Chapter 5.

Measuring regional business employment dynamics from micro-aggregation of administrative data[1]

This chapter presents the OECD DynEmp Regional project, a distributed micro-data project aimed at analysing confidential administrative micro-data on employment dynamics at the level of TL2 regions and metropolitan areas. The chapter presents within-country differences in plant-level employment dynamics for the TL2 regions of Costa Rica, Finland, France and Sweden and the metropolitan areas of France and Sweden. The role of plant characteristics, such as age or size, and regional characteristics, in particular regional productivity, agglomeration economies, innovation and the rural-urban continuum, is also examined. The chapter finally discusses the methodological challenges and the solutions implemented in the first version of the dynemp_reg *routine, providing a detailed description of the inputs and outputs of the statistical programme and of the micro-aggregation procedure.*

Introduction

Business employment dynamics are at the core of the creative destruction process, with important effects on resource reallocation and productivity growth. A growing body of evidence reveals significant differences in employment and business dynamics across countries and over time. These differences are particularly relevant when analysing the employment growth trajectories of start-ups and young firms, which importantly contribute to job creation (Criscuolo, Gal and Menon, 2014; Calvino, Criscuolo and Menon, 2015). Policy settings and structural factors have been found to be important drivers explaining cross-country differences in employment dynamics and heterogeneous performance of entering versus incumbent firms across countries (Calvino, Criscuolo and Menon, 2016; OECD, 2016a).

As of today, only a limited body of research is able to disentangle the regional and local specificities underlying economy-wide trends in business dynamics using comparative and representative data for a significant number of countries at a reasonable and unique level of detail. This chapter contributes to building new evidence in this area both by proposing a replicable methodology and new evidence at the regional (TL2) and metropolitan level for a set of four countries.

In particular, the *DynEmp Regional database* allows focusing separately on the business and employment dynamics of the manufacturing and non-financial business services sectors both in a cross-section and over time. This separate focus, combined with longitudinal information on employment, entry/exit, age and size classes, represents a unique characteristic of the data collected using the methodology proposed in this chapter.

As argued in other chapters of this report, analysing employment and business dynamics at regional and local level is particularly relevant for different reasons. Region-specific shocks may induce adjustment mechanisms that may be different from those related to economy-wide shocks. For instance, a shock in a particular region might be associated with higher migration in neighbouring regions (that are not necessarily in the same country) than a national shock would induce across countries (Decressin and Fatas, 1995). Symmetrically, regional characteristics may be associated with heterogeneous responses to economy-wide shocks. For instance, rural areas may have economic and employment dynamics that are different from urban or metropolitan areas, possibly due to agglomeration economies, spatial wage disparities and localised knowledge spillovers (see, among others, Rosenthal and Strange [2004]; Puga, 2010). Furthermore, more specialisation in the production of goods and services is likely to occur at the regional level, and this possibly influences employment dynamics and new business formation in a way that is not observable in economy-wide studies. Finally, the pronounced heterogeneity across firms, industries and regions can be addressed more appropriately by using micro-level analysis that provide detailed firm characteristics.

In this context, this chapter presents the OECD DynEmp Regional project, a new distributed micro-data project aimed at analysing confidential administrative micro-data on employment dynamics at the regional and local level by means of micro-aggregation. This chapter is part of a larger effort co-ordinated by the OECD Directorate for Science, Technology and Innovation to provide new evidence on employment dynamics, innovation activities and productivity across countries exploiting firm-level data.[2]

The aim of this chapter is to showcase the potential of the regionalisation of the DynEmp algorithm. In particular, it provides descriptive and econometric evidence based on the first wave of data collection, illustrating within-country employment dynamics in TL2 regions of Costa Rica, Finland, France and Sweden and metropolitan areas in France and Sweden.

This evidence highlights the heterogeneity in the employment distribution within countries and regions. It focuses on employment growth dynamics by plant characteristics (namely size and age), suggesting that young plants show a higher employment growth performance. It analyses the role of different groups of plants (entering, incumbents and exiting) and their relative contribution to net job creation in TL2 regions and metropolitan areas, suggesting that – for instance – in the Finnish non-financial market services sector, for every 100 jobs, in the average TL2 region about 1 new job is created by incumbents, about 3 by entrants and about 4 jobs are destroyed by exiting plants.

This chapter also provides some more formal analysis that investigates the role of young plants for regional employment growth and the regional characteristics associated with their location, suggesting that frontier regions, more innovative regions, and – especially in services – urban regions appear to have a higher share of employment in young units. In fact, agglomeration economies appear to be particularly relevant in non-financial market services.

Finally, this chapter discusses the methodological challenges together with the solutions proposed and implemented in the algorithm. In particular, the chapter provides details on the characteristics of the statistical programme, the required input data and outlines the output data produced.

As discussed in more detail below, the regional DynEmp algorithm is modular and flexible and could potentially be further extended in breadth to cover all broad sectors of the economy and in depth at lower levels of sectoral aggregation (e.g. at the level of 3- and 4-digit sectors) and/or at more detailed regional levels, were confidentiality not an issue. This is a unique feature that could be extremely valuable to investigate the evolution of regional disparities in business dynamism, the growth performance of young firms, etc., and their determinants both structural and policy-driven. Finally, the development of a comparable cross-country database of business dynamics across regions could be very useful in the evaluation of policies aimed at supporting entrepreneurship and employment growth in lagging regions.

This chapter is organised as follows. The next section describes the OECD DynEmp project, its aims and different phases, and focuses on the challenges and the proposed solutions to its regionalisation. The following section describes more in depth the OECD DynEmp Regional programme (*dynemp_reg*), together with the characteristics of the input required and the output databases produced. The chapter then provides an overview of the preliminary version of the *DynEmp Regional database* and presents preliminary evidence based on the data collected. It also discusses the current challenges faced. The final section concludes and discusses the next steps of the DynEmp Regional project.

DynEmp: Methodology and challenges

This section provides an overview of the OECD DynEmp project and its regional extension (DynEmp Regional), discussing the main methodological challenges faced and the solutions implemented.

The OECD DynEmp project

The OECD DynEmp project is based on a distributed data collection exercise aimed at creating a harmonised cross-country micro-aggregated database on business employment dynamics from confidential micro-level sources at national level.[3] The primary sources of data are national business registers and for some countries, such as Brazil, social security records.

Box 5.1. **The Dynemp methodology and the data collected**

The DynEmp project is led by the OECD Directorate for Science, Technology and Innovation and supported by a network of national experts who run the common algorithm centrally developed by the OECD DynEmp team on the confidential micro-data to which they have access. The distributed micro-data approach involves developing a computer code by the OECD DynEmp team (the preferred statistical programme for this purpose is Stata by StataCorp LP) and then running this code in a decentralised manner by national experts from statistical agencies, academia, ministries or other public institutions who have access to the national micro-level data. The micro-aggregated data generated by the centrally designed but locally executed programme codes are then sent back for comparative cross-country analysis to the OECD. These data reduce confidentiality concerns as they aggregate information at a sufficiently high level, and achieve a high degree of harmonisation, as the definition of the extracted information is the same, ensured by the centrally written computer routine. The experts also implement country-specific disclosure procedures in order to ensure that confidentiality requirements are respected.

The first phase of the project was implemented in the first half of 2013 and was called DynEmp Express. This first phase was based on a simplified statistical routine which led to the collection of a database at the national level covering 18 countries (see Criscuolo, Gal and Menon [2014]). The second phase of the project, called DynEmp v.2, aimed at building a database which contains more detailed data on the within-sector contribution of start-ups and young firms to employment growth, in order to analyse the role played by national policies and framework conditions for employment growth (see, e.g. Calvino, Criscuolo and Menon [2015], [2016]). At the time of writing, 23 countries have been successfully included in the DynEmp v.2 database (Australia, Austria, Belgium, Brazil, Canada, Costa Rica, Denmark, Finland, France, Hungary, Italy, Japan, Luxembourg, the Netherlands, Norway, New Zealand, Portugal, South Africa, Spain, Sweden, Turkey, the United Kingdom and the United States). A new version of the DynEmp programme (dynemp3) is under development at the time of writing. Two parallel projects also led by the OECD Directorate for Science, Technology and Innovation, MultiProd and MicroBerd, have been also recently launched. While DynEmp focuses on employment dynamics, MultiProd is a distributed micro-data project aimed at analysing the dynamics of productivity and allocative efficiency across countries, and MicroBerd is aimed at investigating the heterogeneity and the policy drivers of business R&D investment exploiting cross-country harmonised micro-aggregated data mainly coming from national representative surveys and administrative tax records (for further details see www.oecd.org/sti/ind/multiprod.htm and www.oecd.org/sti/rd-tax-stats.htm).

The advantages of using harmonised micro-aggregated data from business and establishment registers for the study of business and employment dynamics are manifold. First, micro-aggregating these data provides a unique picture on the universe of businesses or establishments, which is rarely provided by any other micro-data source. Second, the cross-country use of business registers allows separate identification of the different channels of employment growth, distinguishing between gross job creation and gross job destruction, and between the extensive (firm entry and exit) and the intensive (post-entry growth) margins of employment growth. Furthermore, the role of firm age and size can be examined separately and jointly. Finally, each of these elements can be compared across countries, sectors and over time.

The extension of the DynEmp methodology to the analysis of subnational employment and business dynamics at the plant level is the focus of the DynEmp Regional project presented in this chapter.

DynEmp Regional: Methodological challenges and solutions

The international comparison of employment dynamics at the regional and local levels involves a considerable number of methodological and conceptual challenges. An overview of these, together with the approaches taken by the DynEmp Regional project to face them, is discussed more in detail in this section and in Annex 5.A1, while in this section a brief overview is presented.

These challenges combine more general issues associated with the construction of internationally comparable employment and business dynamics indicators, building upon the experience developed at the OECD in the framework of the DynEmp project, with concerns specific to the local and regional analyses. These include the choice of the statistical unit of reference (firms vs. plants), measurement issues – such as the identification of entrants, the choice of the relevant employment variable, the measurement of location – and confidentiality challenges.

Unit of analysis

Analysing employment dynamics at the local level requires the availability of information on the location of economic activity. The statistical unit of reference at which the analysis is carried out is particularly important in this context.

A discussion on the challenges in the choice of the unit of analysis, and the bias induced when measuring location at the firm level and not at the plant level, is discussed in OECD (2016b).[4] Given these issues, in order to minimise the above described bias, the DynEmp Regional analysis is carried out at the plant level, exploiting the information available in the data on the plants' location.

In order to effectively design and implement the DynEmp Regional routine, the first step was to conduct a questionnaire to assess data availability. The questionnaire was sent to a considerable number of experts in OECD member and non-member countries that are part of the DynEmp and MultiProd networks. Relevant responses to this questionnaire, with particular reference to the availability of plant-level data with location information, are summarised in Table 5.1.

As is evident from the table, the scope of the analysis is limited to those countries for which suitable plant-level data including location information are available.

Out of 24 countries contacted, plant-level data suitable for the analysis of employment dynamics at the local level seem available for 12 countries. For the majority of these countries (7 out of 12), these data include information on the postal codes of plants, and for 3 additional countries – Brazil, France and Japan – postal code information can be retrieved by exploiting additional data sources or correspondence tables. The two remaining countries, Denmark and Spain, for which plant-level data for local employment analysis are available, report location information only at higher levels of aggregation (such as TL/NUTS regions).

Table 5.1. **DynEmp Regional questionnaire synopsis**

Country	Plant-level data	Location information	Postal code information
Australia	No	..	..
Austria	..	..	..
Belgium	No	..	..
Brazil	Yes	Yes	Possible merge
Canada	Yes	Yes	Yes
Costa Rica	Yes	Yes	Yes
Denmark	Yes	Yes	No
Finland	Yes	Yes	Yes
France	Yes	Yes	Possible merge
Germany	Yes	Yes	Yes
Hungary	No	..	..
Italy	..	..	..
Japan	Yes	Yes	Possible merge
Luxembourg	Partial	..	..
Netherlands	Partial	Yes	Yes
New Zealand	..	..	..
Norway	..	..	..
Portugal	..	..	..
South Africa	No	Partial	Partial
Spain	Yes	Yes	No
Sweden	Yes	Yes	Yes
Turkey	No	..	..
United Kingdom	Yes	Yes	Yes
United States	Yes	Yes	Yes

Notes: Responses are recorded at the time of writing this chapter. ..: not available.

Source: Authors' elaboration based on questionnaire responses provided by members of the DynEmp and MultiProd network.

Measurement challenges

A number of measurement challenges need to be tackled when analysing employment and business dynamics across countries. An extensive discussion of some of the main issues in this context, including: 1) the definition of entry and exit; 2) the measurement of location by means of units' postal code; 3) the choice of the employment variable; 4) changes in the sectoral classifications over time, is presented in Annex 5.A1.

More general challenges associated with employment and business dynamics measurement in a cross-country perspective are also discussed in detail in Criscuolo, Gal and Menon (2014, 2015) in the context of the DynEmp project.

Confidentiality challenges

An important challenge when working with highly representative administrative data is the confidentiality of these sources. In this framework, the *dynemp_reg* programme implements a number of strategies to support national experts in participating countries to comply with the national confidentiality requirements.

Aggregation levels for the output databases are designed keeping in mind the trade-off existing between the levels of detail of the information collected and the extent to which this information can be further subjected to confidentiality blanking. In addition,

the *dynemp_reg* programme directly deals with confidentiality when the blank option is specified, performing a simple blanking of cells below a given number of units. Participants to the project are then asked to check the data produced by the programme and to blank cells according to their national confidentiality rules.

Furthermore, all statistics concerning median values are calculated as the average of the five "central" units in the distribution of interest so that no information referring to an individual unit is disclosed. The number of central units can be increased at the request of those participants who may require stricter confidentiality thresholds.

Finally, an additional optional variable is provided to support participants in blanking related to dominance. In particular, if the option dominance is included, the programme also computes additional statistics on the share of employment or turnover accounted for by the biggest *N* unit in the cell, where *N* can be input by the user and is set to one by default.

The DynEmp Regional programme (*dynemp_reg*)

Building upon the discussion of the challenges to the regional extension of DynEmp carried out in the previous section, this section focuses more in detail on the DynEmp Regional routine.

The *DynEmp Regional database* is based on highly representative administrative sources with a longitudinal dimension. This is a unique feature of this data collection. In particular, the input data required to run the *dynemp_reg* programme must be a longitudinal database and should include the universe of local units (establishments or plants) belonging to the sectors included in the data. In the input database, individual units need to be identified by a unique longitudinal local unit identifier (id) that has to be constant over time.

Only five variables are required to run the *dynemp_reg* routine. They include a suitable unit employment variable; the calendar year to which the time-varying variables refer to, the birth year of the unit, the 3-digit sector identifying the main economic activity of the unit, following the ISIC Rev. 4 or NACE Rev. 2 classification, and the postal code of the local unit (see also the discussion in Annex 5.A1.).

The first version of the *dynemp_reg* programme allows micro-aggregation of data along different dimensions. The combination of these dimensions defines a cell, which is the reference unit in a micro-aggregated setting. The aggregation levels considered in the output database are the following (see also Table 5.2 for a more schematic visualisation):

- Level 2: TL2 regions,[5] macro sector[6] (plus the total sector category), size classification and age classification
- Level 3: metropolitan areas, macro sector (plus the total sector category), size classification and age classification
- Level 5: metropolitan areas, units part of a multi-plant vs. single-plant firm, macro sector (plus the total sector category), size classification and age classification.

Regional levels of aggregation are classified following a methodology developed by the OECD Centre for Entrepreneurship, SMEs, Local Development and Tourism (see Annex 5.A1. for further details). The aggregation in *dynemp_reg* is based on an external

correspondence table between postal codes and regional levels of aggregation (TL2 and metropolitan areas, conditional on their availability).

Table 5.2. **Output databases by aggregation level**

Level ref. number	Sector	Region	Metropolitan areas	Multi-plant vs. single plant firm	Group of units	Size	Age
2	7 macro sectors; all	Yes (TL2)	No	No	All	4 classes	2 classes
3	7 macro sectors; all	No	Yes	No	All	4 classes	2 classes
5 (optional)	7 macro sectors; all	No	Yes	Yes	All	4 classes	2 classes

Source: DynEmp Regional database, http://www.oecd.org/sti/ind/dynemp.htm.

Incumbents are classified into different age and size categories. The size classes considered in aggregation Levels 2, 3 and 5 are: 0-9; 10-49; 50-249; 250+; and 99 (missing).[7] The age classes considered in aggregation Levels 2, 3 and 5 are: 0-5; 6+; and 99 (missing). A discussion on the definition of entry and exit is carried out in Annex 5.A1. A considerable number of statistics are collected at these levels of aggregation, including average employment growth, total employment and gross job flow indicators.

Data description: Sources and coverage

This section focuses on the description of the composition of the preliminary version of the *DynEmp Regional database*. It presents its coverage, main underlying micro-data sources and the available output databases at the time of writing.

As discussed above, data required to run the *dynemp_reg* programme consist of the population (or highly representative) plant-level databases with information on the location of the plant.

Table 5.3 summarises the data sources for countries currently included in the *DynEmp Regional database*. It must be emphasised that the coverage is country-specific. The reader should bear this in mind when comparing descriptive results across countries.

Table 5.3. **Underlying micro-data sources and available output levels**

Country	Source(s)	Period	Geographical variable(s) used	Available output data
Costa Rica	Registro de variables económicas – REVEC	2005-15	Postal code	Level 2
Finland	Business Register on Establishments	2000-15	Postal code	Level 2
France	DADS; Fichier annuel de démographie d'entreprise (créations, transferts et stocks d'établissements)	2005-13	"Code commune" matched with postal code	Level 2, 3, 5
Sweden	RAMS	2000-15	Postal code	Level 2, 3

Note: Data for some countries are still preliminary.

Source: DynEmp Regional database (accessed in February 2017), http://www.oecd.org/sti/ind/dynemp.htm.

Table 5.4 summarises the number of TL2 regions and metropolitan areas by country, based on the preliminary version of the *DynEmp Regional database* available at the time of writing.

Table 5.4. **TL2 regions and metropolitan areas by country**

Country	Number of TL2 regions	Number of metropolitan areas
Costa Rica	6	..
Finland	5	..
France	22	15
Sweden	8	3

Notes: Data for some countries are still preliminary. Only TL2 regions or metropolitan areas with non-missing information are reported. TL2 regions for France exclude DOM and TOM (oversea territories), only the largest 15 metropolitan areas are considered.

Source: DynEmp Regional database (accessed in February 2017), http://www.oecd.org/sti/ind/dynemp.htm.

Preliminary evidence from four countries

This section contextualises the analysis in the framework of recent OECD research and shows some preliminary evidence on employment and business dynamics at the regional and local level, based on the data available at the time of writing. It also discusses the challenges and limitations of the approach proposed.

Previous evidence from the DynEmp project (see Criscuolo, Gal and Menon [2014]; Calvino, Criscuolo and Menon [2015], [2016]; Blanchenay et al. [2017]) has documented significant heterogeneity across countries along different dimensions.

First, this evidence points to the crucial role of start-ups and young firms (rather than small firms in general) for job creation across a significant number of OECD and non-OECD countries. Second, policies and structural factors, especially in the field of access to finance, bankruptcy regulations and contract enforcement, are shown to have an important role in explaining these cross-country differences, and the heterogeneity in the performance of entering versus incumbent firms (Calvino, Criscuolo and Menon, 2016; see also Adalet McGowan, Andrews and Millot [2017] for evidence from different data sources). Finally, this evidence illustrates a steady decline in business dynamism across countries, which markedly accelerated during the crisis with later resilience that has been only partial.[8] Countries and sectors, however, show a fair degree of heterogeneity. For instance, the high-tech services sectors show a particularly pronounced decline in business dynamism (Blanchenay et al., 2017; see also OECD [2017]).

The granularity of the *DynEmp Regional database* allows focusing on within-country differences along some of the dimensions discussed above.

A number of figures and regression tables are presented and discussed below. Given the preliminary nature of the *DynEmp Regional database*, these figures and tables are to be interpreted with caution and should be seen as a first exploration of this source of cross-region cross-country information on regional and local employment dynamics, and not yet as an exhaustive analysis that fully exploits the potential of the *DynEmp Regional database*.

A considerable part of the analysis focuses on the dynamics of plants with less than 250 employees. This is consistent with the important role of young units (that are predominantly small) highlighted by recent OECD research and confirmed by the analysis below. This focus appears also interesting from a regional perspective. As a matter of example, in European countries small and medium enterprises (SMEs) and their competitiveness are a priority for the European Union's Cohesion Policy and part of the 11 Thematic Objectives for 2014-20 (see European Commission [2016]).[9] Finally, excluding cells associated with larger plants somehow reduces the potential issues introduced

by stringency and differences in confidentiality blanking rules across countries. See below for further discussion.

Preliminary analysis at the TL2 level

Figure 5.1 presents the within-region employment distribution, focusing on the share of 0-9, 10-49 and 50-249 plants with respect to all small and medium plants in the country-region, focusing separately on manufacturing and non-financial market services. The panels presented in Figure 5.1 allow a qualitative overview of the size distribution of plants within regions.

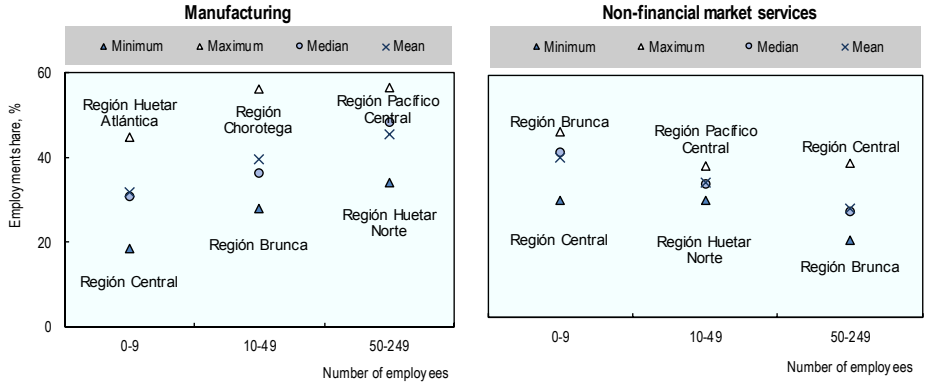
The share of medium sized units (with 50-259 employees) in the non-financial market services sector is significantly higher in the TL2 region where the capital city is located. The share of smaller units, instead, appears to some extent higher in more peripheral regions.

Still focusing on plants with less than 250 employees, Figure 5.2 shows the heterogeneity in average employment growth rates of incumbents, by size class across TL2 regions over the available years. A significant degree of both within-country and cross-country differences is evident. While it is not easy to determine a clear trend in Costa Rica, in Finland (and to some extent in the French manufacturing sector) the smallest plants appear to experience faster employment growth, both in manufacturing and in non-financial market services. Sweden seems to experience instead opposite dynamics, especially in the services sector.

Figure 5.1. **Regional employment shares by plant size (small and medium units)**

Significant regional variation in the employment weight of micro, small and medium-sized plants

A. Costa Rica, 2005-15



B. Finland, 2000-15

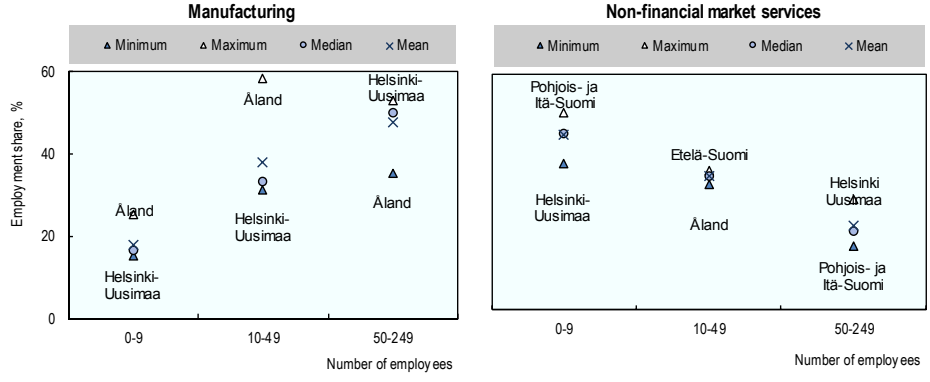
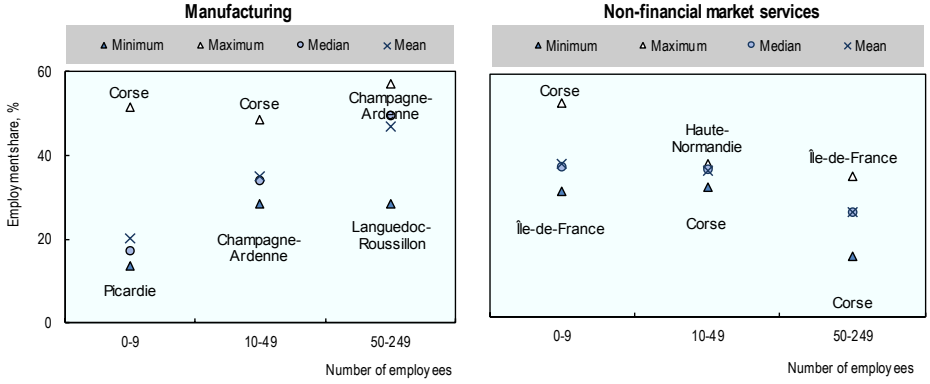


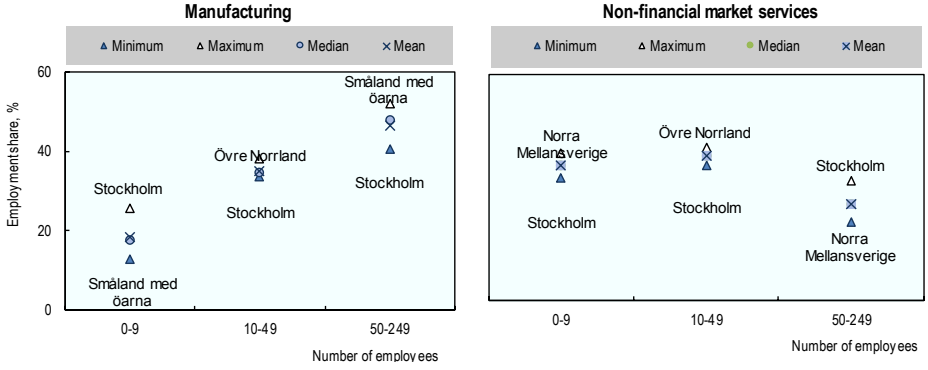
Figure 5.1. **Regional employment shares by plant size (small and medium units)** (*continued*)

Significant regional variation in the employment weight of micro, small and medium-sized plants

C. France, 2005-13



D. Sweden, 2000-15



Notes: For each country, employment shares are calculated as employment in the macro sector, size class (0-9, 10-49, 50-249 employees), TL2 region over employment of all plants with less than 250 employees in the TL2 region, macro sector. Statistics are computed on average over available years. Owing to methodological differences, figures may deviate from officially published national statistics. Data for some countries are still preliminary.

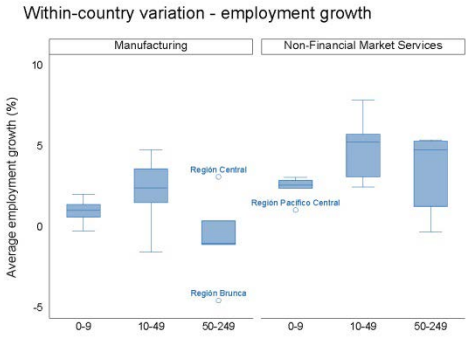
Source: DynEmp Regional database (accessed in April 2017).

StatLink http://dx.doi.org/10.1787/888933626440

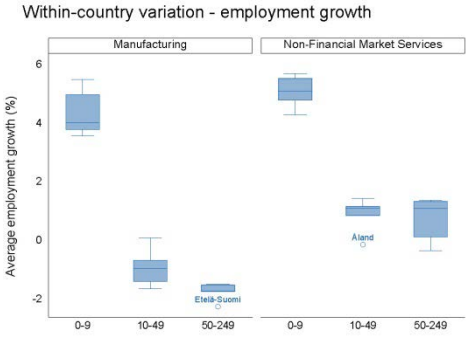
Additional analysis was carried out for Finland, where turnover growth is also available, and presented in Figure 5.3. Focusing on turnover, small units have a more limited growth performance in most TL2 regions, especially in the non-financial market services sector. Units in non-financial market services appear to have generally higher turnover growth rates when compared to plants in the manufacturing sector. This is particularly true for medium units (with 50-249 employees).

Figure 5.2. **Within-country variation in small and medium units' employment growth**

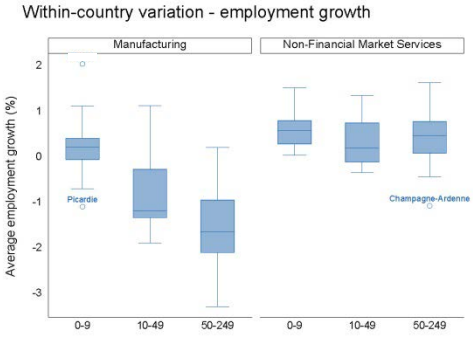
A. Costa Rica, 2005-15



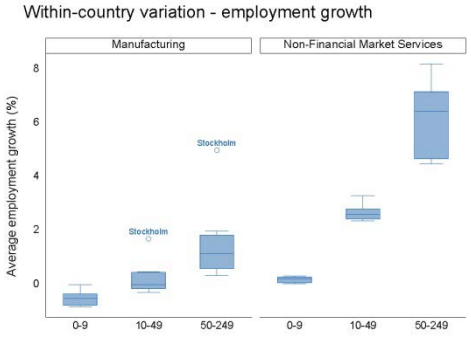
B. Finland, 2000-15



C. France, 2005-13



D. Sweden, 2000-15



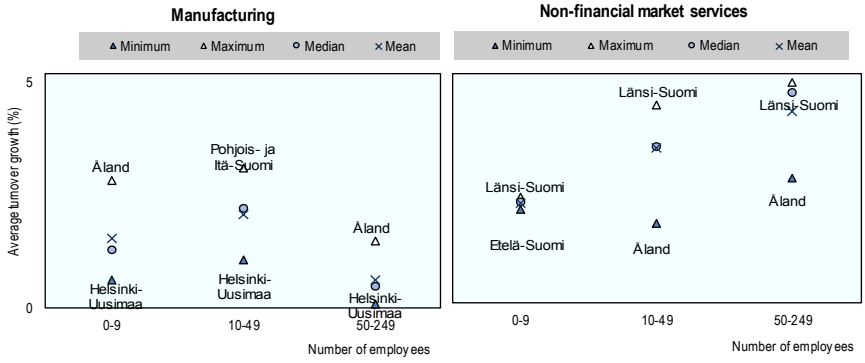
Notes: Average employment growth of incumbents in the country, macro sector, size class (0-9, 10-49, 50-249 employees), TL2 region. The box plot shows the within-country heterogeneity across TL2 regions. Statistics are computed on average over the available years. Owing to methodological differences, figures may deviate from officially published national statistics. Data for some countries are still preliminary.

Source: DynEmp Regional database (accessed in February 2017), http://www.oecd.org/sti/ind/dynemp.htm.

StatLink http://dx.doi.org/10.1787/888933626459

Figure 5.3. **Within-country variation in small and medium units' turnover growth**

Finland, 2000-15



Notes: Average turnover growth of incumbents in the country, macro sector, size class (0-9, 10-49, 50-249 employees), TL2 region. Statistics are computed on average over available years (see Table 5.3 for detailed coverage). Owing to methodological differences, figures may deviate from officially published national statistics. Data for some countries are still preliminary.

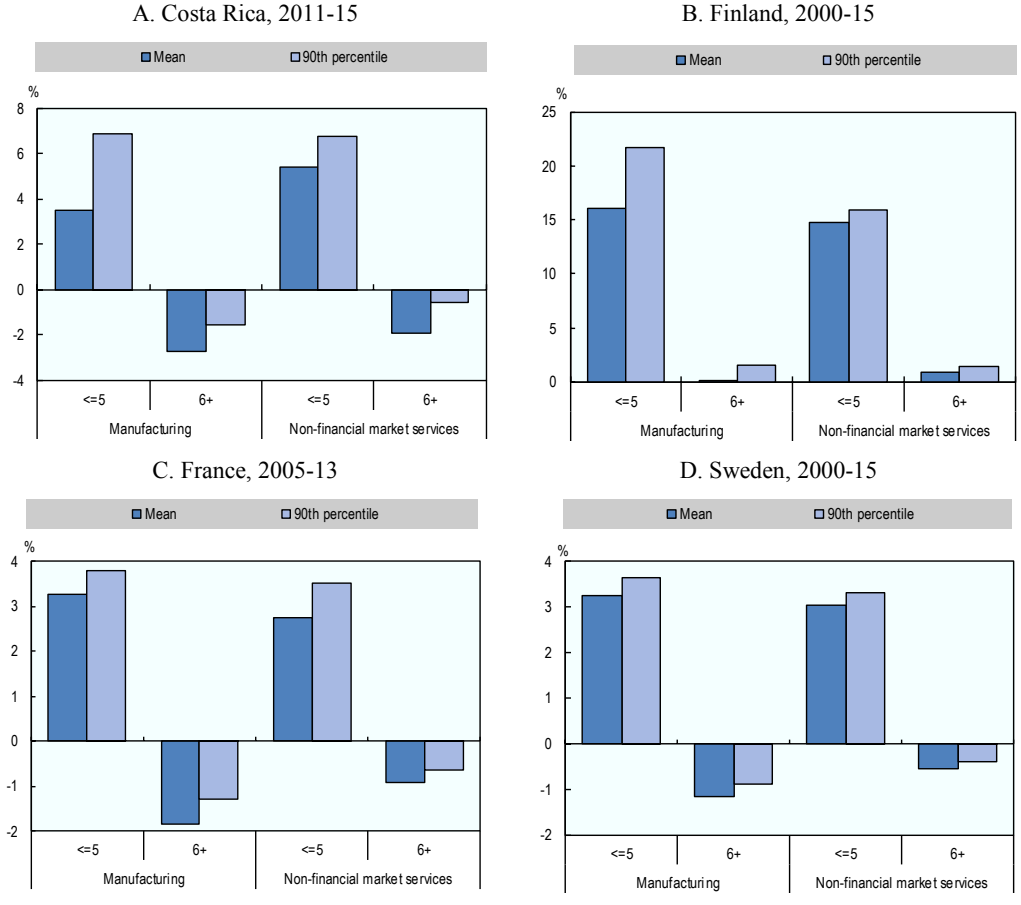
Source: DynEmp Regional database (accessed in April 2017), http://www.oecd.org/sti/ind/dynemp.htm.

StatLink http://dx.doi.org/10.1787/888933626478

Figure 5.4 disentangles instead the different employment growth performances of young (five years old or less) versus older small and medium incumbent plants, across TL2 regions. Statistics are reported for the average TL2 region in the country-macro sector and for the TL2 region at the 90th percentile in terms of employment growth performance. The figure confirms that, in all countries and macro sectors considered, younger small-medium plants grew faster than older ones, consistently with previous evidence from the DynEmp project at the national level (see Criscuolo, Gal and Menon [2014]). Focusing on the difference between the average TL2 region and the region at the 90th percentile of the employment growth distribution allows a visualisation of within-country differences in employment growth profiles of incumbents. Within-country differences between the mean and the 90th percentile are evident, especially in Costa Rica. Furthermore, older small-medium plants in Finland seem to experience a positive (despite very limited) employment growth, while this is not the case for Costa Rica, France or Sweden.

Figure 5.4. **Employment growth in small and medium units, young vs. old**

Within country variation in employment growth



Notes: Average employment growth of incumbents in the country, macro sector, age class (five years old or less; six years old or more), TL2 region. Statistics are reported for the average TL2 region in the country-macro sector and for the TL2 region at the 90th percentile. They are computed on average over available years. Complete information on unit age in Costa Rica is available starting from 2011. Owing to methodological differences, figures may deviate from officially published national statistics. Data for some countries are still preliminary.

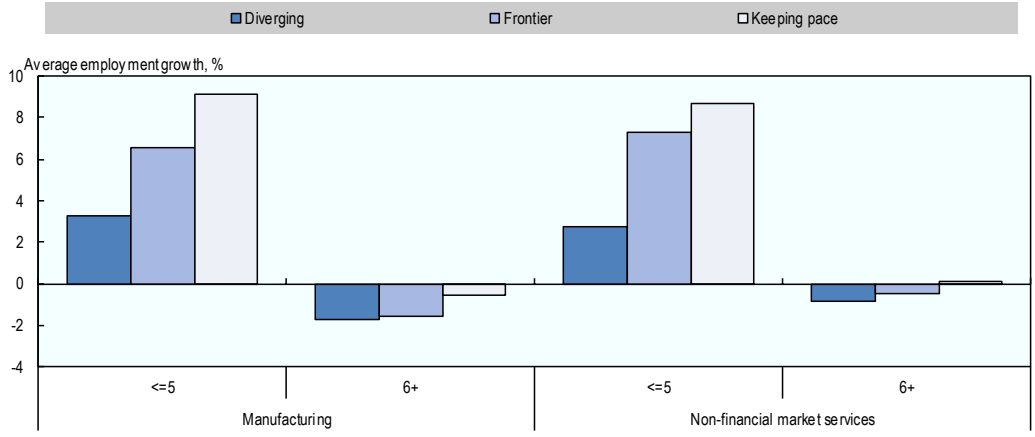
Source: DynEmp Regional database (accessed in February 2017), http://www.oecd.org/sti/ind/dynemp.htm.

StatLink http://dx.doi.org/10.1787/888933626497

The evidence that young firms grow faster on average is confirmed when performing a more formal exercise, regressing cell-level average employment growth on age class, size class, macro sector, country, TL2 region and year dummies.[10] This exercise shows that older units (six years old or more) experience a significantly lower employment growth (about 6% less) with respect to younger plants, conditional on other cell and region characteristics. However, since this simple framework does not allow inferring any causal relation, this should just be interpreted as a robust association.[11]

Figure 5.5 extends the analysis on the employment growth patterns of small and medium units separately focusing on frontier, keeping pace and diverging regions.[12] In the figure, the bars refer to the average region in each age class, type of region, macro sector. For instance, the first bar focuses on the average employment growth of young small and medium units in diverging TL2 regions in the manufacturing sector, the second focuses on frontier regions, and the third on keeping pace regions. The bars plot averages across the TL2 regions in each group. Young small and medium incumbents (less than five years old) are again characterised by higher average employment growth with respect to their older counterparts. In both manufacturing and non-financial market services, keeping pace regions exhibit higher average employment growth, especially when focusing on young units.

Figure 5.5. **Employment growth in small and medium units, young vs. old/diverging, keeping pace and frontier regions**



Notes: Average employment growth of incumbents in the macro sector, age class (five years old or less; six years old or more), type of TL2 region (frontier, keeping pace and diverging). Average values across TL2 regions in the different types are reported. They are computed on average over available years (see Table 5.3 for detailed coverage). Costa Rica is excluded because the productivity classification is not available. Owing to methodological differences, figures may deviate from officially published national statistics. Data for some countries are still preliminary.

Source: DynEmp Regional database (accessed in April 2017), http://www.oecd.org/sti/ind/dynemp.htm.

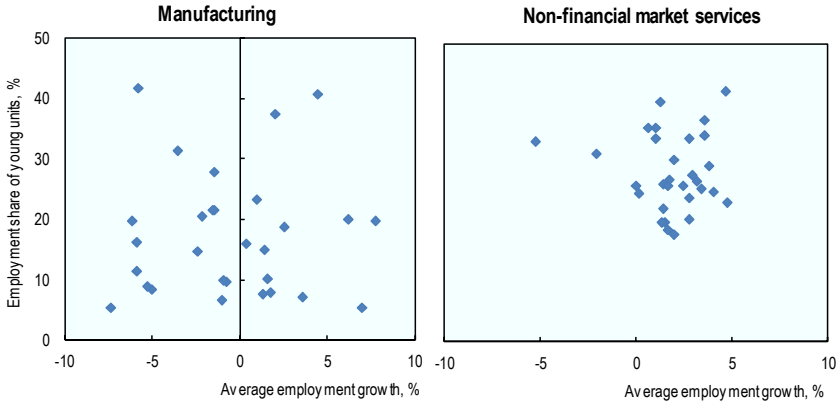
StatLink http://dx.doi.org/10.1787/888933626516

Building upon the descriptive analysis presented so far, Figure 5.6 investigates the yearly correlation between the employment share of young (five years old or less) units and the overall average employment growth, across different TL2 regions by macro sector of activity (manufacturing and non-financial market services). Here the attention is no longer restricted to small and medium units, as was done in some of the analysis presented above. As shown in Figure 5.2, non-financial market services tend to have higher employment growth rates across TL2 regions in different countries. Furthermore,

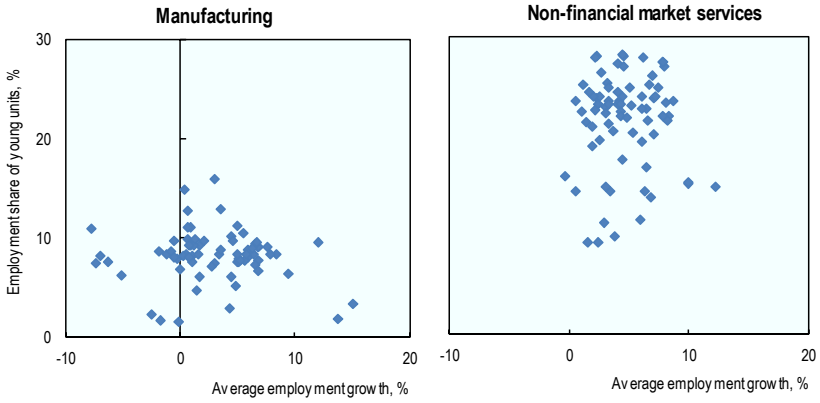
more evidently, they also tend to have higher shares of young units across regions. From this simple chart there appear significant differences in the share of young units between manufacturing and non-financial market services, but it is not straightforward to evince a clear relationship between share of young units and average employment growth within sectors.

Figure 5.6. **Share of young units and average employment growth**

A. Costa Rica, 2011-15



B. Finland, 2000-15



C. France, 2005-13

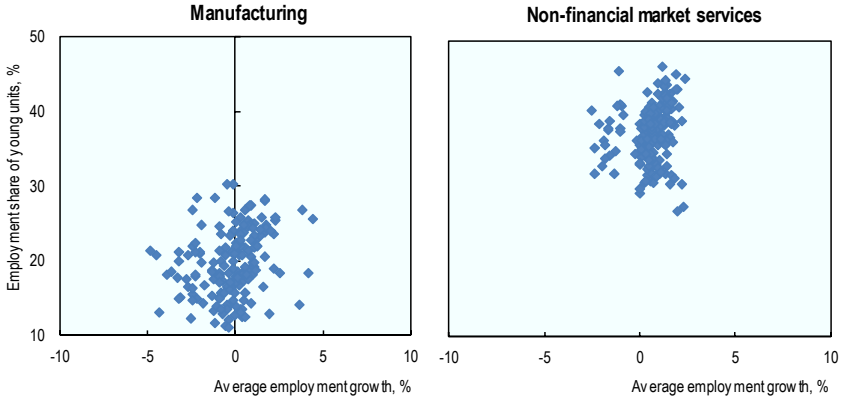
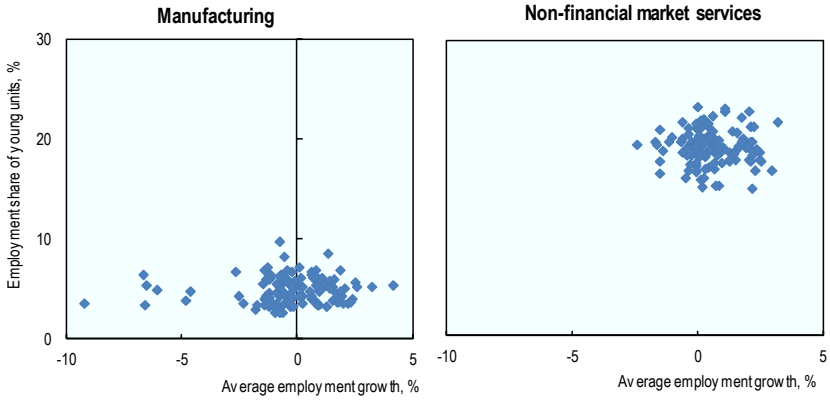


Figure 5.6. **Share of young units and average employment growth** (*continued*)

D. Sweden, 2000-15



Notes: Scatter plot illustrating the correlation of the employment share of young (five years old or less) units in the country, year, macro sector, TL2 region and the average employment growth of the country, year, macro sector, TL2 region. Complete information on unit age in Costa Rica is available starting from 2011. Owing to methodological differences, figures may deviate from officially published national statistics. Data for some countries are still preliminary.

Source: DynEmp Regional database (accessed in February 2017), http://www.oecd.org/sti/ind/dynemp.htm.

StatLink http://dx.doi.org/10.1787/888933626535

Further econometric analysis has been carried out in order to investigate this issue more in depth, with mixed results. In particular, when regressing the average employment growth rate of a given TL2 region-macro sector-year on the (employment) share of young units in the previous year (in the same region and macro sector) controlling for macroeconomic shocks (using year dummies) and unobserved regional characteristics (using TL2 fixed effects), a small positive and statistically significant coefficient (0.06) results from the analysis.[13] However, this correlation becomes statistically insignificant once the econometric model accounts for unobserved sectoral characteristics.[14]

Additional regression analysis was carried out to investigate the relationship between the share of employment in young units in a TL2 region and its characteristics. In particular, we looked at the relationship between the share of employment in young units and: 1) whether a region is mostly rural, mostly urban or intermediate; 2) whether a region is a frontier region, catching-up, diverging or keeping pace region; 3) both regional characteristics combined (see OECD [2016a] for additional details on these classifications).

Finally, we tried to proxy for the role of agglomeration economies for the development of an entrepreneurial economy, by including in these regressions two proxies for agglomeration. The first is population density and the second is "plant density", i.e. number of plants per square meter in the TL2 region.

Therefore, the following model is estimated, separately looking at the share of employment in young units in the manufacturing and non-financial market services sectors:

$$y_{c,r,t} = char_{c,r} + agglomeration_{c,r,t} + year_t + country_c + \epsilon_{c,r,t}\,, \quad (1)$$

where y is the share of employment in young units; *char* indicates a set of regional characteristics dummies; *country* and *year* are a set of country and year dummies; the subscript c indicates country, r TL2 regions and t time; *agglomeration* is either the

logarithm of population density or of plant density,[15] which is added as a control after estimating the baseline models. Results are reported in Table 5.5 (Panel A focuses on manufacturing and Panel B on non-financial market services, reference categories are mostly rural regions and diverging regions).[16]

Table 5.5. **Share of employment in young units and regional characteristics**

A. Manufacturing sector

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intermediate	0.656		0.563	0.385	0.102			0.744	0.645
	(0.454)		(0.507)	(0.545)	(0.579)			(0.557)	(0.589)
Mostly urban	1.323***		0.133	0.944*	0.487			0.324	0.226
	(0.418)		(0.517)	(0.500)	(0.569)			(0.552)	(0.578)
Frontier		3.322***	3.446***			3.426***	3.230***	3.691***	3.577***
		(0.669)	(0.855)			(0.695)	(0.737)	(0.856)	(0.875)
Keeping pace		0.545	0.785*			0.527	0.562	0.768*	0.769*
		(0.378)	(0.454)			(0.390)	(0.394)	(0.455)	(0.458)
Log plant density				0.257		-0.0584		-0.197	
				(0.210)		(0.154)		(0.168)	
Log population density					0.462**		0.0410		-0.0819
					(0.218)		(0.161)		(0.189)
Observations	356	356	356	356	356	356	356	356	356
R-squared	0.840	0.850	0.851	0.841	0.842	0.850	0.850	0.851	0.851

B. Non-financial market services sector

	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Intermediate	2.994***		2.753***	2.013***	1.982***			2.363***	2.369***
	(0.364)		(0.406)	(0.473)	(0.469)			(0.507)	(0.498)
Mostly urban	3.102***		1.817***	1.669***	1.576***			1.419***	1.378***
	(0.338)		(0.406)	(0.423)	(0.420)			(0.438)	(0.438)
Frontier		3.929***	3.402***			1.814***	1.828***	2.694***	2.788***
		(0.512)	(0.691)			(0.563)	(0.591)	(0.750)	(0.749)
Keeping pace		-0.643**	0.264			-0.287	-0.258	0.322	0.338
		(0.321)	(0.385)			(0.333)	(0.342)	(0.396)	(0.400)
Log plant density				0.795***		0.881***		0.383**	
				(0.166)		(0.138)		(0.176)	
Log population density					0.844***		0.933***		0.385**
					(0.178)		(0.158)		(0.188)
Observations	356	356	356	356	356	356	356	356	356
R-squared	0.921	0.916	0.929	0.926	0.926	0.923	0.923	0.929	0.929

Notes: The reference categories are mostly rural regions and diverging regions. Data on regional characteristics are not available for Costa Rica, which is excluded from the estimation sample. All regressions include year and country dummies. Robust standard errors in parentheses. The tables do not report the regression constants. *** p<0.01, ** p<0.05, * p<0.1.

Focusing on the results for the share of employment in young units, which should proxy for the presence of entrepreneurs in the region, and on the rural-urban gap, the estimates suggest that the share of employment in young units is significantly higher in the mostly urban TL2 regions, even if in the manufacturing sector this effect tends to disappear once the level of agglomeration in the region is accounted for. When focusing on the results based on the productivity classification, frontier regions appear to have a

higher share of employment of young units with respect to diverging regions, both in manufacturing and non-financial market services. When combining the two regional classifications, results appear stable for the non-financial market services sector, while only the regional productivity dummies remain significant in the manufacturing sector. Proxies for agglomeration (population density or plant density) are always positive and significant in non-financial market services.[17]

Additionally, we re-estimated equation (1) using as dependent variables the average employment growth rate of young units and the relative net job creation by entering units.[18] Results are reported in Annex 5.A1 (Tables A.1 and A.2) and provide qualitatively similar insights, slightly weaker when focusing on employment growth rates (in line with the descriptive evidence reported in Figure 5.5).

In a similar spirit, we focused on the relationship between regional innovation and employment in young units. Firstly, we estimated equation (1) replacing the regional characteristics dummies with a proxy of regional R&D intensity at TL2 level (total R&D expenditure over total employment in the region, persons aged 15-64). Results, reported in Table 5.6 (Panel A), suggest that high regional innovativeness, proxied by our R&D intensity indicator, is associated with a higher share of employment in young units.[19] Focusing on the manufacturing sector and changing the innovativeness proxy (using regional patent stock) produces similar results (Table 5.6, Panel B). However, when including both innovation indicators, only R&D intensity remains significant.

Table 5.6. **Share of employment in young units and regional innovation**

A. R&D in manufacturing and non-financial market services

	(1)	(2)	(3)	(4)
	Manufacturing		Non-financial market services	
Log R&D intensity	0.159***	0.161***	0.146***	0.142***
	(0.0378)	(0.0436)	(0.00985)	(0.0116)
Log population density		-0.00296		0.00804
		(0.0204)		(0.00667)
Observations	273	273	273	273
R-squared	0.843	0.843	0.934	0.934

B. R&D and patents in manufacturing

	(5)	(6)	(7)	(8)
	Manufacturing			
Log R&D intensity			0.147**	0.146**
			(0.0664)	(0.0656)
Log patents	0.0756***	0.0845***	0.00392	0.00553
	(0.0198)	(0.0273)	(0.0306)	(0.0344)
Log population density		-0.0202		-0.00328
		(0.0214)		(0.0231)
Observations	347	347	272	272
R-squared	0.851	0.852	0.841	0.841

Notes: Dependent variables in log. Data on regional characteristics are not available for Costa Rica, which is excluded from the estimation sample. All regressions include year and country dummies. Robust standard errors in parentheses. The tables do not report the regression constants. *** p<0.01, ** p<0.05, * p<0.1.

Secondly, we looked at whether a higher share of employment in young units at time *t-1* is somehow associated with a higher patent stock in the region, at time *t*. Preliminary

results from this type of exercise suggest the existence of a positive link between the stock of young units in a region and its innovation output.[20]

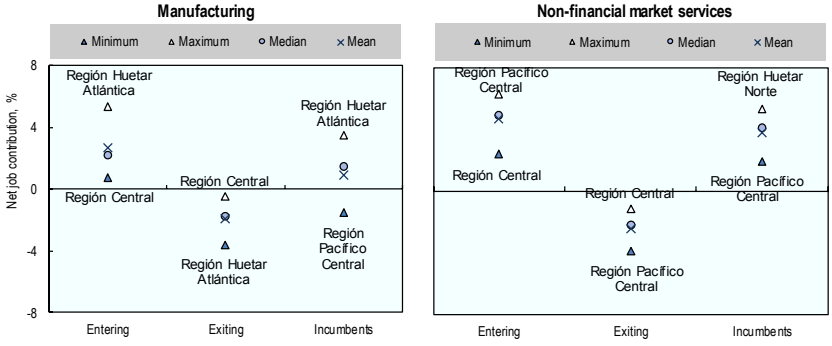
Further analysis oriented at corroborating these findings, which are at the moment simple associations with no causal interpretation, would be an interesting avenue for further research.

The granularity of the *DynEmp Regional database* was further exploited, disaggregating the data by groups of units (entering, incumbents and exiting). Along these lines, Figure 5.7 focuses on the relative net job creation by different groups of plants in TL2 regions. The measure represents net job creation in the TL2 region by the group of plants normalised by the total employment in the TL2 region, by macro sector of activity. Similarly to Figure 5.6 and to the regression analysis, this figure does not restrict anymore the focus on small and medium units.

The figure shows that, for instance, in the Finnish non-financial market services sector (Panel B), for every 100 jobs in the average TL2 region, about 1 new job is created by incumbents, about 3 by entrants and about 4 jobs are destroyed by exiting plants.

Figure 5.7. **Relative net job creation in TL2 regions by group**

A. Costa Rica, 2005-15



B. Finland, 2000-15

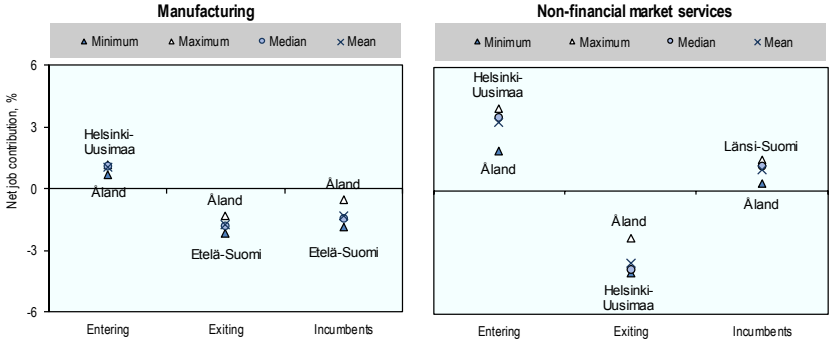
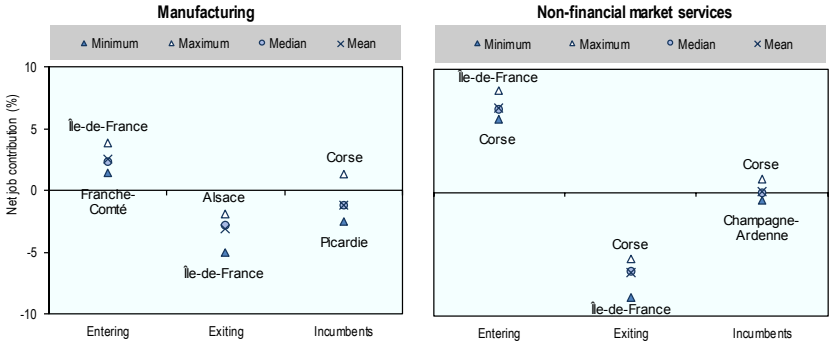
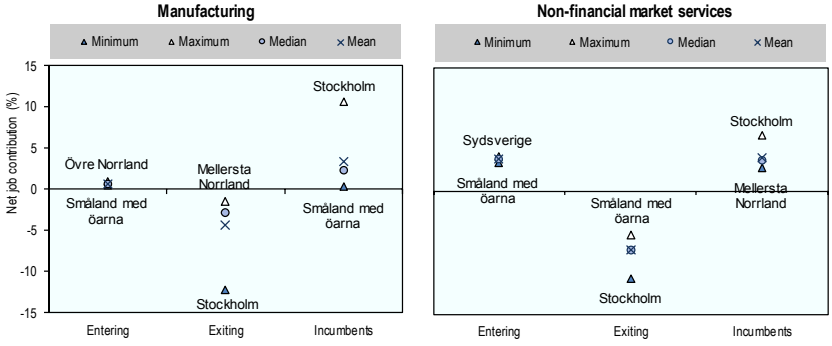


Figure 5.7. **Relative net job creation in TL2 regions by group** (*continued*)

C. France, 2005-13



D. Sweden, 2000-15



Notes: Relative net job creation is calculated as net job creation in the country, TL2 region, macro sector, group of plants (entering, incumbents and exiting) over average total employment (between time *t* and *t*-1) in the country, macro sector, TL2 region. The mean, median, top and bottom TL2 regions are shown. Figures are computed on average over available years. Owing to methodological differences, figures may deviate from officially published national statistics. Data for some countries are still preliminary.

Source: DynEmp Regional database (accessed in April 2017).

StatLink http://dx.doi.org/10.1787/888933626554

Interestingly, looking at sectoral dynamics of the average TL2 regions, non-financial market services confirm to be more dynamic than the manufacturing sector in all countries under scrutiny. Furthermore, another interesting regularity appears to be the fact that – still looking at average regions – entering units, especially in non-financial business services, generally tend to outperform incumbents, more importantly in the TL2 regions of the capital city in Finland and France. This does not clearly occur in Sweden.[21]

Focusing on within-country sector figures, France seems to show more homogeneous dynamics, with more limited differences between the TL2 region at the top and at the bottom of the relative net job creation distribution in each given group of plants. Other countries, in particular Costa Rica and to a lesser extent incumbent and exiting units in the Swedish manufacturing sector, seem instead to exhibit more heterogeneous dynamics, which would be hidden looking at the average region only.

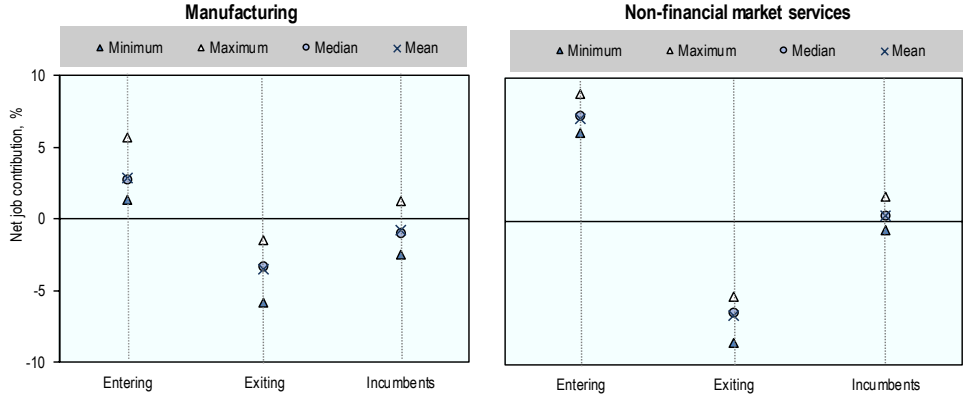
Descriptive analysis for metropolitan areas

Whenever data availability allowed, we extended part of the previously presented analysis shifting the focus from TL2 regions to metropolitan areas. In particular, Figures 5.8 and 5.9 focus on the contribution to net job creation occurring at the level of metropolitan areas. This output database is available only for France and Sweden.

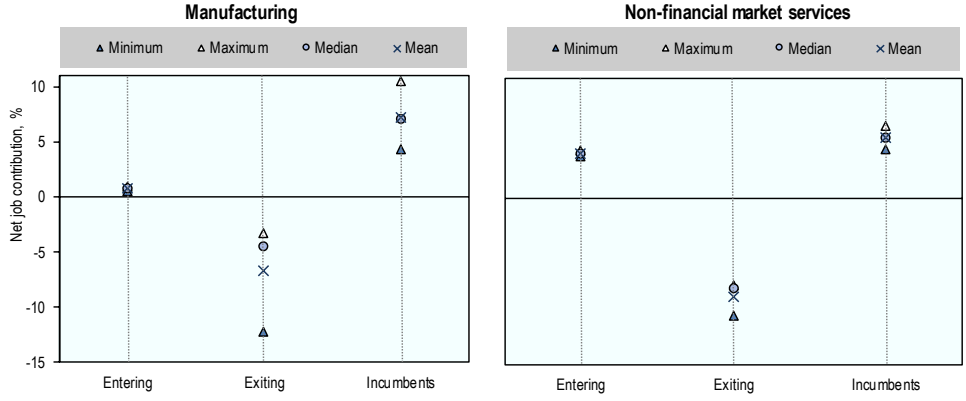
Similarly to Figure 5.7, Figure 5.8 focuses on the relative net job creation by different groups of plants (entering, incumbents and exiting) in metropolitan areas. The measure represents here net job creation in the metropolitan area by the group of plants normalised by the total employment in the metropolitan area, by macro sector of activity.

Figure 5.8. **Relative net job creation in metropolitan areas by group**

A. France, 2005-13



B. Sweden, 2000-15



Notes: Relative net job creation is calculated as net job creation in the country, metropolitan area, macro sector, group of plants (entering, incumbents and exiting) over average total employment (between time *t* and *t*-1) in the country, macro sector, metropolitan area. The mean, median, top and bottom metropolitan areas are shown. Figures are computed on average over available years. Owing to methodological differences, figures may deviate from officially published national statistics. Data for some countries are still preliminary.

Source: DynEmp Regional database (accessed in February 2017), http://www.oecd.org/sti/ind/dynemp.htm.

StatLink http://dx.doi.org/10.1787/888933626573

The figure shows that, for instance, in the Swedish manufacturing sector, for every 100 jobs, in the average metropolitan area about 7 new jobs are created by incumbents, less than 1 by entrants and about 7 jobs are destroyed by exiting plants. In the top

metropolitan area, instead, more than ten new jobs are added by incumbent plants in the manufacturing sector. Different dynamics seem evident when focusing on metropolitan areas in France, where entrants both in manufacturing and non-financial market services exhibit a relative net job creation higher than incumbent plants.

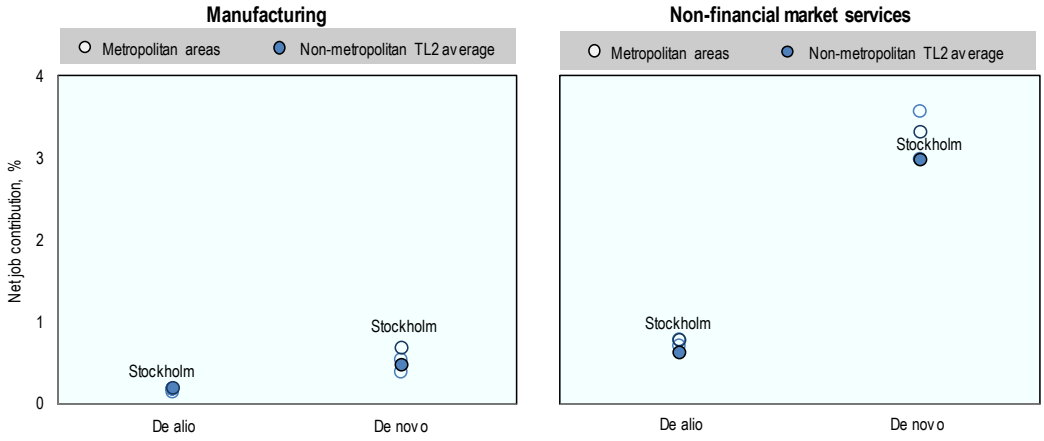
For France, additional unreported analysis has allowed to distinguish the relative net job creation by entering and exiting units separating plants part of a single-plant firm from those that are part of a multi-plant firm. Overall, on average, plants that are part of single-plant firms appear responsible for most of the dynamics observed in the overall figures, especially in the manufacturing sector. Additional analysis along these lines would be an interesting avenue for future research.

Finally, *de novo* entrants appear to contribute more significantly to net job creation in the metropolitan areas, especially in the non-financial market services sector.[22] This is reported in Figure 5.9, which focuses on the role of entering plants and further analyses their relative net job creation in metropolitan areas by disentangling *de novo* and *de alio* entrants in Sweden, where this further level of aggregation is available.

Comparing these figures to the same measures computed on "non-metropolitan" TL2 regions (see the figure note for details) illustrates that *de novo* entrants in non-financial market services have a higher relative contribution to net job creation in two out of three metropolitan areas. The magnitude of the differences seems less relevant in the manufacturing sector.

Figure 5.9. **Relative net job creation in metropolitan areas – *de novo* vs. *de alio* entry**

Sweden, 2000-15



Notes: Relative net job creation (reported in white) is calculated as net job creation in the country, metropolitan area, macro sector, group of plants (*de novo* vs. *de alio* entering units) over average total employment (between time *t* and *t*-1) in the country, macro sector, metropolitan area. Relative average net job creation from TL2 regions that have less than 10% of zip codes associated with a metropolitan area (labelled "Non-metropolitan TL2s") is reported in blue. Owing to methodological differences, figures may deviate from officially published national statistics. Data for some countries are still preliminary.

Source: DynEmp Regional database (accessed in February 2017), www.oecd.org/sti/ind/dynemp.htm.

StatLink http://dx.doi.org/10.1787/888933626592

Challenges and limitations

Despite the considerable harmonisation efforts, a few challenges and limitations need to be taken into account when interpreting the preliminary findings of the DynEmp Regional project.

Despite the careful choice of the levels of aggregation at which the analysis was carried out, blanking still represents one caveat to consider when analysing the results. This can be particularly relevant for figures that do not restrict the attention to the smallest units in the economies under scrutiny (for instance, Figure 5.7). Furthermore, differences in blanking rules (due to primary or secondary disclosure) need to be further taken into consideration, particularly when comparing small cells across countries (notably those cells including the largest units, especially in scarcely populated sectors or regions).

As discussed in the section "The DynEmp Regional programme", the identification of TL2 regions and metropolitan areas is based on a matching of microdata with an external correspondence table between zip codes and local levels of aggregation. Differences in the quality of such matching across countries need to be kept in mind as an additional caveat when carrying out cross-country analyses using the *DynEmp Regional database*.

From a more methodological perspective, the current structure of the DynEmp Regional programme does not allow taking self-employment into account. This could be an interesting direction for further development of future versions of the routine.

Finally, one of the challenges and important directions for further development is associated with the study of plant transition dynamics. In fact, longitudinal databases on employment dynamics can be used to study the regional employment growth performance of cohorts of plants over time (see for instance Calvino, Criscuolo and Menon [2015] for an example at national level). Due to time constraints, it was not possible to implement a module that computes such micro-aggregated statistics across different countries in the first version of the DynEmp Regional programme.

However, the module has been developed and implemented in one country. Some results for France are reported in Annex 5.A1 (Table 5.A1.3). These are aimed at showcasing the potential of this type of analysis. The results are based on a preliminary version of a transition database that follows cohorts of plants starting in pre-determined years (2001, 2004, 2007 and 2010) for three, five or seven years, conditional on data availability. Data are aggregated along different dimensions, including TL2 regions, macro sectors of activity, years, age classes, size classes at the beginning of each period and size classes at the end of each period.

A simple linear regression model was estimated, in order to investigate how survival (in terms of employment) and post-entry employment growth of entrants is associated with some regional characteristics (see Table 5.A1.3). Results suggest that entrants in mostly urban regions and, to a lesser extent, in mostly intermediate regions, tend to grow faster than entrants in mostly rural regions. However, no statistically significant effect of regional characteristics appears instead detectable when focusing on the survival shares of entrants.[23]

Additional refinements of the DynEmp Regional code, aimed at a cross-country implementation of this type of dynamic analyses, are left to future research.

Conclusions and next steps

This chapter has presented the OECD DynEmp Regional project, a new distributed micro-data project aimed at analysing confidential administrative micro-data on employment dynamics at the regional and local level by means of micro-aggregation.

The results presented in this chapter have documented the heterogeneity in the employment distribution within countries and regions. They have confirmed that young plants contribute disproportionally to net job creation across regions and show on average a higher employment growth performance, as found in national-level analysis. The evidence reported in the chapter also highlights the importance of agglomeration economies for entrepreneurial efforts and post-entry employment growth in non-financial market services. The weight of entrepreneurial firms appear significantly higher in regions at the frontier, suggesting that on average these regions might be also more dynamic, reflecting a stronger creative destruction process.

At the time of writing, the DynEmp Regional project is still at its initial stages. A number of additional refinements might be carried out in the following versions of the code. These might include further customisation of the statistical routine to allow the participation of additional countries (such as, for instance, Denmark and Spain, whose microdata include only broader location information); possible addition of other indicators (possibly including a focus on concentration indices); possible additional refinements aimed at finalising the "Transition Matrix" code that would allow following cohorts of plants over time across countries, focusing on their employment and growth performance and the relation of this on contextual regional characteristics; possible methodological refinements in order to separately account for self-employment if and whenever possible.

Next steps entail matching the indicators of business dynamics at the regional and metropolitan level developed by the code to additional existing databases containing information on different relevant characteristics of TL2 and metropolitan areas, to refine the current analysis and to further understand the drivers of subnational business dynamics and employment growth.

Notes

1. The authors would like to thank Alfonso Alfaro Ureña, Fredrik Andersson, Giuseppe Berlingieri, David Bullon Patton, Isabelle Desnoyers-James, Arlina Gómez, Nick Johnstone, Mika Maliranta, Tayutic Mena, Carlo Menon, Francisco Monge, Joaquim Oliveira Martins, Rudy Verlhac and other members of the DynEmp network (reported in Table 5.A2.1 in Annex 5.A2). Access to French data benefited from the use of the 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).

2. As part of this strand of work, the OECD is leading three projects, DynEmp, MicroBerd and MultiProd, which rely on countries' confidential micro-data to carry out comparable cross-country analysis on employment dynamics, investment in research and development (R&D), and productivity (see Box 5.1).

3. See https://www.oecd.org/fr/sti/dynemp.htm for additional information and details.

4. Analysing local employment dynamics at the firm level is likely to induce a bias, attributing employment of all the plants of a multi-plant firm to the location of the firm's headquarter. This bias depends on the proportion of multi-plant firms in a given economy, and on the distance of the plants that are part of these multi-plant firms from their headquarters.

5. The Territorial Level 2 consists of macro-regions. See the most updated territorial grids for OECD countries at: http://stats.oecd.org/wbos/fileview2.aspx?IDFile=cebce94d-9474-4ffc-b72a-d731fbdb75b9 for further details.

6. This corresponds to the STAN A7 classification in ISIC Rev. 4 (agriculture, forestry and fishing; mining and quarrying; manufacturing; electricity, gas, water and waste; construction; non-financial market services; non-market services).

7. Size is defined on the average of employment at time *t*-1 and *t* for incumbents, on employment at time *t*-1 for exitors, and on employment at time *t* for entrants.

8. Proxies of business dynamism include indicators of new business formation, job flows indicators, churning, etc. See Decker et al. (2016) for additional discussion, mainly focusing on the United States.

9. The SME definition is not identical to what result from our focus on plants with less than 250 employees, but a considerable degree of overlapping is expected. See OECD (2005) for further details.

10. Recall the definition of a cell provided in the section entitled "The DynEmp Regional programme". See Criscuolo, Gal and Menon (2014) for further discussion and analysis at the national level.

11. The regression table is omitted for brevity but available upon request.

12. See OECD (2016a) for additional details on this classification. The group of catching-up regions is excluded as it identifies only one small insular region in Finland (Aland).

13. Here the attention is no longer restricted to units with less than 250 employees.

14. Similar dynamics result when using the second lag of share of young units. Regression tables are not reported for brevity, but they are available upon request. This departs from unreported preliminary analysis that focuses on the relationship between average employment growth and share of young units at the country-industry-year level. This analysis suggests that this relationship remains positive and significant even once accounting for industry unobserved effects. Further exploration of these dynamics appears as a fruitful avenue for further research.

15. Note that population and plant density seem to essentially capture the same thing as they are very highly correlated, with a correlation coefficient above 0.6, and thus cannot be included together in the same regression model.

16. The group of catching-up regions is excluded from estimation as it identifies only one small insular region in Finland (Aland).

17. Unreported additional robustness exercises carried out for one country qualitatively confirm this role of agglomeration also when using a different range of proxies.

18. See the notes to Figure 5.7 for more details on the definition.

19. The results on R&D intensity also appear robust to the inclusion of TL2 regional dummies. Reverse causality issues are limited by the fact that the share of R&D is significantly higher in larger firms (see, for instance, OECD [2015]).

20. We included country and year fixed effects and focused on the manufacturing sector only. These results are preliminary and available upon request.

21. Replicating these figures using a common set of years for all countries (2013) provides qualitatively similar insights for average regions.

22. *De novo* entry corresponds to real economic entry, while *de alio* entry is entry due to a merger, an acquisition or a change in legal status.

23. The baseline category is mostly rural regions. Focusing on other covariates, the 2010 cohort of entrants appears to have experienced higher employment growth with respect to the 2007 one; the length of the time period over which units are observed is, unsurprisingly, negatively associated with survival and positively with growth; non-financial market services consistently appear more dynamic than the manufacturing sector (with higher growth and lower survival of entrants).

References

Adalet McGowan, M., D. Andrews and V. Millot (2017), "The walking dead? Zombie firms and productivity performance in OECD countries", *OECD Economics Department Working Papers*, No. 1372, OECD Publishing, Paris, http://dx.doi.org/10.1787/180d80ad-en.

Blanchenay, P. et al. (2017), "Cross-country evidence on business dynamics over the last decade: From boom to gloom?", *OECD Science, Technology and Industry Working Papers*, OECD Publishing, Paris, forthcoming.

Calvino, F., C. Criscuolo and C. Menon (2016), "No country for young firms?: Start-up dynamics and national policies", *OECD Science, Technology and Industry Policy Papers*, No. 29, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jm22p40c8mw-en.

Calvino, F., C. Criscuolo and C. Menon (2015), "Cross-country evidence on start-up dynamics", *OECD Science, Technology and Industry Working Papers*, No. 2015/06, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jrxtkb9mxtb-en.

Criscuolo, C., P.N. Gal and C. Menon (2015), "dynemp: A routine for distributed microdata analysis of business dynamics", *Stata Journal*, Vol. 15/1, pp. 247-274.

Criscuolo, C., P.N. Gal and C. Menon (2014), "The dynamics of employment growth: New evidence from 18 countries", *OECD Science, Technology and Industry Policy Papers*, No. 14, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jz417hj6hg6-en.

Decker, R. et al. (2016). "Declining business dynamism: What we know and the way forward", *American Economic Review*, Vol. 106/5, pp. 203-07, American Economic Association, http://dx.doi.org/10.1257/aer.p20161050.

Decressin, J. and A. Fatas (1995), "Regional labor market dynamics in Europe", *European Economic Review*, Vol. 39/9, pp. 1627-1655, Elsevier B.V., https://doi.org/10.1016/0014-2921(94)00102-2.

European Commission (2016), "European Structural and Investment Funds 2014-2020: 2016 summary report of the programme annual implementation reports covering implementation in 2014-2015", European Union, http://ec.europa.eu/regional_policy/en/policy/how/stages-step-by-step/strategic-report (accessed in February 2017).

OECD (2017), "Key issues for digital transformation in the G20: Report prepared for a joint G20 German Presidency/OECD conference", OECD, Paris, https://www.oecd.org/g20/key-issues-for-digital-transformation-in-the-g20.pdf.

OECD (2016a), *OECD Regional Outlook 2016: Productive Regions for Inclusive Societies*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264260245-en.

OECD (2016b), "A business demography database for OECD regions: Data, methodology and applications", paper presented at the 31st Working Party on Territorial Indicators, Paris, 7 November 2016.

OECD (2015), *OECD Science, Technology and Industry Scoreboard 2015: Innovation for Growth and Society*, OECD Publishing, Paris, http://dx.doi.org/10.1787/sti_scoreboard-2015-en.

OECD (2005), *OECD SME and Entrepreneurship Outlook 2005*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264009257-en.

Puga, D. (2010), "The magnitude and causes of agglomeration economies", *Journal of Regional Science*, Vol. 50/1, pp. 203-219, http://dx.doi.org/10.1111/j.1467-9787.2009.00657.x.

Rosenthal, S.S. and W.C. Strange (2004), "Evidence on the nature and sources of agglomeration economies", in *Handbook of Regional and Urban Economics*, Vol. 4, pp. 2119-2171.

Annex 5.A1. Measurement challenges and implemented solutions

Measurement challenges and implemented solution

Measurement of entry and exit

Cross-country analyses of firm dynamics require a harmonised definition and comparable measures of firm entry and exit. This can be particularly challenging. This subsection discusses the approach taken in the DynEmp project, with a particular focus on the specific challenges associated with the use of plants as a unit of analysis.

In the *dynemp_reg* programme, entrants are tagged by taking advantage of the birth year reported in the source database. Few corrections to this birth year are automatically implemented,[1] unless an option in the programme that forces the use of the exact source birth year is specified. If the data are left-censored and the user specifies this in the programme, the calculation of the age variable will take this into account, keeping the unit's age class missing until the plant becomes old enough to be tagged in a certain age class with certainty.

Two important challenges arising when defining entry are worth further discussion. The first one concerns whether a new plant is related to the entry of a new firm or is instead part of an existing multi-plant firm. In this respect, a procedure integrated in the *dynemp_reg* routine allows distinguishing these two cases, whenever the optional firm identifier variable is specified. In particular, when this variable is specified, the output is further disaggregated, providing separate statistics for single plants or plants part of a multi-plant firm.[2] The second challenge is related to whether the registration of a new plant (and therefore its entry) corresponds to economic (*de novo*) entry or is rather related to other events such as a merger or an acquisition (*de alio* entry). The *dynemp_reg* programme optionally allows the separate account of *de novo* and *de alio* entry, whenever the relevant information is available in the source data.

Consistently with the previous phases of the DynEmp project, the exit event is defined internally by the *dynemp_reg* routine based on the last year in which the unit appears. In particular, the exit variable is equal to one in the year following the last time a unit appears in the data with positive employment.

Similarly to the case of entry, the *dynemp_reg* programme optionally allows the separate account of units that exit due to a change in legal status whenever the relevant information is available in the source data.

Measurement of location

The analysis of employment and business dynamics at the local level requires the presence in the source data of a location variable. In order to choose the most appropriate location variable to use in the *dynemp_reg* code, we took advantage of the preliminary questionnaire on data availability, described above.

The majority of countries for which data allow analysis at the plant level (7 out of 12) responded that information on the postal code of plants is available in their microeconomic database. Few countries have only information on city or municipality codes (such as Brazil

and Japan), while other countries have only more aggregated geographic information (such as Denmark and Spain).

The *dynemp_reg* code therefore integrated a procedure to identify regional and geographical levels of aggregation (TL2 regions and metropolitan areas) starting from the postal code of units, based on an external correspondence table for each country. These correspondence tables have been developed by the OECD Centre for Entrepreneurship, SMEs, Local Development and Tourism.[3]

A few challenges need to be mentioned when developing a cross-country analysis that needs to deal with postal codes to identify plant location over time. These include the heterogeneity in postal code formats across countries; the possible change over time of the postal code for a given plant; possible reforms in the postal code systems.

The *dynemp_reg* routine implements an automatic cleaning of the postal code variable, aiming at assigning a time invariant zip code to each plant. In particular, it creates a zip variable which takes the value of the most frequent (modal) zip code whenever such zip code changes over time or is missing, for a given plant. In case the mode is not available, the variable takes the zip code reported in the plant's latest available year. For reference, the automatically generated log file also reports the percentage of zip codes modified by the automatic cleaning procedure and the share of zip codes matched with the external correspondence table.

Measurement of employment

Measurement of employment across countries is a challenging task. In the first version of the *dynemp_reg* code, the following recommendations are made to project participants.

Employment records should preferably be based on both headcounts and full-time equivalent if available, in two separate runs. The employment variable should measure employees, if available. If the available employment variable measures instead total employment, the programme should be run anyway but the DynEmp team should be informed about this. Employment can refer to either a yearly average or to a precise point in time (given that the focus is on longitudinal growth, this should not make a significant difference). The programme will run regardless of whether the employment variable is expressed as an integer or a decimal number. It is assumed that no additional rounding beyond that to unity is applied on the variable in the data.

Further refinements are planned for the subsequent versions of the *dynemp_reg* programme, as discussed in the final section of this annex.

Changes in industrial classification

Another measurement challenge which applies when working with panel databases on business and employment dynamics concerns the changes in industrial classifications. In particular, a major change occurred between 2008 and 2009 in a significant number of countries due to the adoption of the new ISIC Rev. 4 (or NACE Rev. 2 in Europe) classification. In this context, many former industries were split into several parts, and others merged into a single industry.

For instance, the activities classified under the industry "Printing and publishing" (code 22) in ISIC 3.1 (NACE Rev. 1.1 in Europe) were split into five different 2-digit industries in ISIC Rev. 4 (NACE Rev. 2). Some of these industries are in manufacturing,

while some others are in services. As highlighted by the previous example, changes were not of the one-to-one type but n-to-m types. This applies to all levels of industry classification (i.e. 2-, 3- and 4-digit).

An additional challenge in this framework concerns the fact that units also change their activity from time to time, irrespective of changes in the classification system. However, it is typically more convenient to work with a constant industry identifier over time as this simplifies many types of analyses that make use of the industry dimension. For instance, a constant industry classification simplifies the definition of entry and exit as there is no need to follow which activity the unit enters or exits.

The *dynemp_reg* programme, building upon the experience developed within the DynEmp and MultiProd projects, implements a probabilistic industry conversion system, which is described in detail by Criscuolo, Gal and Menon (2015) and in Berlingieri, Blanchenay and Criscuolo (2017). Its main rationale is to convert the input data according to the ISIC Rev. 4 classification before micro-aggregating them, taking advantage of probabilistic weights calculated in the years of overlapping industry classifications. Whenever overlapping years are not available, the routine takes advantage of an external correspondence table. Finally, in order to have a time-invariant industry code for each plant, the programme assigns to each unit its modal industry or its most recent industry code if the mode is not available.

Definitions

The *dynemp_reg* routine computes a few intermediate unit-level variables, which are subsequently used to calculate summary statistics at different aggregation levels in the final micro-aggregated dataset. The programme runs independently of whether the employment input data are expressed as an integer or a decimal number (it would round them up in the latter case).

The formula used to calculate the employment growth rate is presented in the following equation:

$$\gamma_{i,t} = \frac{L_{i,t} - L_{i,t-1}}{\frac{1}{2}(L_{i,t} + L_{i,t-1})}$$

Note that $L_{i,t}$ stands for employment of unit i in year t. The formula is commonly used in the business dynamics literature as it has the advantage of not being biased by mean reversion dynamics (see Davis and Haltiwanger [1999], among others). The index is also scale neutral (i.e. it does not depend on the employment level at the beginning of the period) and is bounded between -2 and +2.

The unit's year of birth is the first year of activity of the unit and, as previously discussed, is needed to calculate the unit's age. If the data are left-censored and the user specifies this in the programme, the calculation of the age variable will take this into account.

In all output databases three different groups of plants can be identified: entering units, exiting units and incumbents.

In the *dynemp_reg* programme, incumbents, entering units (entrants) and exiting units (exitors) are defined as follows. For each time interval (t-1, t), an entrant is a unit that is

not there in *t*-1 but is there in *t*; an exitor is a plant that is not there in *t* and is there in *t*-1. An incumbent is a unit that is there in *t*-1 and *t*.

Additional tables

Table 5.A1.1. **Average employment growth of young units and regional characteristics**

A. Manufacturing sector

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intermediate	-0.246		-0.170	-0.181	-0.149			-0.148	-0.134
	(0.266)		(0.280)	(0.307)	(0.294)			(0.322)	(0.323)
Mostly urban	-0.591*		-0.483*	-0.500	-0.445			-0.460	-0.442
	(0.313)		(0.266)	(0.387)	(0.360)			(0.334)	(0.347)
Frontier		-0.553	-0.161			-0.364	-0.259	-0.132	-0.104
		(0.447)	(0.448)			(0.565)	(0.570)	(0.563)	(0.588)
Keeping pace		0.183	0.210			0.150	0.129	0.208	0.204
		(0.408)	(0.452)			(0.428)	(0.432)	(0.458)	(0.466)
Log plant density				-0.0622		-0.106		-0.0234	
				(0.155)		(0.171)		(0.207)	
Log population density					-0.0812		-0.130		-0.0357
					(0.139)		(0.160)		(0.202)
Observations	356	356	356	356	356	356	356	356	356
R-squared	0.820	0.820	0.821	0.821	0.821	0.820	0.820	0.821	0.821

B. Non-financial market services sector

	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Intermediate	0.0634		0.0605	-0.161	-0.0878			-0.122	-0.0288
	(0.143)		(0.147)	(0.170)	(0.165)			(0.179)	(0.171)
Mostly urban	0.347**		0.196	0.0190	0.119			0.00906	0.0940
	(0.152)		(0.151)	(0.197)	(0.192)			(0.188)	(0.188)
Frontier		0.618***	0.455*			0.221	0.387	0.122	0.312
		(0.217)	(0.239)			(0.307)	(0.294)	(0.320)	(0.304)
Keeping pace		0.151	0.136			0.218	0.193	0.163	0.153
		(0.199)	(0.212)			(0.208)	(0.206)	(0.217)	(0.217)
Log plant density				0.182***		0.165**		0.180*	
				(0.0694)		(0.0781)		(0.0980)	
Log population density					0.126*		0.103		0.0898
					(0.0763)		(0.0828)		(0.102)
Observations	356	356	356	356	356	356	356	356	356
R-squared	0.943	0.943	0.943	0.944	0.943	0.944	0.943	0.944	0.943

Notes: The reference categories are mostly rural regions and diverging regions. Data on regional characteristics are not available for Costa Rica, which is excluded from the estimation sample. All regressions include year and country dummies. Robust standard errors in parentheses. The tables do not report the regression constants. *** p<0.01, ** p<0.05, * p<0.1.

Table 5.A1.2. **Relative net job creation by entering units and regional characteristics**

A. Manufacturing sector

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intermediate	0.147		0.150	0.119	0.0815			0.197	0.192
	(0.0948)		(0.105)	(0.117)	(0.125)			(0.120)	(0.127)
Mostly urban	0.177**		-0.0205	0.138	0.0781			0.0290	0.0273
	(0.0864)		(0.112)	(0.109)	(0.126)			(0.122)	(0.129)
Frontier		0.527***	0.614***			0.566***	0.555***	0.677***	0.681***
		(0.146)	(0.191)			(0.152)	(0.163)	(0.192)	(0.197)
Keeping pace		0.129	0.206**			0.122	0.124	0.201**	0.197**
		(0.0823)	(0.0985)			(0.0831)	(0.0834)	(0.0977)	(0.0979)
Log plant density				0.0265		-0.0222		-0.0510	
				(0.0473)		(0.0340)		(0.0406)	
Log population density					0.0549		-0.0124		-0.0420
					(0.0488)		(0.0350)		(0.0455)
Observations	356	356	356	356	356	356	356	356	356
R-squared	0.651	0.663	0.666	0.651	0.652	0.663	0.663	0.667	0.666

B. Non-financial market services sector

	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Intermediate	0.450***		0.418***	0.195*	0.198*			0.227*	0.239**
	(0.0982)		(0.105)	(0.114)	(0.114)			(0.122)	(0.121)
Mostly urban	0.479***		0.286***	0.106	0.0986			0.0903	0.0809
	(0.0882)		(0.0992)	(0.102)	(0.104)			(0.104)	(0.107)
Frontier		0.609***	0.519***			0.0561	0.101	0.170	0.231
		(0.153)	(0.188)			(0.176)	(0.175)	(0.207)	(0.203)
Keeping pace		-0.0790	0.0566			0.0141	0.0141	0.0853	0.0915
		(0.115)	(0.125)			(0.116)	(0.116)	(0.126)	(0.127)
Log plant density				0.207***		0.230***		0.188***	
				(0.0385)		(0.0396)		(0.0477)	
Log population density					0.210***		0.225***		0.180***
					(0.0423)		(0.0412)		(0.0495)
Observations	356	356	356	356	356	356	356	356	356
R-squared	0.850	0.848	0.855	0.861	0.859	0.859	0.858	0.861	0.860

Notes: The reference categories are mostly rural regions and diverging regions. Data on regional characteristics are not available for Costa Rica, which is excluded from the estimation sample. All regressions include year and country dummies. Robust standard errors in parentheses. The tables do not report the regression constants. *** p<0.01, ** p<0.05, * p<0.1.

Table 5.A1.3. **Post-entry growth and survival of entrants and regional characteristics**

A. Post-entry employment growth

	Post-entry employment growth
Year=2010 dummy	0.0485***
	(0.0157)
J=5 dummy	0.0431***
	(0.0141)
Non-financial market services	0.0349***
	(0.0131)
Mostly intermediate dummy	0.0438***
	(0.0159)
Mostly urban dummy	0.0618***
	(0.0151)
Observations	132
R-squared	0.207

B. Survival share (in terms of employment)

	Survival share (employment)
Year=2010 dummy	0.000490
	(0.0139)
J=5 dummy	-0.123***
	(0.0147)
Non-financial market services	-0.0725***
	(0.0119)
Mostly intermediate dummy	-0.0189
	(0.0154)
Mostly urban dummy	-0.0140
	(0.0152)
Observations	132
R-squared	0.516

Notes: Dependent variables are: post-entry employment growth in Panel A; survival share of entrants (in terms of employment) in Panel B. The reference categories are year=2007 for the years dummies; manufacturing for the macro sector dummies; mostly urban regions for the regional characteristics dummies. Robust standard errors in parentheses. The tables do not report the regression constants. *** p<0.01, ** p<0.05, * p<0.1.

Notes

1. These corrections include replacing the birth year with the first year with positive employment when the birth year is within the observed period, or when the birth year is later than the first year of appearance with positive employment.
2. In the first version of *dynemp_reg* this additional optional disaggregation level is available only for the metropolitan part of the output, in order to reduce possible issues related to residual confidentiality.
3. Further methodological details for metropolitan areas are available in OECD (2012) and for TL2 regions at: http://stats.oecd.org/wbos/fileview2.aspx?IDFile=cebce94d-9474-4ffc-b72a-d731fbdb75b9.

References

Berlingieri, G., P. Blanchenay and C. Criscuolo (2017), "The great divergence(s)", *OECD Science, Technology and Industry Policy Papers*, No. 39, OECD Publishing, Paris, http://dx.doi.org/10.1787/953f3853-en.

Criscuolo, C., P.N. Gal and C. Menon (2015), "dynemp: A routine for distributed microdata analysis of business dynamics", *Stata Journal*, Vol. 15/1, pp. 247-274.

Davis, S.J. and J. Haltiwanger (1999), "Gross job flows", Chapter 41 in Ashenfelter, O. and D. Card (eds.), *Handbook of Labor Economics*, Volume 3, pp. 2711-2805, Elsevier Science BV, North Holland.

OECD (2012). *Redefining "Urban": A New Way to Measure Metropolitan Areas*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264174108-en.

Annex 5.A2.
Contributors to the DynEmp regional data collection

Table 5.A2.1 summarises the contributors to the DynEmp regional data collection. The table includes only countries for which data had been received at the time of writing. A more detailed version of the table, including all contributors to the DynEmp and MultiProd project, is available in OECD (2017).

Table 5.A2.1. **Contributors table**

Country	Contributor(s)	Institution(s)
Costa Rica	Alfonso Alfaro Ureña, David Bullon Patton, Arlina Gómez, Tayutic Mena, Francisco Monge	Banco Central de Costa Rica (BCCR) and COMEX
Finland	Mika Maliranta	Research Institute of the Finnish Economy (ETLA)
France	DynEmp team	OECD
Sweden	Fredrik Andersson	Statistics Sweden (SCB)

Reference

OECD (2017), *Business Dynamics and Productivity*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264269231-en.