statistics Canada, which combines the tax data sources below from the Canadian Revenue Agency based on a Business Register to produce a final linked file at the enterprise level:

- Statements of Remuneration Paid (T4 slips) issued by employers, both incorporated and unincorporated, to employees for income tax purposes;

- Corporation Income Tax Returns (T2) provided by incorporated businesses when they file taxes;
- Goods and Services Tax (GST) remittances that are mandatory for all businesses with annual revenues above CAD 30 000;
- Payroll Account Deductions (PD7) that businesses are required to file concerning employment deductions (e.g. for pensions or employment insurance), which serves as the basis for the employment data from the Survey of Employment, Payrolls and Hours.

The capital stock measure relies on data on tangible assets.

Chile

The data source is the Chilean Tax office database (*Servicio de Impuestos Internos de Chile*, SII). The SII dataset combines information from four different tax forms filled by each Chilean company on behalf of its employees:

- Form 22 for sales and net income of companies that can be subject to capital taxation, based on current or accrued revenues (annual information).
- Form 29 for sales, investment, value added and intermediates (monthly information). This form also shows the industrial sector of each firm according to the ISIC.
- Form 1887 for employees and labour costs (annual information). While the statement is annual, it contains information on the monthly payroll of firms, allowing to identify, for any given month, the employment status of an individual worker.
- Form 1879 for self-employed and labour costs (annual information of remunerations of professionals or other lucrative activity, directors or advisors of corporations).

Denmark

The main data source is the Accounts Statistics for Non-Agricultural Private Sector from Statistics Denmark. This is merged with the General Enterprise Statistics to get more detailed information about employment for instance.

The primary data sources of the Accounts Statistics for Non-Agricultural Private Sector are:

- Questionnaires sent to 7 500 business enterprises; stratified results are calculated for this population of units;
- Pharmacy accounts from the Drugs Administration Agency (230 pharmacies);
- The Central Customs and Tax Administration (92 500 enterprises);
- The Business Register (this information only covers turnover and employment plus background information such as enterprise code number, kind of activity, name and address).

Only active enterprises with an annual performance corresponding to at least half a year's work for one person or a certain sector-specific turnover are included. Firms with more than 50 persons engaged are a census.

The General Enterprise Statistics covers all enterprises in all sectors with information about the number of enterprises and their employees. In addition, it contains accounting information and information about turnover and exports for most of the activities. The statistics integrates information from two other business statistics, which are compiled for different unit types. The two statistics comprise the Accounts statistics, which are compiled at the enterprise level, and the VAT statistics, which are compiled at the administrative level used by the Danish Central Customs and Tax Administration.

The employment variable is expressed in full-time equivalents and in head counts in the Accounts Statistics for Non-Agricultural Private Sector and in the General Enterprise Statistics.

Finland

The MultiProd data are aggregated from the Structural Business Statistics data that, in principle, cover exhaustively all firms in the Finnish business sector. The Structural Business Statistics data are collected directly from firms, that typically employ at least 20 persons, through surveys and then complemented by exploiting the Tax Administration's corporate taxation records and Statistics Finland's Business Register that has near-perfect coverage.

Firms with less than one person engaged have been dropped from the sample. Finland uses full-time equivalent units as labour input, so that these dropped firms include firms run by a half-time owner, for example; from the data, firms that have less than one person cover 57% of the total population. As a result, the MultiProd data cover the entire population of firms with at least one person engaged. Moreover, the data includes only firms that produce positive value added, which is needed to measure the log of labour productivity.

France

The dataset used for MultiProd combines information from five main data sources:

- *Système Unifié de Statistiques d'Entreprises* (1995-2007);
- *Fichier Approché des Résultats ESANE* (2008-15);
- *Enquête sur les Liaisons Financières entre Sociétés*;
- *Déclaration Annuelle de Données Sociales*;
- SIRUS (*Contours des Entreprises Profilées*).

The FICUS (*Fichier Complet Unifié de SUSE*) dataset covers all private sector firms in France, as recorded on the SUSE (*Système Unifié de Statistiques d'Entreprises*). Since 2008, the dataset has been replaced by the FARE (*Fichier Approché des Résultats ESANE*).[8] Both datasets combine information retrieved from the annual statements of enterprises supplied by the General Tax Directorate (Industrial and Commercial Profits System, Non-Commercial Profits system and Agricultural Profits) with results from the annual sectoral surveys (EAE, ESA) carried out by INSEE. Among other characteristics, they provide information on turnover, employment, value added, investments, the 4-digit sector each firm belongs to and geographic locations.

The LIFI (*Liaisons Financières entre Sociétés*) survey aims to identify and outline the financial linkages across enterprises to reconstruct the business groups operating in France. A business group is defined based on capital holding links between them on 31 December. To guarantee consistency in the data, only groups with more than 50 employees (the size of a medium firm) are considered.

The DADS (*Déclaration Annuelle de Données Sociales*) is constructed by the French National Institute of Statistics and Economic Studies (INSEE) from the compulsory social security declarations made annually by all employers for each of their employees. The dataset covers the whole population of private sector employees, providing information on wages and occupational classification that can be aggregated at the firm level. For the purpose of this analysis, we use data manually re-aggregated at the firm level, retrieved from the DADS-Postes dataset.

The *Système d'Identification au Répertoire des Unités Statistiques* (SIRUS) defines an enterprise as the smallest combination of legal units that constitutes an organisational unit for the production of goods and services, with a certain autonomy of decision, notably for the allocation of its current resources. This dataset allows to aggregate single legal units to the enterprise level, as defined by Eurostat. Since 2001, at different moment in time, a certain number of legal units linked to five large companies (Renault, Peugeot, Accor, Saint Gobain and SEB, called *groupes historiques*) have been aggregated to the group head in FICUS/FARE. For the aim of this project, the *Contours des entreprises profilées* dataset is used to aggregate the *groupes historiques* in the years before the *profilage* and to keep them up-to-date over time. Starting from 2012, the INSEE has started extending the work beyond the *groupes historiques* and carried out a larger number of consolidations at the enterprise level (around 100 enterprises per year). Given the lack of information prior the consolidation and the changing number of targeted enterprises, the dataset has been kept at the legal unit level throughout the entire period in order to guarantee consistency over time.

There is a break in series in 2008 due to a better coverage of small firms, the change of industrial classification from NAF Rev.1 to Rev.2, and the introduction of a new production process for the Structural Business statistics. The break has been minimised thanks to the use of DADS for all the employment related variables, as well as the selection of all firms with at least one employee.

Germany

The MultiProd dataset combines information from several surveys administered by the German statistical office. Regarding the mining and manufacturing sectors, the data sources are the following:

- The annual report of companies from the manufacturing, mining, and quarrying sectors (JBU). Information is available in the form of annual results for all German firms employing at least 20 persons and operating in the corresponding industries. Participation of firms in the survey is mandatory.
- The cost structure survey in the manufacturing, mining and quarrying sectors (KSE), which is a production survey based on a sample of firms with 20 or more persons engaged. All businesses with 500 or more employees are included in the population. The sample of smaller companies with 20 to 499 employees is modified every four years.

- The investment survey in the manufacturing, mining and quarrying sectors (IEU) that includes all enterprises in the sectors concerned having at least 20 employees. Investment is used to construct the variable Capital using the PIM.

The data source for services is the AFiD-Panel Structure survey in the services sector, covering firms in NACE sections H, J, M, N and division S95. Firms are surveyed if their annual turnover is higher than EUR 17 500. New samples were drawn in 2003, 2008, 2011 and 2014.

Hungary

The data source for Hungary is the corporate tax database of the National Tax and Customs Administration (NAV). It contains the financial statements of companies with double-entry accounting.

Before the accession of Hungary to the EU, a number of small enterprises were allowed to use single-entry book keeping and, hence, were not included in the database. The criteria to file a double book keeping tax report has been lowered; since 2004, firms reporting at least HUF 50 million (around EUR 200 000) turnover in two consecutive years are included.

Tobacco and oil industries are not taken into account in the dataset, since from 2001 there have been changes in accounting rules that prevent the correct computation of value added and gross output.

Ireland

The data sources are the Census of Industrial Production (CIP), the Annual Services Inquiry (ASI) and the Business Register (BR) from the Central Statistics Office.

The CIP contains data on firms in the manufacturing, utilities, mining and quarrying industries, while the ASI covers firms in market and non-market services, excluding financial services (NACE Rev.2 sectors G, H, I, J, L, M, N and divisions 92, 93, 95 and 96). Up to and including 2007, the CIP surveys all enterprises with ten or more employees and a portion of enterprises with between three and nine employees; from 2008 onwards, all industrial enterprises with more than three persons engaged are included. The ASI surveys all firms with more than 20 employees and uses a stratified random sample to survey enterprises reporting between 2 and 20 persons engaged.

The CSO has recently revised the methodology for the compilation of structural business statistics (including CIP, ASI and BR), with the new approach applied back to 2008. Accordingly, the final sample in the analysis, which is comprised of a panel of firms spanning from 2006 to 2014, required a number of transformations to ensure consistency in the series.

The mining and quarrying sector was dropped from the CIP dataset as it contained less than the minimum number of observations MultiProd requires to run.

To give an idea of the coverage of the sample provided, around 4% of the entire population of firms as reported in the BR are represented by the combined CIP+ASI sample each year, with these firms accounting for around 40% of employment reported in the BR. This sample is then augmented by MultiProd with the entire population of firms from the BR to re-weight the sample of surveyed enterprises into a representative sample for estimating key output statistics.

For MFP estimation, firm-level capital stocks are calculated based on the perpetual inventory method (PIM) using annual firm level investment. The PIM relies on an opening capital value for each firm, annual investment and depreciation.

Italy

The data source used for MultiProd is the financial statements of limited companies with employees (*Bilanci società di capitali con dipendenti*) integrated with the SBS Frame from 2011. It is a panel dataset that contains a representative sample of the population of Italian limited corporations operating in the sectors of mining, manufacturing, utilities and private non-financial services from 2001. It integrates also the information from the Asia business register, the survey on large enterprises (SCI), and foreign trade data (COE). The panel is composed of the universe of the approximately 400 000 limited companies having at least one employee (over approximately 1.2 million of firms with employees and 4.3 millions of firms in total). Since the financial statements are not available for the whole population, the results are re-weighted using the whole population of firms from the Asia register.

The Asia register (*Archivio statistico delle imprese attive in Italia*) covers all sectors except agriculture and finance. Total employment in Asia is given by the sum of independent and dependent workers. In 2017, independent workers represented 23.2% of the total employment. They are mainly employers (about 28% of independent workers in 2017) and self-employed with no employees (about 72% of independent workers in 2017). Until 2009, firms with total employment smaller than one used to be classified as having one worker, in order to have at least one worker per firm. From 2010, however, firms with total employment smaller than 0.5 have started to be reported as having zero employment (i.e., correctly rounded). In order to have a consistent dataset for the entire period, from 2010 onwards it has been replicated a modification similar to that applied up to 2009 (firms with less than one employment unit rounded up at 1). Given that the Asia business register has been used to re-weight the previous dataset, which includes only firms with at least one employee, all firms with one person engaged have been excluded from the final dataset. The underlying assumption is that firms with one person engaged are very likely to have zero employee, with the only person included being an independent worker. The only cases not taken into consideration when applying this correction are quite rare: i) cases in which there is one employee and zero independent workers (which are going to be dropped from the BR and not in the PS); ii) cases in which there is more than one independent worker but zero employees (kept in the BR but not available in the PS).

The SCI survey (*Sistema dei Conti delle Imprese*) annually reports detailed information on the income statement and on the balance sheet, on employment, on labour costs, on investments and on other relevant characteristics of all Italian companies with 100 employees and more (20 employees prior to 2008). Companies with 200 and more workers carrying out a significant secondary activity are sent additional ad hoc questionnaires in order to collect the separate information for the various economic activities in which the Company operates. The survey reported to the year 2013 involved 10 527 companies in the industry and services with 100 employees and beyond, present in the Reference Archive (Asia). The SCI survey concerns enterprises that operate in all economic sectors except agriculture, finance and public administration.

In Italy, two different variables for persons engaged and employees are available: therefore, the productivity measures are computed adopting as employment variable persons engaged, whereas wages are computed as total cost of labour (for paid workers) over number of employees.

Before 2007, Italian firms were allowed to deduct the accelerated depreciation from taxes. The tax reform introduced in 2007 abolished this possibility. This might cause a jump up in the measure of capital between 2007 and 2008.

Japan – first data source

Data come from the Basic Survey of Japanese Business Structure and Activities carried out by the Ministry of Economy, Trade and Industry (METI). Only enterprises with 50 or more employees and with paid-up capital or invested funds of at least JPY 30 million are included. From 1994 to 1997, only firms in the Manufacturing, Mining and Wholesale and retail industry were covered, however the coverage of the non-manufacturing sector has progressively expanded over time. Since 2007, most of the non-financial market services sector is covered.

Between 2005 and 2006, the survey question concerning the total cost of labour changed. Until 2005, the variable used as total cost of labour was including total wage payments, annual bonuses and retirement allowance; however, no information on social insurance payments and expenses for employees' recreation activities was available. From 2006 onward, this variable changed: it started to include only total wage payment and annual bonuses, but not anymore retirement allowance. However, a new item has been added, called "welfare allowance", including retirement allowance and various social insurance payments and expenses for employees' recreation activities. In order to make the series more consistent, those two pieces of information have been summed. Therefore, from 2006, the variable used as total cost of labour includes total wage payments, annual bonuses, retirement allowance and social insurance payments. Hence, the cost of labour series has a jump in 2006, due to the fact that from that year onward it includes also social insurance payments. To deal with this issue, the variable used for labour cost used before 2006 has increased of a 15%, in order to impute social insurance.

Japan – second data source

Data come from the Census of Manufactures carried out by the Ministry of Economy, Trade and Industry (METI).

The Census of Manufactures is conducted on all establishments in the manufacturing sector with four or more employees; up to 2008, the universe of establishments in the manufacturing sector in years ending with 0, 3, 5 and 8 were surveyed. A long questionnaire is sent to establishments with 30 or more employees, while a shorter version is for those with 29 or fewer employees, in both cases excluding head offices that are not involved in the manufacturing processing or repair of industrial products. The data on establishments have been aggregated at the firm level.

In 2007, the Census frame was reviewed in comparison with the list of manufacturing establishments from the Establishment and Enterprise Census. In 2011, the Census of Manufactures was not conducted, and the necessary items were surveyed by the Economic Census for Business Activity in 2012.

The Census provides the number of persons employed as a total number of sole proprietors and unpaid family workers and regular workers at the end of the calendar year. It excludes temporary workers, i.e. persons that are employed on agreements shorter than a month or on daily agreements. In MultiProd, the employment variable is the number of persons engaged in head counts.

Information on capital variables such as investment and the book value of capital is not included for the years 2001 to 2004, 2006 to 2010 and 2012 to 2014 for plants with less than 30 employees.

Luxembourg

The dataset for Luxembourg is provided by STATEC (National Institute of statistics and economic studies). It is sourced from the Structural Business Statistics, complemented with data from the *Répertoire des Entreprises* (Business Register). The population covers enterprises (legal units) with main economic activity in NACE Rev.2 sections B to N (excluding K) and division S95.

Enterprises with more than 45 employees or an annual turnover (excluding VAT) of more than EUR 7 million, as well as any legal units legally related to one of these enterprises, are exhaustively selected. A sample of smaller enterprises, which do not meet the above criteria and which have a turnover of at least EUR 100 000, is drawn randomly by strata and according to a principle of rotation. As of 2015, an additional source is used to impute units which are not part of the sample or for which no recent survey data are available.

The Business Register covers all firms except sole proprietorships without employees and subject to less than EUR 10 000 in VAT.

Netherlands

The data source is the Production Statistics (*Productiestatistieken*) from Statistics Netherlands, which in the version accessed by MultiProd is restricted to enterprises with ten or more employees. Enterprises with less than ten employees are a selected sample that cannot be made representative since after 2004 these enterprises are surveyed only in certain activities and benchmark years. They are therefore excluded from the MultiProd sample.

The Production Statistics is a stratified sample survey whose target population is determined using the Business Register (*Algemeen Bedrijvenregister*). The sampling design has the following characteristics:

- The sample design includes all firm that existed in year t including new firms but not firms that have exited (by contrast the population includes exits as well; before 2001 those were included in the sample as well but were dropped as a consequence of low response). The sample is drawn on December 1st; de facto this sample consists of firms that are alive on this date.
- Firms with more than 50 persons employed are a census, below this threshold they are sampled. After the sample redesign of 2010, there is a limited number of larger activities codes for which only firms with more than 100 persons employed are a census.

In 2006, there has been a redesign of the BR, concerning the derivation of the population and population units. In 2009, a transition to a new enterprise definition was implemented, by which some units were consolidated into a single unit. This improved delineation of enterprise groups aims at increasing the linkability of the BR with external administrative data sources. Both breaks cause a break in firm-level based time series.

Since the Production Statistics only represents a sample of firms, the results are re-weighted using the entire population of firms from the business register.

New Zealand

The dataset used for MultiProd is built from within the Integrated Data Infrastructure, administered by Statistics New Zealand, and uses a mix of a stratified sampling survey and tax records. In particular, four component datasets have been used:

- The Longitudinal Business Frame (LBF), which is the BR with industry/sector;
- Linked Employer-Employee Data (LEED), constructed from administrative tax data, which include a monthly head count of employees and annual information on working proprietors;
- Tax-filed company accounts (IR10) collected by Inland Revenue;
- The Annual Enterprise Survey (AES), collected by Statistics New Zealand.

The LBF and the LEED cover the whole population.

The IR10 covers mainly small and medium firms and is not mandatory. In particular, firms can choose to send in a copy of their accounts instead (this option is commonly taken by larger firms), whereas sole-proprietors/self-employed can send in an alternative form instead, which does not contain all the components needed for MultiProd (mainly smaller firms).

The Annual Enterprise Survey is a stratified sample survey, targeted at large businesses, which covers only around 4 or 5% of firms. There has been a substantial decrease in the number of units sampled from 2009, in parallel with an increase use of administrative data.

IR10 and AES are used together to compute output, capital and intermediate consumption. The combination of IR10 and AES covers in total around 70% of the population, with some overlap between the two sources but also gaps with firms that do not complete either source (mainly in the "medium" firm population). Therefore, observations have been weighted using weights derived from LBF and LEED.

Norway

Data come from the Accounts statistics merged with information from the business register. All incorporated firms in all economic sectors are included.

All private limited companies in Norway are obliged to submit annual accounts to the Register of Company Accounts maintained by the Brønnøysund Register Centre every year. The accounts statistics contain data from both the income statement and the balance sheet.

Self-employed are not included.

Portugal

The name of the data source used is the Integrated Business Accounts System (*Sistema de Contas Integradas das Empresas*) from Statistics Portugal (INE). Microdata is collected by means of administrative sources (compulsory) supported by business annual fiscal reports to the national fiscal administration (*Informação Empresarial Simplificada* - Simplified Business Information). It includes the population of active companies. The full sample does not have any unit size threshold.

The datasets used covers Sections A to S, excluding K (financial and insurance activities) and O (public administration) of NACE Rev.2. Data refer only to companies and not to all enterprises; individual enterprises (including both sole proprietors and self-employed) are indeed excluded, since there are no information on gross value added for individual enterprises.

The variable used for employment is total persons engaged, expressed in head counts, but data on employees are also available.

The capital variable is only available from 2010 onwards. Therefore, for the purpose of MultiProd, the 2010 value of tangible and intangible assets has been used as initial value for the PIM, deflating it to its 2004 value.

Sweden

The data source used for MultiProd is SBS administrative tax data from Statistics Sweden. The primary source is the standardised accounting statements (SRU) collected from the Swedish Tax Agency, which is the information submitted by enterprises in an appendix to the income tax declaration. All active enterprises within the Swedish Business Register (approximately one million enterprises) are covered.

The variable used for employment is the average number of employees, converted to full-time employees in accordance with what is reported in the enterprises' annual reports.

Switzerland

The data source is the value added statistics (*Wertschöpfungsstatistik*) from the Swiss Federal Statistical Office (SFSO).

The value added statistics are based on a stratified sample survey carried out annually by companies in the secondary and tertiary sectors, except banking and insurance. Non-employer businesses are not included. The value added statistics include companies in the following legal forms:

- Partnerships (individual enterprises, simple companies, collective and limited partnerships);
- Corporations (public companies, limited liability companies, cooperatives, foundations and associations).

The framework for drawing the sample of the value added statistics is established from the enterprises listed in the Business Register, and also in the Structural business statistics (STATENT), the latter is mainly based on data from the register of Old-age and Survivors' Insurance (AHV).

All businesses with more than 50 employees are surveyed. Only companies employing at least three persons (including the owner) are included in the sampling frame.

Detailed questionnaires are sent to enterprises with more than 20 persons employed, smaller firms receive a shorter version.

All types of workers are included. In order to improve the comparability of the results, the number of part-time employees is converted into full-time equivalent on the basis of three different occupation rates.

3. The DynEmp project

3.1. Overall description

Business and employment dynamics – the entry and growth of successful firms, and the downsizing and exit of the least productive ones – are at the core of the processes of creative destruction, resource reallocation and productivity growth. DynEmp is a project aimed at supporting the design of better policies for employment and productivity growth, by providing new empirical evidence.

Measuring employment dynamics and its economic effects in terms of job creation is not an easy task and appropriate data, taking into account both the age and the size of businesses, are necessary. Very few databases allow researchers to follow cohorts of firms (or plants) over time, despite the wide recognition that this is crucial when studying business dynamics, especially in the case of entrants.

The DynEmp project provides a unique approach for this type of investigation. It is based on a distributed data collection to create a harmonised cross-country micro-aggregated database on employment and business dynamics from confidential micro-level data, where the primary sources of firm and establishment data are national business registers and social security data. The project is supported by a network of national experts who run common statistical routines developed by the OECD on confidential microdata to which they have access (see Criscuolo, Gal and Menon (2014[2])). The experts also implement additional country-specific disclosure procedures to ensure that the confidentiality of the underlying microdata is respected.

The advantages of using harmonised micro-aggregated data from business registers for the study of business employment dynamics are manifold. First of all, the different channels of employment changes can be identified separately, distinguishing between gross job creation and job destruction, and between the role of firm entry and exit (the so-called extensive margin) and post-entry growth (the so-called intensive margin). Furthermore, the role of firm age and size can be examined. Last but not least, each of these elements can be compared across countries, sectors and over time.

The first phase of the project (DynEmp Express) was launched in the first half of 2013. It was based on a simplified statistical code which led to the collection of a database based on firm-level sources covering 18 countries over the period 2001-11 (see Criscuolo, Gal and Menon (2014[2])).

A number of significant extensions were implemented in the second phase of the project entitled DynEmp v.2. This version provided a more disaggregated analysis of the growth patterns of incumbents and start-ups, following cohorts of entrants for three, five, and seven years after their entry in the market. The dataset also allowed for a more granular analysis at the 2-digit sectoral level (see Calvino, Criscuolo and Menon (2015[3]), and Calvino, Criscuolo and Menon (2016[4])).

The current version of DynEmp (DynEmp3) is the third round of data collection launched in 2017. Novelties include additional adjustments of the employment variables to proxy total employment in the calendar year, additional aggregations, variables and distributed regressions. Size classes have been refined to maximise the cross-country comparability of some statistics (see Calvino and Criscuolo (2019[5])).

Recently, the DynEmp code has been modified to analyse employment and business dynamics at the regional and local level (see (OECD, 2017[6])), as well as to run on cross-sectional databases that do not allow a panel dimension.

3.2. Methodology

3.2.1. Basic requirements of the DynEmp3 input data

The input data must respect a number of basic requirements:

- The data source must include the universe of units at the country level in the covered sectors. Typical data sources are business registers, social security records or tax records.
- Panel dimension: individual units need to be identified by a unique longitudinal identifier (id) that has to be constant over time. If the unit exits the data source, its identifier must not reappear again, since the statistics on job dynamics for a given year are based also on information on contiguous years.[9]
- Other required variables are:
  - A measure of employment at the unit level;
  - The calendar year to which the time-varying variables refer to;
  - The birth year of the unit;[10]
  - 3-digit sector identifying the main economic activity of the unit, following the ISIC Rev.4 (or NACE Rev.2) classification.[11]

If the data source contains information on different statistical units, the program should preferably be executed at the firm / enterprise level rather than at the plant / establishment level.

Regarding the employment variable, the input data can either refer to the number of employees or to the number of persons engaged, with a preference for the latter variable. Likewise, head count measure should be prioritised, even though the program can be run on data expressed in full-time equivalent.

3.2.2. Main optional variables

The dynemp3 routine allows the inclusion of a number of optional variables, including:

- Turnover
- Ownership (independent, domestic or foreign group)
- Trade status (no trade, importer, exporter, or both)
- Left censoring
- Entry/exit change (*de novo* vs. *de alio*)

When turnover is available, it allows the calculation of several moments of unit-level turnover growth (mean, median, and standard deviation) and labour productivity as well.

The inclusion of the variables ownership and/or trade status in the input data produces two further possible levels of aggregation.

Left censoring allows to take into account the fact that in some data sources birth information is collected only starting from a given year, and this has implications for the age classification.

Whenever information on non-economic entry/exit is available, the entry/exit change options allow to take this into account and create additional aggregations.

3.2.3. Data harmonisation procedures

The dynemp3 program carries out some adjustments, basic consistency checks and cleaning of the data:

- It adjusts the employment variable if it records employees (and not the total number of persons engaged / total employment). The adjustment implemented is total employment = employees + 1. This corresponds to assuming on average one proprietor / owner.
- It shifts year-related variables to match the calendar year as much as possible.
- It replaces negative employment to missing.
- It drops units that have missing or zero total employment in all periods.
- It interpolates employment records that are disproportionally smaller/bigger than those of the previous and following year.
- It replaces industry classification that varies over time with the modal 3-digit sector the unit's activity is classified by. In case of multiple modes, the program chooses the most recent mode.
- It replaces the birth year variable with the minimum value in case it changes over time for the same unit, and it changes the birth year variable to the modal value of the same unit if it is missing in a few years.
- It switches the birth year to the first year with positive employment when the birth year is within the observed period but the unit has zero or missing employment. This correction can be optionally excluded.

3.2.4. The DynEmp output

The dynemp3 routine produces a set of statistics based on micro-level (firm or plant) employment dynamics. The information is aggregated to the level of sectors, age classes, and size classes. In addition, there are two further optional levels of aggregation at the level of ownership status (independent, domestic or foreign group) and engagement in foreign trade (no trade, importer, exporter, or both).

The output files can be classified into five groups:

- One dataset containing transition matrices of selected group of firms (entrants, exiting and incumbents), classified along the age and size dimension, over a 3, 5, 7, 10 or 14 year time horizon.
- 15 datasets containing different aggregations of annual flows data, including statistics on gross job creation and destruction, average growth rate, average and median size, and age.

- Up to nine excel tables containing the output of the distributed regressions of growth rate and probability of exit on size and age dummies; average growth, share of stable, growing, shrinking firms and overall frequencies on size and age dummies; coefficients of the Pareto distribution of size (employment), descriptive statistics on three 3-digit sectors (activities of holding companies, head offices, and temporary employment agencies).
- Up to eight text files reporting a tabulation of gaps in the data by macro sector of activity.
- A number of graphs reporting density estimations and histograms of the employment growth distribution at different points in time.

3.3. DynEmp metadata

3.3.1. Overview of the dataset

Information in this section only refer to the latest phase of data collection, DynEmp3.

As of June 2019, 17 countries have been included in the DynEmp3 cross-country database: Austria, Belgium, Brazil, Canada, Costa Rica, Finland, France, Hungary, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden and Turkey.

A number of additional countries have also participated in the third wave of data collection: Australia, Chile, Denmark, Estonia and Luxembourg.

These countries have run the DynEmp3 code but sent a partial version of the output, or their data presented significant confidentiality challenges, or their results are still under evaluation. Therefore, their output has not been yet included in the cross-country dataset dated January 28th 2019, and additional details on their metadata are not reported in this version of the metadata paper, but might be incorporated in a future vintage of the database.

Table 3.1 lists the names of the national data sources, and the unit of analysis used.[12]

Table 3.1. National data sources

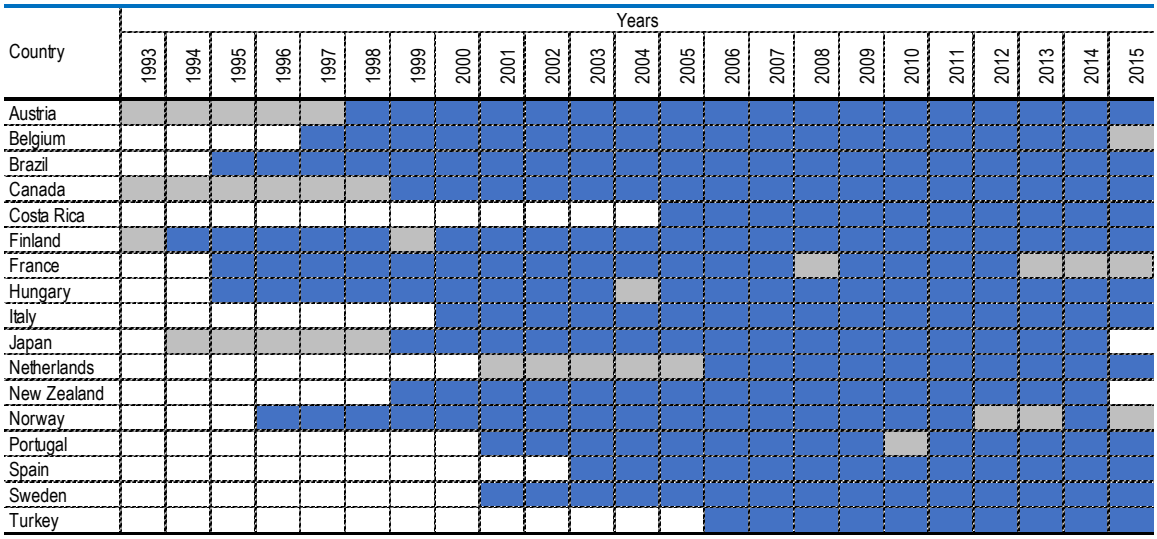
Country	Statistical unit	National data sources
Austria	Establishment	WIFO INDI-DV (social security data from the Main Association of Austrian Social Security Institutions)
Belgium	Firm	National Social Security Office data; Central Balance Sheet Office (National Bank of Belgium) and VAT data for turnover
Brazil	Firm	RAIS (Fiscal/tax register), Ministry of Labour and Employment
Canada	Firm	Statistics Canada's Longitudinal Employment Analysis Program (LEAP)
Costa Rica	Firm	*Registro de variables económicas*-REVEC, Banco Central de Costa Rica
Finland	Firm	Statistics Finland's Business Register on Companies
France	Firm	FICUS/FARE fiscal dataset, INSEE
Hungary	Firm	Corporate Income Tax data (CIT) of National Tax and Custom Administration
Italy	Firm	ASIA (Business Register) + Frame SBS from 2011 onward, ISTAT
Japan	Establishment	Census of Manufactures, METI
Netherlands	Firm	Business Register; FATS; Fiscal register, CBS
New Zealand	Firm	Integrated Data Infrastructure, Statistics New Zealand
Norway	Firm	Establishment population register; Accounts statistics
Portugal	Firm	*Relatório Único / Quadros de Pessoal* (BR), Ministry of Labour, Solidarity and Social Security
Spain	Firm	Central Business Register, INE
Sweden	Firm	Statistics Sweden's Register-based labour market statistics (RAMS)
Turkey	Firm	Entrepreneur Information System from the Ministry of Industry and Technology

Notes: Access to French data used in this work benefited from the use of Centre d'Accès Sécurisé aux Données (CASD), which is part of the "*Investissements d'Avenir*" programme (reference: ANR-10-EQPX-17) and supported by a public grant overseen by the French National Research Agency (ANR).
Access to the anonymised data for New Zealand was provided by Statistics New Zealand in accordance with security and confidentiality provisions of the Statistics Act 1975 and secrecy provisions of the Tax Administration Act 1994. The results in this report have been confidentialised to protect individual businesses from identification. See Appendix for the full disclaimer.

Data for most countries are collected at the firm level, with the exception of Austria and Japan that used establishment-level data.

The time coverage of the database, described in Table 3.2, varies according to country, broadly covering the period between 1999 and 2015.

Table 3.2. Time coverage



Note: The coverage is based on the level 2 output database. Years in grey have been excluded from the analysis due to ongoing checks in or specific features of the underlying data. In most analyses, the first year is also excluded because two years are required to compute job flows.

Table 3.3 lists the sectoral classifications and coverage of the source data. The industry aggregation used in the DynEmp3 database follows the OECD STAN (SNA) A38 industry classification. DynEmp typically covers the whole economy with the exceptions of the financial sector (ISIC Rev.4 64-66), the public sector (ISIC 84), and households and other activities (ISIC 97-99). For the time being, the focus of the analytical work is restricted to manufacturing and non-financial market services, i.e. sections C, G to J, L to N of ISIC Rev.4. The analysis by A38 industry often excludes "Coke and refined petroleum", "Real estate", and "Basic pharmaceuticals".

Table 3.3. Sectoral classifications and coverage

Country	Classification followed	ISIC Rev.4 industries covered[13]
Austria	ÖNACE 2008 (NACE Rev.2)	All
Belgium	NACE-BEL 2008	All
Brazil	CNAE 1 up to 2007; CNAE 2 from 2006	All
Canada	NAICS	All except ISIC Rev.4 Sections P and Q
Costa Rica	CAECR 2011 (based on ISIC Rev.4)	All
Finland	NACE Rev.1.1 and 2	All
France	Nomenclature d'Activités Française (NAF) Rev.1 and 2	All
Hungary	TEÁOR 03 up to 2007; TEÁOR 08 from 2008	All
Italy	Ateco Rev.2	All except ISIC Rev.4 Sections A and O
Japan	JSIC Rev.10 to 13 (change in 2002, 2008 and 2014)	Manufacturing industries only
Netherlands	NACE Rev.1.1 up to 2008; NACE Rev.2 from 2009	All
New Zealand	ANZSIC06	All
Norway	NACE Rev.1.1 up to 2007; NACE Rev.2 from 2008	All
Portugal	NACE Rev.1.1 up to 2006; NACE Rev.2 from 2007	All
Spain	NACE Rev.1.1 up to 2008; NACE Rev.2 from 2009, both classifications available for 2007 and 2008	All except ISIC Rev.4 Section O
Sweden	NACE Rev.1.1 up to 2006; NACE Rev.2 from 2007	All
Turkey	NACE Rev.2	All

Table 3.4 summarises entry and age information. It specifies, for each country, if age is calculated as the year of incorporation, the year of first appearance in the data source, or the year of birth as reported in the data source. When multiple data sources are combined, age can be determined using several measures.

Table 3.4. Entry and age information

Country	First appearance	Incorporation	Birth
Austria	✓		
Belgium	✓		
Brazil	✓		
Canada	✓		
Costa Rica	✓		
Finland	✓	✓	
France	✓	✓	
Hungary	✓		
Italy	✓		
Japan	✓		
Netherlands	✓		
New Zealand			✓
Norway			✓
Portugal			✓
Spain	✓		
Sweden			✓
Turkey	✓	✓	✓

Table 3.5 summarises which employment variable has been used for DynEmp3 (number of persons engaged or employees) and the employment recording unit. The majority of countries use head count to record employment, except Canada, Finland, Hungary and the

Netherlands. In Canada, the measure that is used is an average labour unit (ALU), which is a measure of employment derived from the wages paid to employees divided by the average wage of those employed in firms in the same industry and region, and of the same size. In Spain where both employment variables are available, the number of employees is in head count, whereas the number of persons engaged is in full-time equivalent.

Analytical work mainly focuses on units with two or more persons engaged (one or more employees).

Table 3.5. Employment data characteristics

Recording units and variables used

Country	Persons engaged	Employees	HC	FTE
Austria		✓	✓	
Belgium		✓	✓	
Brazil		✓	✓	
Canada		✓		ALU (proxy)
Costa Rica		✓	✓	
Finland	✓			✓
France		✓	✓	
Hungary		✓		✓
Italy	✓		✓	
Japan	✓		✓	
Netherlands	✓			✓
New Zealand	✓		✓	
Norway		✓	✓	
Portugal	✓		✓	
Spain		✓	✓	
Sweden	✓		✓	
Turkey	✓		✓	

Note: HC indicates head count and FTE indicates full time equivalent. The table presents, for each country, the recording unit and the variable that have been used in the cross-country database. Nevertheless, more variables and units can be available in the source data. For Belgium, FTE is also available. For Japan, the number of employees is available as well. For Spain, the two employment variables are available, see country-specific information for more details.

Table 3.6. Timing of recording for employment data

Country	Yearly average	Point in time
Austria	✓	
Belgium	✓	
Brazil	✓	
Canada	✓	
Costa Rica	✓	
Finland	✓	
France	✓	
Hungary	✓	
Italy	✓	
Japan		✓
Netherlands		✓
New Zealand	✓	
Norway		✓
Portugal		✓
Spain	✓	✓
Sweden		✓
Turkey	✓	

Table 3.6 describes the timing of recording of the employment variable (yearly average or point in time). Most of the countries use a yearly average. Japan and Norway record employment at the end of the year, Portugal in October and Sweden in November. In the Netherlands, the employment level reflects the most recent information available in the fourth quarter of the pertinent year. In Spain, employment is recorded as a yearly average when the firm is surveyed (typically large units), and as a point in time when the firm is not involved in a survey.[14]

3.3.2. Country-specific information

Austria

The data source is the WIFO INDI DV dataset, that contains anonymised social security data collected by the Main Association of Austrian Social Security Institutions (*Hauptverband der österreichischen Sozialversicherungsträger*), which is the umbrella organisation of 22 social insurance institutions.

Insurance providers submit personal information on individuals who have applied for and who receive social security benefits. Non-employer businesses are not included.

The statistical unit is the establishment.

Employment refers to the number of employees, expressed in head counts.

Belgium

The data come from the register of the National Social Security Office (NSSO), merged with information from the Central Balance Sheet Office (National Bank of Belgium) and VAT data from the Federal Public Service Finance for turnover.

The database covers all employers in Belgium, as they are required to report staff movement to the NSSO, so that all social security agencies receive an electronic notification every time an employee is hired by or leaves an employer. The statistical unit is the firm.

Employment data refer to the number of employees in headcounts. Temporary workers are not included, nor non-employer businesses. Information on the trade status of the firm is available. The distinction between *de novo* versus *de alio* entry is available as well.

Brazil

The data come from the Annual Relation of Social Information (RAIS) compiled by the Ministry of Labour and Employment (MTE). RAIS is a longitudinal matched employer-employee dataset covering by law the universe of formally employed workers in Brazil. Each observation in the source dataset consists of contract-worker-establishment data for a given year.

The data cover formally employed wage earners, either public or private. All tax-registered businesses must declare how many employees they had in December 31, and must also inform labour force movement (hiring and firing) across the year. To further improve data quality and ensure compliance, the MTE cross-tabulates registry information from many other official sources, such as the Ministry of Social Security, the Federal Reserve and the Secretary of Federal Revenues (taxes). As a result, MTE estimates that RAIS is annually declared by 98% to 99% of officially existing firms.

The statistical unit is the firm.

Employment refers to the number of employees, expressed in head counts.

Canada

The data source is the Statistics Canada's Longitudinal Employment Analysis Program (LEAP), which is an administrative database that contains annual employment information for each employer business in Canada. The information is generated from the annual statements of remuneration paid (T4 slips) that Canadian businesses are required to issue to their employees for tax purposes. More precisely, LEAP uses three major sources of information: T4 administrative data received from the Canada Revenue Agency, data from Statistics Canada's Business Register, and data from Statistics Canada's Survey of Employment, Payrolls and Hours (SEPH).

LEAP covers all incorporated and unincorporated businesses that issue at least one T4 slip in any given calendar year, but excludes self-employed individuals or partnerships where the participants do not draw salaries.

The statistical unit is the firm.

For employment, the average labour unit (ALU) is used. It is constructed by dividing a firm's payroll by the average wage paid by firms in the same province, industry and size class. As such, the measures capture the amount of employment if firms were paying their workers an average wage and employing them for the same duration as other comparable firms. While it is not a full-time equivalent count, it can be thought of as a "normal" equivalent count.

Costa Rica

The data source is REVEC (*Registro de variables económicas*), which is a database maintained by the Central Bank of Costa Rica, gathering different external data sources. The one allowing the update of business dynamics is the *Registro Nacional*, which registers certain acts and contracts related to the existence, validity and representation of different classes of legal persons. Other administrative data sources such as social security records are used to update the economic activity of the legal units. The statistical unit is the firm.

Employment refers to the number of employees, expressed in head counts, working at least 20 hours a week, either temporarily or permanently. Self-employed are not covered by the data.

Finland

Data come from the Business Register on Companies maintained by Statistics Finland. The data sources of the Business Register are several administrative records (the most important being the Tax Administration) and Statistics Finland's direct inquiries to enterprises.[15]

The Business Register covers all enterprises, corporations and private practitioners of trade that are liable to pay value added tax or have paid employees. The VAT threshold is EUR 8 500. The estimated coverage of the Business Register for enterprises falling below this VAT threshold is about 95%.

The statistical unit is the firm.

Employment data record persons engaged comprising all types of employees and self-employed persons and are expressed in full-time equivalent.

Information on the trade status of the firm is available.

France

The data source is the FICUS/FARE fiscal dataset maintained by INSEE, the French statistical office.

The FICUS (*Fichier Complet Unifié de SUSE*) dataset covers all private sector firms in France, as recorded on the SUSE (*Système Unifié de Statistiques d'Entreprises*). Since 2008, the dataset has been replaced by the FARE (*Fichier Approché des Résultats ESANE*).[16] Both datasets combine information retrieved from the annual statements of enterprises supplied by the General Tax Directorate (Industrial and Commercial Profits System, Non-Commercial Profits system and Agricultural Profits) with results from the annual sectoral surveys (EAE, ESA) carried out by INSEE. Among other characteristics, they provide information on turnover, employment, value added, investments, the 4-digit sector each firm belongs to and geographic locations.

The statistical unit is the firm.

Employment refers to the number of employees, expressed in head counts.

A redesign in the data collection process occurs in 2008 (from FICUS to FARE).

On the 1st of January 2009, the French government introduced a new self-employed worker status (« *auto-entrepreneur* »); this resulted in an increase in the number of self-employed.

Hungary

The data source for Hungary is the corporate tax return data of the National Tax and Customs Administration (NAV). The birth year information comes from the Institution Registry of Hungarian NSI via merging datasets, and refers to first appearance (company establishment).

Data from NAV contain the financial statements of companies with double-entry accounting.

The statistical unit is the firm.

The employment variable records the number of employees expressed in full-time equivalent.

The distinction between *de novo* versus *de alio* entry is available.

The criteria to file a double book keeping tax report has been lowered in 2004.[17]

Italy

The data comes from Asia (the Italian Business Register), complemented with the statistical register Frame SBS from 2011 onwards.

The Asia register (*Archivio statistico delle imprese attive in Italia*) covers all enterprises carrying on economic activities contributing to gross domestic product at market prices, in the fields of industry, trade and services.

In 2013, ISTAT has developed a new statistical register for the annual production of profit-and-loss accounts of small and medium enterprises. This register (called Frame SBS) massively uses firm-level administrative and fiscal data as primary sources of information to estimate key structural business statistics, while sample data on SMEs are used for estimating those variables which cannot be directly obtained from administrative sources.

The statistical unit is the firm.

Total employment in Asia is given by the sum of independent and dependent workers. Employment does not include temporary workers nor part-time workers.

Information on the trade status and the ownership of the firm is available from 2011.

Japan

Data come from the Census of Manufactures carried out by the Ministry of Economy, Trade and Industry (METI).

The Census of Manufactures is conducted on all establishments of the manufacturing industry with four or more employees.

Up to 2008, all establishments of the manufacturing industry in years ending with 0, 3, 5 and 8 were surveyed. A long questionnaire is sent to establishments with 30 or more employees, while a shorter version is for those with 29 or fewer employees, in both cases excluding head offices that are not involved in the manufacturing processing or repair of industrial products. In 2007, the Census frame was reviewed in comparison with the list of manufacturing establishments from the Establishment and Enterprise Census. In 2011, the

Census of Manufactures was not conducted, and the necessary items were surveyed by the Economic Census for Business Activity in 2012.

The statistical unit is the establishment.

The Census provides the number of persons employed as a total number of sole proprietors and unpaid family workers and regular workers at the end of the calendar year. It excludes temporary workers, i.e. persons that are employed on agreements shorter than a month or on daily agreements. In the cross-country database, the employment variable is the number of persons engaged in head counts.

Netherlands

The data source is the General Business Register (*Algemeen Bedrijvenregister*) of Statistics Netherlands, which is mainly based on the New Trade Register (NHR) of the Chamber of Commerce since 1 April 2014. From 2006 onwards, the employment figures are derived from social security information. Before 2006, it was derived from Chambers of Commerce information, which is of a less quality, especially concerning changes in employment.

Registration in the NHR is compulsory for every company and almost every legal entity. Coverage might be less exhaustive in sectors not covered by DynEmp (mostly in public administration and non-market services).

The statistical unit is the firm.

Employment concerns persons engaged, and refers only to those persons on the firm's payroll (no secondments or agency workers). It is in full-time equivalent and includes all paid workers, including paid working proprietors and family members. The number of full-time equivalents is rounded to integers.

There are several changes and redesigns in the time series:

- In 2006, there has been a redesign of the BR, concerning the derivation of the population and the population units.
- In 2010, a transition to a new enterprise definition was implemented, by which some units were consolidated into a single unit. This improved delineation of enterprise groups aims at increasing the linkability of the BR with external administrative data sources.
- In 2013, there has been an adjustment in the derivation of persons employed, concerning self-employed only who work a few hours in their own firm. This increased the number of self-employed in the BR.
- Starting in April 2014, the Business Register uses a new main source: the New Trade Register (NHR) maintained by the Chamber of Commerce. This concerns an extension of the legal obligation to enrol in this register. Although the legal obligation was changed in 2008, it only has an effect on the measured population in the BR from 2014 onwards when the BR was aligned to the NHR.

Information on the trade status of the firm is available from 2009 (for goods only). The ownership variable is available from 2007.

New Zealand

The dataset used for DynEmp is built from within the Integrated Data Infrastructure, administered by Statistics New Zealand. In particular, three component datasets have been used:

- The Longitudinal Business Frame (LBF), which is the Business Register with industry/sector;
- Linked Employer-Employee Data (LEED), constructed from administrative tax data, which include a monthly head count of employees and annual information on working proprietors;
- The Business Activity Indicator (BAI), which is constructed by Statistics New Zealand based on Goods and Services Tax (GST) records collected by Inland Revenue.

The LBF and the LEED cover the full population of firms.

The BAI fully covers all GST-liable products, including exports (despite being zero-rated for GST). Financial services industries are excluded from the DynEmp analysis of sales, as the main products of this industry are GST-exempt.

The statistical unit is the firm.

Employment concerns persons engaged (employees and working proprietors), expressed in head counts, and refers only to those persons on the firm's own payroll (no secondments or agency workers).

Norway

Data come from the business register, merged with information from the Accounts statistics.

All incorporated firms in all economic sectors are included. Self-employed are not covered.

The statistical unit is the firm.

Employment refers to the number of employees, expressed in head counts and recorded as end-of-year values.

Portugal

The data source is *Quadros de Pessoal*, a longitudinal matched employer-employee dataset collected by the Cabinet of Strategy and Planning (GEP) from the Ministry of Labour, Solidarity and Social Security.

Quadros de Pessoal draw on a compulsory annual census of all firms in Portugal that employ at least one worker, the Single Report (*Relatório Único*). The data also cover individual information on all staff working for each firm in a reference month (October). The variables available in the source data set include the firm's location, industry, total employment, sales, ownership structure, and legal setting.

The statistical unit is the firm.

The employment variable includes all types of workers, including unpaid family workers, employers when exercising functions in the company (e.g. the owner/partner-manager),

active members of cooperatives and interns with a working contract. Personnel on short-term leave (such as sickness, maternity, strike or holidays) are included, whereas personnel on long-term leave (such as military service) are not reported. Civil servants, the self-employed and domestic services are not covered, and the coverage of agriculture is low given its low share of wage earners.

There is a change in 2010 as Portugal took steps to simplify administrative procedures. One of the measures was to create a new model of annual report synthesising all the information in a single document that initially had to be provided by employers in separate documents at different times of the year.

Spain

Data come from the Central Business Register (*Directorio Central de Empresas*, DIRCE) maintained by the Spanish statistical office, which is updated from several administrative sources and surveys (mainly tax, social security and legal sources, structural and short-term surveys of enterprises).

The database integrates all Spanish enterprises, including their local units situated within the country, into one single information system.

The statistical unit is the firm.

Two employment variables are available. One records the number of employees (in head counts) and the other the number of persons employed (in FTE). This latter variable is the addition of non-paid persons (estimated in FTE) and the number of paid workers (equal to number of employees, in head counts). All types of workers are included and non-employer businesses are covered by the data. In the cross-country database, the employment variable used is the number of employees in head counts.

When the firm is object of a survey, employment datum is a yearly average. When the firm is not involved in a survey, employment datum comes from administrative sources and it is mainly a point in time (1st of December of each reference year when the data source is Social Security).

Information on the ownership of the firm is available.

Sweden

Data come from the register-based labour market statistics (RAMS) produced annually by Statistics Sweden. They are comprehensive statistics based on, among other things, income statements that employers are, by law, required to send to the Swedish Tax Authority. The standard accounting statements of the Swedish Tax Authority are used for the collection of data for self-employed persons.

RAMS describes employment in the form of gainful work and provides a full-covering image of the population, aged 16 to 74 years, related to the labour market. The register contains a link between the individuals and the establishments and companies to which the persons work.

The statistical unit is the firm.

The employment variable in RAMS includes all persons who are assessed to have carried out an average of one hour's work per week during the month of November. Even those

who were temporarily absent during the measurement period, for instance due to illness, are included in the frame.

Between 2003 and 2004, the method for how individuals become self-employed changed. It was modified again between 2010 and 2011 for persons over 64 years. Both changes increased the number of self-employed people.

Information on the trade status and the ownership of the firm is available.

Turkey

The data come from the Entrepreneur Information System (EIS) compiled by the Ministry of Industry and Technology. It is constructed from the linkage of information from the Statistical Business Register and different administrative sources, of which:

- Data from Revenue Administration, the Turkish Tax Authority (balance sheet and income tables; corporate and income tax declarations; data from forms BA and BS);
- Social Security Institution data. Normally, the employer must notify the authorities of new starters immediately by submitting a notice of employment.

The dataset includes information for almost all segments of the economy. However, it omits information on the banking sector, as well as on firms that either make purchases for less than TRY 160 000 or have annual sales below TRY 220 000.

The statistical unit is the firm.

Employment refers to the number of persons engaged, expressed in head counts.

Appendix

Disclaimer for New Zealand data

The New Zealand results in this report are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI), managed by Statistics New Zealand.

The opinions, findings, recommendations, and conclusions expressed in this report are those of the author(s), not Statistics NZ or MBIE.

Access to the anonymised data used in this study was provided by Statistics NZ under the security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation, and the results in this report have been confidentialised to protect these groups from identification and to keep their data safe.

Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy impact assessment for the Integrated Data Infrastructure available from www.stats.govt.nz.

The results are based in part on tax data supplied by Inland Revenue to Statistics NZ under the Tax Administration Act 1994. This tax data must be used only for statistical purposes, and no individual information may be published or disclosed in any other form, or provided to Inland Revenue for administrative or regulatory purposes.

Any person who has had access to the unit record data has certified that they have been shown, have read, and have understood section 81 of the Tax Administration Act 1994, which relates to secrecy. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.

Annex Table 1. Contributors to MultiProd and DynEmp

Country	National representative(s)	Institution(s)
Australia	David Hansell	ABS
Austria	Jürgen Janger, Werner Hölzl	WIFO (Austrian Institute of Economic Research)
Belgium	Michel Dumont, Chantal Kegels	Federal Planning Bureau
Brazil	Carlos Henrique Leite Corseuil	IPEA (Institute of Applied Economic Research)
Canada	Pierre Therrien, Jay Dixon, Danny Leung	Industry Canada; Statistics Canada
Chile	Antonio Martner Sota	Ministry of Economy, Development and Tourism
Costa Rica	Tatiana Vargas, Alonso Alfaro-Urena	Ministry of Foreign Trade; Central Bank of Costa Rica
Denmark	Dorte Høeg Koch	Ministry of Industry, Business and Financial Affairs
Estonia	Karel Lember	Ministry of Economic Affairs and Communications
Finland	Mika Maliranta	ETLA (Research Institute of the Finnish Economy)
Germany	Marie Le Mouel	DIW (German Institute for Economic Research)
Hungary	Mihály Szoboszlai	Central Bank of Hungary
Ireland	Luke Rehill, Brendan O'Connor, Javier Papa	Department of Finance; Department of Business, Enterprise and Innovation
Italy	Stefano Costa	ISTAT
Japan	Kenta Ikeuchi	RIETI (Research Institute of Economy, Trade and Industry)
Luxembourg	Chiara Peroni, Bruno Rodrigues	STATEC
Netherlands	Michael Polder	Statistics Netherlands
New Zealand	Lynda Sanderson	Ministry of Business, Innovation and Employment
Norway	Arvid Raknerund, Diana-Cristina Iancu	Statistics Norway
Portugal	Paulo Jorge Martins Dias, Ricardo Alves, Guida Nogueira	Ministry of Labour, Solidarity and Social Security; Ministry of Economy
Spain	Valentín Llorente García	INE
Sweden	Fredrik Andersson	Statistics Sweden
Switzerland	Marius Ley	Swiss Federal Office of Statistics
Turkey	Faik Yücel Günaydın	Ministry of Industry and Technology

Endnotes

[1] OECD STI/PBD. The authors would like to thank all members of the DynEmp and MultiProd networks for having provided information on national data collection methodology (see Annex Table 1 for the list of contributors). They are also very grateful to Giuseppe Berlingieri, Chiara Criscuolo, Nicolas Gonne, Alexander Himbert and Rudy Verlhac for useful comments and inputs.

[2] The OECD currently follows this approach in another ongoing project, MicroBeRD, which provides information on R&D activity in firms from official business R&D surveys (http://oe.cd/rdtax).

[3] For countries that are not present in the STAN database, information from other sources like national accounts, WIOD and EU KLEMS has been used for obtaining deflators and depreciation rates, whereas the STAN IO tables have been used for labour and intermediates shares.

[4] As the program does not implement the conversion from country-specific industry classifications into ISIC, if countries use a different classification, they are requested to convert it beforehand.

[5] The list of sectors is given in Table 2.3.

[6] Except ISIC Rev. 4 sections K, O, T and U that are not covered by MultiProd.

[7] Previously known as the Expanded Analytical Business Longitudinal Database (EABLD).

[8] Micro-firms that have opted for the Micro-BIC and Micro-BNC tax regimes (reporting less than EUR 32 600 annual turnover without taxes) are included in FARE, not in FICUS. In order to ensure comparability over time, those micro-firms are dropped from the analysis.

[9] Ongoing work aimed at developing a version of the statistical routine that allows the reappearance of a same unit identifier under certain conditions is in progress, in order to extend the coverage of the DynEmp3 cross-country database.

[10] This can be outside (i.e. earlier than) the period covered by the business register. It can be missing for some units, in which case the first year of appearance with positive employment is assumed to be the birth year. For those units where this coincides with the first year of the database - and birth year is missing -, the birth year is left missing.

[11] If the sectoral classification is at a finer level than 3 digit, it is automatically converted to 3 digit. The program can also deal with the dataset being partially or completely classified according to the ISIC Rev.3.1 (NACE Rev.1.1) classification.

[12] Some of the data sources used in the current version of the DynEmp3 cross-country database may change in the future. For France, work is in progress to integrate additional information from DADS in a future vintage of the DynEmp3 cross-country database.

[13] Except ISIC Rev. 4 sections K, O, T and U that are not covered in DynEmp.

[14] See the country-specific sub-section for additional details.

[15] Inquiries are sent yearly to most multi-establishment enterprises and single-establishment enterprises with over 20 employees. Other enterprises receive inquiries less often on a rotation basis.

[16] Micro-firms that have opted for the Micro-BIC and Micro-BNC tax regimes (reporting less than EUR 32 600 annual turnover without taxes) are included in FARE, not in FICUS. In order to ensure comparability over time, those micro-firms are dropped from the analysis.

[17] Since 2004, firms reporting at least HUF 50 million (around EUR 200 000) turnover in two consecutive years are included.