Job creation, firm creation, and de novo entry*

Karen Geurts and Johannes Van Biesebroeck
KU Leuven

[Extremely preliminary and incomplete]

May 9, 2014

Abstract

Firm dynamics and recorded firm changes in large-scale data sets are different realities. By separating the two and concentrating on *de novo* entrants and true failures in an economic sense, we find highly regular post-entry employment patterns of young firms. Exit rates decrease with age within every size class and decrease with size in each of the five years after entry. At the same time, young firms that survive have high employment growth rates, which also monotonically decrease with age in every updated size class. More remarkably, we find that Gibrat's Law is violated for very young firms, yet as a positive relation between current size and growth: conditional on age, large firms grow more rapidly than small firms. These patterns are highly obscured, or even reversed, when all creations of new firm identifiers are interpreted as entrants, and all closures of existing identifiers as exits. An advanced method of firm record linking allows us to accurately identify employer firms that enter *de novo*, and to consistently estimate firm-level survival and employment growth after entry, even when firms change identifier, merge or split-up.

* The authors are grateful to the Belgian National Social Security Office (NSSO) for providing the data and in particular to Peter Vets for developing the employee flow record linking method in accordance with our research needs. We are grateful to Youri Baeyens and Antonio Fiordaliso of Statistics Belgium for linking the NSSO data to the Official Belgian Business Register and providing the official firm record linkages created for the Eurostat Structural Business Statistics.

I. Introduction

Firm-level employment dynamics cannot be viewed separately from firm dynamics. Firms enter and eventually exit from the industry, but in between they can be restructured or change legal identifier for a variety of reasons. Simply looking at the group of entrants and exits in an administrative sense, i.e. creations and closures of firm identifiers, could give a highly misleading impression of the growth patterns of young firms and of their contribution to job creation.

We supplement the official procedures for firm record linking with employee-flow information to identify a well-defined group of *de novo* entrants. *De novo* entrants correspond to the creation of a new employer firm starting from scratch.[1] We separate them out from *de alio* entrants, which are new firm identifiers created after an administrative, legal or organizational change of an established firm. The type of *de novo* entrant closely corresponds to the creation of a new firm in models of firm dynamics discussed below. Our firm linking procedure similarly allow us to distinguish between exit, the firm's decision to stop its activity, and a *transfer* of activities to a new firm identifier.

We find that *de novo* entrants all enter at a small scale and show patterns of survival and employment growth that are remarkably regular, even over the volatile recession period of 2008. Young firms exhibit high exit rates but the ones that survive have high growth rates, and both measures strongly decrease with age in the very first years after entry. These patterns hold both overall and within distinct size classes of firms, measured as current size at a given age. Five years after entry, exit and growth rates have come close to those of older firms of the same size.

Within cohorts of the same age, larger firms have higher survival rates than smaller ones and, more noticeably, also have higher growth rates, both in the set of surviving and all firms. Thus Gibrat's Law is violated for very young firms, yet as a *positive* relation between current size and growth. The increasing growth rates by size are especially strong in the very first years after entry, and move towards a more proportional distribution as an entry cohort matures. After five years, size differences in growth are similar to those of older firms.

The contribution of entrants to job creation in the economy depends on their absolute importance, the initial firm size distribution and changes in the distribution over time, and the illustrated exit and growth patterns. We highlight two findings. First, the initial contribution of entrants to job creation is very small. It is, however, strongly overestimated when *de alio* entrants are not adequately filtered out: initial employment of this subset should not be considered as job creation, because it is merely the result of a *transfer* of existing jobs to a new firm identifier. Second, firm size at startup has some predictive power for subsequent job creation: employers that enter with more than 5 employees better succeed to maintain the initial employment level than smaller firms. More specifically, an entrant cohort's initial employment declines in the first five years after entry, but it decreases more rapidly in the subset of firms that enter with less than 5 employees than in the set of larger entrants.

Our main findings are consistent with a model of passive learning where firms are assumed to enter without knowing their relative cost efficiency (Jovanovic, 1982). The distribution of true costs among the potential firms is yet known to all. With the passage of time, the firm learns about its efficiency level through the observation of its costs. Based on this prior information, the firm re-estimates its cost efficiency and adjusts its profit-maximizing output level for the next period. Firms that discover they are more efficient, survive and

[1] which will be identified by a set of scattered employees that are brought together at the start-up of the firm

grow, while the inefficient decline and eventually fail. As a result, the average efficiency of the survivors improves from period to period, inducing less failure as the cohort ages. Also growth rates of survivors decrease with age, because there is less need for further size adjustments when estimates about cost efficiency become more precise as more evidence comes in. Yet because the most efficient firms grow, sizes of firms of the same cohort diverge. We find that, among employer firms, this selection process is concentrated in the very first years after entry and then rapidly fades out: exit and employment growth rates are especially high in the first year of existence and strongly decline with age at a decreasing rate. Between the fourth and fifth year after entry, the decrease in growth and exit rates is only minor.

The inverse relationship between exit and size within a given age cohort follows from this process of 'noisy' selection. Larger firms are the ones that have received favorable cost information in previous periods, and subsequent information is less likely to be unfavorable enough to induce exit. Firms that have remained small, by contrast, are exactly those with higher expected costs and so have lower value of staying in the industry.

The positive relation between growth and size in the early years after entry is consistent with the passive learning model which assumes that all firms enter at a similar small size, and sizes diverge as some firms find that they are more efficient and grow, while the less efficient ones stay small. If adjustment costs prevent young firms to fully adjust their size instantaneously in response to an observed cost efficiency level then, for expanding firms, current size will understate the desired size. In other words, if expansion can only be made gradually, larger firms, which are the ones that have received favorable cost information in the previous periods, will also be likely to be growing in the future, and a positive correlation between size and growth will be observed. When revisions of estimated cost efficiency become smaller with the passage of time and firms approach their optimal scale, differences between growth rates across size classes decrease and firms move towards a more proportional growth distribution. This is exactly the pattern we find among de novo entrants.

Important empirical studies have found regularities in how exit and growth of firms depends on size and age, that are consistent with a model of passive learning. While most studies investigate these patterns in a population of firms of all ages, our results complement this line of research by a narrow focus on the very first years after entry. Most post-entry patterns we find confirm the regularities found also among firms of different , but the positive size-growth relationship is a specific feature of young firms in the very first years after entry.

Declining exit rates with age and current size have been found in many studies (Caves, 1998). Dunne, Roberts and Samuelson (1989a) (hereafter DRS) and, more recently, Haltiwanger, Jarmin and Miranda (2013) have shown that exit rates are also inversely related to age when size is held constant, and that exit rates decline with size even within distinct age classes. The same authors further find that growth rates of surviving firms decline with age given size. Concentrating particularly on the post-entry period, Mata, Portugal and Guimaraes (1995) similarly concludes that that current size has a negative effect on the failure rate.

Empirical evidence about the existence of an inverse relationship between growth and size, that contrasted with the commonly accepted view that firms grow proportional to their size (Sutton, 1997), motivated Jovanovic to develop his theory of noisy selection. Within this framework, the higher growth rates smaller firms exhibit follow from the fact that they are, on average younger. Conditional on age, however, predicted growth rates of survivors are the result of a Martingale process, thus the inverse relation between growth and size relationship does not hold within specific age groups. Evans (1987) and DRS (1989) yet find that growth rates of surviving firms are also inversely related to size given age. Haltiwanger et al. (2013), by contrast, find no (or a mild positive) relation between size and growth rates of surviving firms once controlling for age. While these studies examine the size-growth relationship in the population of firms of all ages, our analysis concentrates on the early years after entry. As our results suggest, the size-growth relationship in that period is entirely different from the one at later age: larger young firms grow faster than smaller ones, but these size differences are limited to the very first years after entry and move towards a more proportional distribution about age five.

Several studies, investigating plant, not firm creation, have pointed out that post-entry patterns differ between de novo entrants and plants created by established firms, or plants created by merger or acquisition (Mata et al. 1995; Baldwin and Gorecki, 1987; DRS, 1989). Eriksson and Kuhn (2006) and Muendler, Rauch and Tocoian (2012) and show specific post-entry dynamics for spin-offs of existing firms. Our distinction between de novo and de alio entrants goes beyond that, first because we focus on entirely new employer firms, that have, by definition, no other plants on the Belgian territory, and second, because our set of de novo entrants excludes firms that emerge from an established firm for whatever the reason may be. It may be an organizational change, such as a merger or split-up of activities, but we also exclude entrants that are transformations of established firms for legal, liability, tax or administrative reasons. This type of transformations represent the most important share of what we have labeled as de alio entrants. Moreover, our firm linking procedure allows us to consistently follow firm-level employment growth of de novo entrants that are themselves involved in some kind of transformation after entry, which we have labelled as a *transfer* of activities.

Our results show that the specific post-entry patterns young firms exhibit are highly obscured, or even reversed in an analysis of firm-level data where de novo and de alio entrants are mixed up and transfers after entry are disregarded. Because such events predominantly occur in larger size classes, they strongly bias the estimated relationship between firm age and size and growth. First, de alio entrants, which are transformations of an existing firm, lack the typical post-entry dynamics of young firms and show similar features as incumbents. They introduce a downward bias in overall exit and growth rates of young firms and hide the positive size-growth relationship firms exhibit at early ages. Second, young firms that transfer activates to a new identifier after entry, introduce spurious exits and firm-level employment shocks. This leads to irregularities in estimated post-entry patterns.

The next section describes the data set we employ. In section III, we describe the firm record linking method and how it affects the identification of firm-level entry and exit. Section IV discusses exit and growth patterns, and how they are related to firm age and size. In section V, we show the impact of our findings on overall job creation by entrants. Section VI concludes.

II. Data

The analysis in this paper is based on a firm-level data set maintained by the Belgian National Social Security Office (NSSO). It covers the full population of firms with at least one employee in Belgium. Firms are identified by means of their official Belgian enterprise number, the CBE number (Crossroad Bank for Enterprises). This unique number, that new enterprises receive upon registration, is used by all government administrations. Firms keep the CBE number for the rest of their lifetime, also when the legal status or ownership changes. In contrast to some other countries, the Belgian enterprise number does not suffer from this kind of administratively induced changes. As a result, administrative entrants in our data set are readily identified as new CBE numbers entering the NSSO data base, i.e. the first period of observation they record a positive number of jobs.

The NSSO data set is based on quarterly declarations of employers filling in social security information about their employees. Employees are identified as well by a unique identification number.[2] We will exploit the linked employer-employee information in the NSSO data to trace the continuity of firms over changes in identification number and organizational structure.

We also make use of continuity links provided by Statistics Belgium. The links are based on information in the Crossroad Bank for Enterprises. This comprehensive data base merges firm-level data from different administrations such as the national register of legal entities, the trade register, VAT declarations, and NSSO. The information is merged at the firm level by means of the CBE number.

We restrict our empirical analysis to a subset of the NSSO data base. For comparison with other research, only firms in the private non-farm sector are included. We also exclude highly subsidized industries where firm and job creation received strong impetus from government programs in the period of observation. Excluded are Human health and social work activities, an industry in which more than 70% of expenditures are publicly financed; and industries that make use of service vouchers, a government scheme for household help established in 2004, which subsidizes 70% of the wage cost.[3] Since price competition in these activities is weak or absent, firm performance is not representative of that in competitive markets we want to study. Table A.1 in annex lists the industries that are included in the analysis, as well as their classification into six main industries that we use throughout this paper.

The analysis covers the period 2003-2012. The data include 178,000 active firms and 2,070,000 employees on average per year. Aggregate employment in this set of firms increased by 0.9 per cent per year till 2008, dropped by 2.5% between 2008 and 2010 and is more or less stable since then.

III. Employee flow method: identifying de novo entrants and exit

Firm linkage problem

Between the point a firm starts up activities and exits the market, its administrative identifier may change for three reasons. In several countries, firms are administratively assigned a new identification number when the ownership or legal status changes. Firms may also induce the ID change themselves for tax or liability reasons, closing down the existing business number and continuing the same activities in a newly registered company. Finally, business numbers may be merged or split up in case of mergers, acquisitions and split-offs.

Such events impede a straightforward analysis of firm entry and exit based on administrative data sources in several ways. Misinterpreting changes in administrative identifiers as closings and start-ups of firms yields an overestimation of measures of entry and exit, and of job creation and destruction associated with it.[4] It also introduces a bias in the size-growth relationship of firms, since the likeliness of being involved in an event

---

[2] The individual identification number in the National Register.

[3] Employment in Health and social work increased with 39% between 2003 and 2012 (from 9% to 11% of total salaried employment); in the same period, the number of employees in the service voucher system increased from 0% to 4% of total salaried employment.

[4] see Spletzer (2000); Baldwin, Beckstead and Girard (2002); Benedetto, Haltiwanger, Lane and McKinney (2007); Abowd and Vilhuber, 2005.

often depends on size. Finally, it hampers comparative analysis, because the incidence of changes in firm identifiers depends on country-specific administrative, legal and tax regulations.[5]

Record linking methods

To overcome these problems, official statistical institutes have developed methods to establish links between the records of individual firms when identifiers are changed, merged or split-up. We make use of the record linking method developed by Statistics Belgium, which follows the OECD/Eurostat guidelines for constructing comparable indicators on firm dynamics.[6] The method first exploits information from additional administrative sources available in the Crossroad Bank for Enterprises, such as data on corporate restructurings, registrations of legal entities, and social security declarations. Second, a probabilistic matching procedure is applied, which makes use of similarities in name, address, and industry code. The links between firm identifiers are used to reclassify some entrants as incumbents and some exits as continuing firms. In this paper, we will denote them as 'identified by the official method'.

A more direct way, however, to track the longitudinal history of a firm over changes in identifier is to follow one of its main production factors, the work force[7]. If a firm changes ID number but continues its activities, the work force of the previous and the new number will be largely the same. Continuity of the work force can thus be used to trace the continuity of the firm. Employee-flow methods for record linking implement this continuity feature by looking at employees registered at different firm identifiers in two consecutive periods of observation. If a large cluster of employees 'moves' between two identifiers in a short period of time, it is unlikely to be the result of individual inter-firm worker mobility but rather an indication of a change, merger or split-up of firm identifiers. Firm identifiers that are linked by such large clustered employee flows are considered as referring to the same, or parts of the same firm.

In addition to the official linkages provided by Statistics Belgium, we apply an employee flow method to identify links between identifiers of the same firm.[8] This method has been developed in collaboration with NSSO.[9] Central in the design of an employee flow method is the minimum cluster size that is imposed to distinguish employee flows that signal links between firm identifiers from the ones that are merely the result of job changes of individual employees. Recent employee flow methods apply a combination of absolute and relative thresholds. The minimum size of the cluster is mostly set at three or five employees, since below, there is a high probability that the employee flow merely represents individual worker mobility. The minimum relative size of the cluster is mostly between 50% and 80% of work force of the firms involved. The exact level of the relative threshold is not critical since the distribution of the relative cluster sizes is strongly U-shaped: most clusters represent either a very small share of the firm's work force, and represent individual worker mobility, or a very large share, and in that case signal changes in firm identifiers.[10]

A general drawback of employee flow methods is that links between firms that are smaller than the minimum cluster size of three or five employees cannot be captured be definition. Our solution to this problem is given by the official record linking method which does capture links between small firms.

---

[5] Bartelsman, Scarpetta and Schivardi (2005); Bartelsman, Haltiwanger, Scarpetta (2009); Vilhuber 2009
[6] Eurostat (2003); Eurostat/OECD (2007)
[7] If one of the main factors of production, the work force, is (partly) identical in two administrative records at two consecutive points in time, there is a high probability that these records relate to (parts of) the same firm.
[8] Geurts and Vets (2013).
[9] The DynaM method has been developed with the aim of producing reliable indicators on firm and employment dynamics in Belgium. More information see www.dynam-belgium.org
[10] Benedetto et al. (2007); Hethey and Schmieder (2013); Rollin (2013).

In the employee flow method used in this paper, the key decision rules to establish a link between different identifiers of the same firm are as follows. First, the clustered employee flow has to be observed between two successive quarterly observations *q-1* and *q*. This short time span is a first indication that sizable clusters are unlikely to be the result of individual worker mobility. Second, only clusters of at least five employees are taken into account. The third rule sets the relative threshold of the cluster: a link is established between two business numbers if the employee cluster represents at least 50 per cent of the work force of both the origin firm in *q-1* (the 'predecessor'), and the destination firm in *q* (the 'successor'). This rule is a direct implementation of the assumption that two successive firm identifiers of which the production factor labor is mostly the same, are identifiers of the same firm. 78 per cent of the employee flow links we identify at entry meet this key rule.

A set of additional rules are developed to identify also links between firm identifiers in case of mergers and split-ups. A common example is a predecessor that transfers only part of its activities to a newly created business number. If this part is smaller than half of the predecessor's work force, the key 50 per cent rule will not capture the event. The work force of the new business number, however, will predominantly consist of employees that are transferred from the predecessor. To capture this kind of split-ups, we impose that the clustered employee flow represents at least 75 per cent of the work force of the new number. Another 18 per cent of the employee flow links we identify at entry meet this rule. Similar decision rules are worked out to capture other forms of organizational restructurings. The full set of rules is presented in the Appendix A.

De novo versus de alio entrants

The links established by both the official and the employee flow method for record linking are first used to identify administrative entrants that do not start *de novo*, but which result from a mere administrative relocation of existing production factors of an incumbent to a new business number. These firms are denoted as *de alio* entrants, i.e. emerging from an established firm. The distinction between *de novo* and *de alio* entrants is based on *quarterly* information, i.e. the presence or absence of a link in the quarter a new business number enters the NSSO data set. Yet all summary statistics and post-entry patterns presented in this paper will be based on employment measured at a fixed *annual* point of observation, more specifically the end of the second quarter of a given year.

Table 1 shows that, in the year of entry, d*e alio* entrants count for 9 per cent of all administrative entrants. The likeliness that a new business number corresponds to the *de novo* entry of a firm decreases dramatically with size: more than half of administrative entrants with more than 10 employees are identified as *de alio*, and above 50 employees, *de novo* entrants are rather unexceptional. In other words, if a new business number is registered with a large number of employees, there is a high probability that it is simply set up to take over (part of) the production factors of an established enterprise.

Table 1 Share of de alio entrants in all administrative entrants (weighted average 2004-11)

		Size class				
	Total	1-4	5-9	10-19	20-49	50+
Total share of de alio entrants	0.09	0.05	0.36	0.59	0.74	0.89
1. Identified by official method	0.06	0.05	0.12	0.16	0.21	0.37
2. Identified by employee flow method	0.05		0.32	0.57	0.72	0.88
by type of employee flow link						
2.a Predecessor and entrant share at least 50% of employees (ID-change)	*0.04*		*0.25*	*0.44*	*0.54*	*0.65*
2.b At least 75% of entrant's employees are transferred from predecessor (Split-up)	*0.01*		*0.05*	*0.10*	*0.16*	*0.22*
2.c Other	*0.00*		*0.01*	*0.03*	*0.03*	*0.02*

Table 1 further shows that the two record linking methods are complementary for the identification of *de alio* entrants. The official method is especially needed in the size class below five employees, where employee flow links are by definition absent. Yet the employee flow method identifies two to three times more *de alio* entrants in larger size classes. The combined linkage results provide us with the maximum accuracy in distinguishing between *de alio* and *de novo* entrants that can be reached with available methods. The employee flow links show that the vast majority of *de alio* entrants emerge from simple ID changes (78%), where the clustered employee flow represents at least 50 per cent of the work force of both the predecessor and the entrant. Another important share arise from split-offs of an incumbent (18%). By definition, *de alio* entrants identified by an employee flow link, have at least 50 per cent of their work force in common with a predecessor. In observation, however, this share is close to 100 per cent for most de alio entrants, with a median of 94 per cent and a mean of 90 per cent.[11]

In what follows we show that *de alio* entrants, although a small group, introduce a strong bias in entry statistics based on administrative data. They exhibit incumbent-like features, which largely obscure the distinct characteristics of de novo entrants if they are not adequately filtered out. The official OECD/Eurostat record linking method helps to avoid these biases. This method fails however to separate out many large *de alio* entrants. The added value of the employee flow method is that it helps to more accurately identify *de novo* entrants, and to more distinctly reveal their specific size distribution and contribution to job creation.

Graph 1 shows how a more accurate identification of *de novo* entrants affects the employment distribution of entrants. Table 2 presents some additional statistics. The results are highly sensitive to different definitions of entrants. The administrative and official method locate a significant share of entrants' employment in larger firms, and assign an average entry size of 3.1 and 2.7 jobs respectively. *De novo* entrants, by contrast, are much smaller (1.9 jobs) and exhibit a strongly right-skewed employment distribution. Entrants defined by the employee flow method show quite similar features. The size characteristics of *de alio* entrants are remarkably different and clearly reflect those of incumbents, the population they emerge from: they are much larger (14.6 jobs) and the mass of employment is located in larger firms. These results are independent of the industry-specific size structure (Table A.3): in both manufacturing and private services, the employment distribution of *de alio's* reflects that of incumbents, while the employment distribution of *de novo's* mirrors it.

[11] The remaining shares mainly represent natural in- and outflow of employees.

Graph 1. Employment distribution of *de novo* and *de alio* entrants (weighted average 2004-2011)

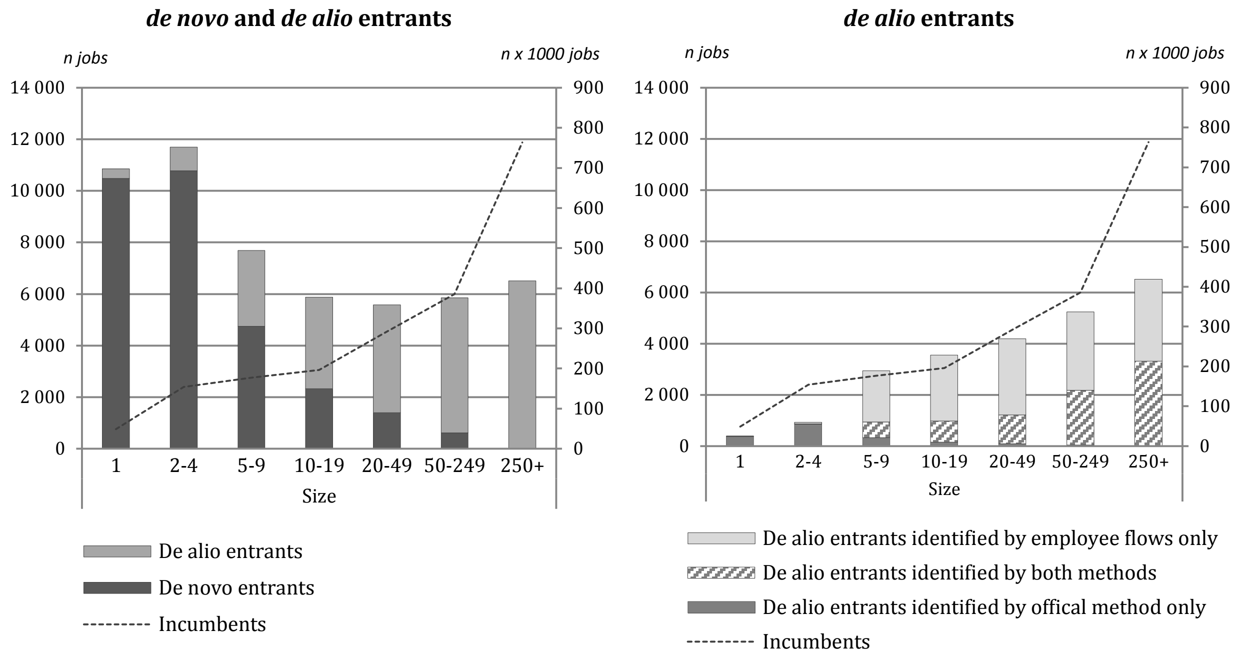


Table 2. Summary statistics different types of entrants (weighted average 2004-2011)

	Share of entrants		Entry rate (%)	Job creation rate (%)	Average size	Relative size vs. incumbents
	Firms	Employment				
All administrative entrants	1.00	1.00	9.3	2.6	3.1	0.26
Narrower definitions of entrants						
- *De novo* entrants by official method	0.94	0.82	8.7	2.1	2.7	0.23
- *De novo* entrants by employee flow method	0.95	0.59	8.8	1.6	2.0	0.16
- *De novo* entrants (all methods)	0.91	0.56	8.4	1.5	1.9	0.16
De alio entrants	0.09	0.44	0.9	1.1	14.6	1.22

Entry rate: % of active firms - Job creation rate: % of total employment - Relative size: average size of entrants compared to average size of incumbents

Not distinguishing between *de novo* and *de alio* entrants when using administrative data sources, means mixing up two distinctly different populations: on the one hand firms that enter the market and are typically small, and on the other hand incumbents that entered an unknown number of years before, have survived till the period of observation, and are typically larger. In section III we show that these different stages of the selection process at which the two types of entrants are located, are clearly reflected in their growth and exit patterns. Confusing *de alio* entrants with firms that started *de novo*, leads to a strong bias in exit and growth rates, and hides the distinct relation between growth and exit and size among young firms.

For entry statistics, the bias *de alio* entrants introduce is especially clear in measures of job creation. Although they represent only 9% of administrative entrants, *de alio* entrants represent almost half of employment of administrative entrants (44%).[12] This employment should not be considered as job creation, since it is mainly the result of an administrative transfer of existing jobs to a new firm identifier. Therefore, not separating out *de alio* entrants leads to a strong overestimation of job creation by entrants. Our narrow identification of *de novo* entrants shows that the magnitude of job creation through entry is very modest: while a large number of new firms enter the market *de novo* every year (8.4% of all active employers), the jobs they create in the year of entry represent only 1.5% of total employment.

> Another aspect in which *de alio* entrants hide the true characteristics of entrants is their relation to the business cycle. Business formation is considered to be procyclical , and especially employment weighted entry is found to covary positively with output growth (Campbell, 1988). De novo entrants strongly reflect this feature (Graph A.1): annual changes in job creation are relatively small and show strong positive correlation with current gdp growth (0.81). This relation is much weaker in the group of all administrative entrants (0.43). The reason is that de alio entrants show almost no correlation with gdp growth (0.13), which is explained by the fact that they are mainly driven by legal, tax or administrative motivations and not by economic events. As a result, misinterpreting them as true entrants introduces substantial spurious variation in job creation by entrants and conceals their strong positive relation with cyclical fluctuations. The official method helps little to reduce this spurious variation; while the employee flow method, using the continuation of the production factor labor as a main criterion for continuing businesses, approximates the annual variation of de novo entrants.

Exits versus transfers

Employee flow methods have been mainly developed to produce more reliable statistics on firm and employment dynamics. They are used to remove 'spurious' firm entries and exits from the data set, as well as 'spurious' job creation and destruction resulting from the mere administrative reshuffling of employment between business numbers (Albaeck and Sorensen, 1998; Korkeamäki and Kyyrä, 2000; Baldwin, Beckstead and Girard, 2002; Persson, 1999). Some studies use employee flows to identify changes in firm structure such as mergers and acquisitions (Mikkelson, Unger, Lebel, 2006; Benedetto et al., 2007), or to examine a specific group of entrants such as spinoffs (Eriksson and Kuhn, 2006; Muendler et al., 2012). By our knowledge, no attempt has been undertaken so far to trace the employment histories of firms from the date of entry over various changes in firm identifiers they might be involved in. Because the aim of our study is to draw a clear picture of growth and exit patterns of young firms, we use the different link methods to follow the growth path of de novo entrants in the years after start-up. This exercise cannot be extended to far. Some entrants, and as we will see some of the most successful ones, are involved in successive identifier changes, mergers or split-offs as they age. To follow the 'true' employment growth of these entrants rapidly becomes complex, and even unfeasible. Therefore we restrict our analysis to the first five years after entry, for which we draw a reasonably accurate picture of firm-level exit and growth patterns.

Young firms may disappear from the data set because they change identification number or are merged to another firm. Miscoding such events as exits leads to an overestimation of failure rates. Surviving young firms may be affected as well by transformations of underlying identifiers, for example when they shift part of activities and associated work force to a newly created business number.[13] Employment of the origin firm will thereupon exhibit a negative shock. Misinterpreting this employment drop as job destruction leads to a negative bias in the firm-level growth rate. Events in which activities are relocated between business

---

[12] In several countries, also changes in ownership and legal structure lead to the assignment of a new business number, which is not the case in Belgium. We assume that the bias in job creation by entrants in these countries is even higher.
[13] The split-up of activities into smaller firms is a common practice in order to remain below the threshold of more stringent legal regulations.

numbers will be denoted as *transfers* of activities. We make use of both the official and the employee flow record linking methods to identify them.

The links between firms involved in a transfer are employed to more accurately estimate exit and growth patterns of *de novo* entrants. This is implemented as following. First, true *exits* are identified by separating them from closings of business numbers which result from a mere *transfer* of existing production factors to another firm identifier. Second, employment of young firms involved in a *transfer* is imputed by making use of employment of the linked successors. The imputation procedure is explained below.

In the first five years after start-up, the business number of half of de novo entrants is closed down. Closings of larger firms are more likely to correspond to a *transfer* than those of smaller ones. About 30 per cent of young firms with at least ten employees at the age they disappear from the data set, do not exit the market. They continue with a different identifier, employing largely the same work force (Table A.4 in Appendix). The employee flow links show that most transfers again correspond to plain ID changes (77%); 19 per cent of de novo entrants are absorbed by another business number; and 4 per cent, especially in larger size classes, are split up into several numbers. Miscoding these events as exits leads to an overestimation of exit rates, especially in size classes above ten.

Table 3. Share of transfers in all administrative exits at age 1 to 5 (avg. 2004-11)

		Size class before exit				
	Total	1-4	5-9	10-19	20-49	50+
Total share of transfers	0.03	0.02	0.17	0.28	0.30	0.29
- of which identified by employee flow method	0.01		0.14	0.26	0.29	0.29
by type of employee flow link						
1. one-to-one ID change	*0.01*		*0.12*	*0.20*	*0.15*	*0.17*
2. split into several business numbers	*0.00*		*0.00*	*0.01*	*0.06*	*0.10*
3. merged to another business number	*0.00*		*0.03*	*0.06*	*0.08*	*0.02*

> To recall, the employee flow method links a firm to a successor when at least 50 per cent of their work force is identical. This key rule is somewhat modified in case of split-ups and mergers. By definition, a closing business number is only linked when at least 50 per cent of its work force is transferred to a successor. In observation, this share is much higher, with an average of 86 per cent and a median of 88 per cent. Most transfers result from simple ID changes.

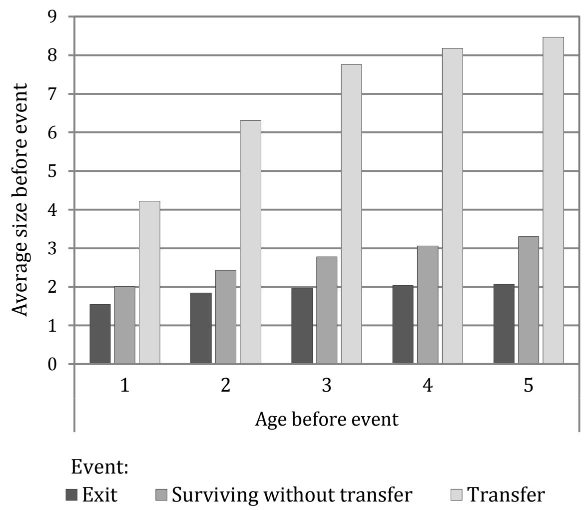
When activities are transferred between firm identifiers, the business number of the predecessor is mostly closed down. Yet 24 per cent of transfers also involve business numbers of de novo entrants which continue after the event. We compare the size of *de novo* entrants in either type of transfer to that of other entrants of the same age. Graph 2 shows the average number of jobs of de novo entrants by age, split up into three groups: firms that will *exit* in the next year, firms that will *survive* without being involved in a transfer, and firms that will be involved in a *transfer*. Survivors are on average larger than exiters, but *de novo* entrants involved in a transfer are two to three times larger than survivors.

> Firms that entered at a larger size or have grown rapidly (show in appendix?) are more likely to be involved in a transfer of activities between business numbers.[14] This does not come as a surprise. Successful young firms are

[14] Entry size of de novo not in transfer (1.9) versus de novo in transfer (4.0)

for example more likely to be taken over by an incumbent. Rapidly growing firms also have an incentive to split-up activities into smaller units in order to remain below the size threshold of more stringent legal regulations.[15] Yet the majority of transfers correspond plain one-to-one ID changes: firms that close down their business number and continue activities with a new identifier, employing more or less the same work force. Tax and liability reasons may be important motivations for ID changes, and again larger firms more often make use of it.

Graph 2. Average size of *de novo* entrants at age before event (avg. 2004-11)



Imputed employment after transfers

To impute the number of jobs of young firms in the years after a *transfer*, we link them to their successors. In case of one-to-one ID changes, the majority of events, imputed employment is given directly by the employment of the successor. The same holds for split-offs or break-ups: the imputed number of jobs in the subsequent periods equals the sum of jobs of the different successors. Only when the entrant is merged to another firm, its future employment levels cannot be observed. In that case, we make the assumption that the average contribution of the entrant to the employment growth of the merged entity equals its employment share at the start of the merger, and that deviations from this average are random. Therefore imputed employment of the firm in the years after the merger is equal to its employment share in the sum of the separate entities just before the merger. The technical details of the imputation procedure are provided in the appendix.

Young firms can be involved in a second or even more transfers in the first five years after entry. The probability of a second transfer, conditional on having experienced a first transfer, is much higher (9.2%) than the probability of a first transfer (3.5%). This indicates that firms, that have once relocated activities to another business number, have a strong incentive to further make use of this possibility to optimize their legal or organizational structure. Employment of a de novo entrant after a second transfer is kept constant at the level of imputed employment just before the transfer. This assumption should be considered as rather

[15] In Belgium, small firms (less than 100 employees and turnover not exceeding: 7,300,000 EUR and balance sheet total not exceeding: 3,650,000 EUR) are not subject to a number of legal obligations such as filing a full annual account, and the installation of a works council

cautious, since from the first transfers we know that employment on average increases after a transfer. A similar imputation as we did for the first transfers was however not feasible within the scope of this paper.

Misinterpreting transfers as exits and shocks to employment yields an overestimation of exit rates and an underestimation of growth rates, especially in larger size classes. Our main findings, however, do not depend on this specific subset of entrants. We provide a robustness check by excluding entrants involved in a transfer from the estimations.

IV. Patterns in firm dynamics

We estimate employment growth and exit patterns of young firms in the five years after entry. The results are based on the set of *de novo* entrants in the private, non-subsidized sector in the period 2004-2012. Our primary interest lies in how net employment growth differs across various aggregations of firms, more specifically, how it varies with size, within a given age class, and with age, conditional on size. We follow DRS (1989) who have highlighted that for the empirical analysis of these relationships, the mean growth rate of a given class of firms can be decomposed into the growth rate of surviving firms, weighted by the probability of survival, minus the probability of exit. Therefore, three measures are estimated: the exit rate, the growth rate of surviving firms, and its summary measure, the growth rate of all firms.

In the Appendix, we also provide results for *de alio* entrants, and results based on data where we do not distinguish between *transfers* and true *exits*. They are included to illustrate how an analysis based on a naive interpretation of raw administrative data, i.e. not taking into account that firms can change identifier between entry and exit, hides the distinct growth and exit patterns of young firms.

a. Measures

Growth and exit rates are estimated as year-on-year differences in a given period (year *t-1* - year *t*), averaged across entry cohorts. The annual point of observation is the end of the second quarter (June 30th). Employment is measured as the number of employees registered at June 30th, or as imputed employment for de novo entrants involved in a transfer. We present the results of employment-weighted regressions, where the weights are equal to the individual firm sizes, such that the estimates are readily interpreted as net employment growth rates of a given class of firms. Specifically, the estimated growth rates represent the net employment growth rate of a given age-size class of firms, and the exit rates represent job destruction rates due to exit.

The year of entry as is defined as the first year a firm records positive employment. It will be denoted as age 1, as opposed to age 0, the (unknown) point in time the firm is established. Clearly, the period between age 0 and age 1 is not observed, but if it is considered as a period in which a first selection process takes place, it partly explains why firms are of different size in the first year of positive employment. The distinction between age 0 and age 1 is motivated by two reasons. First, as the annual point of measurement is the end of the second quarter, the set of entrants in year *t* include all firms that started as an employer between July 1st of year *t-1* and June 30th of year *t*, minus the ones exited before June 30th. Firms which fail in the early months after entry will not be observed if their lifetime does not include June 30. The population of employer entrants thus includes survivors of a first selection process of unequal length. Second, firms may already have been in operation without employees before age 1, either as a legal company or as self-employed. For these firms, the decision to engage employees is already based on prior information about success. Also other reasons may motivate why the year of entry as an employer firm should be regarded as the result of an early selection process. For example, firms that enter at a larger scale have been successful in getting access to larger credits. In summary, the distinction we make between age 0 and age 1 explicitly denotes that there has been a period before entry as an employer during which firms have received some information about their cost efficiency. This distinction will also help to explain why firms of different size exhibit different survival and growth rates already at age 1. The first five years after entry as an employer will be subsequently referred to as age 2 to 6.

> In a conceptually more rigorous definition of entry, age 0 should be regarded as the moment the idea of establishing a firm was born in the mind of the manager. In the period between age 0 and the first year of positive employment, the manager learns about its own efficiency, either through observed costs if the firm is in operation as a legal entity without employees, or through other information channels, such as access to credit or feedback from a professional network. This rigorous concept of entry corresponds to the one employed by Jovanovic (1982), who assumes that all firms have the same prior beliefs about their true costs and therefore enter at the same size.

Exiters in observation period *t-1 - t*, are firms for which *t-1* is the last year of positive employment. As explained in the previous section, firms that change ID number, merge or split up, are not considered as exits, and their growth path after the event is based on imputed employment. The years between entry and exit, firms are denoted as survivors. Survivors may have zero employment in t-1 or t ('dormant' firms). They are treated as outliers and removed from the regressions in the periods concerned.

Following Davis, Haltiwanger and Schuh (1996a), firm-level growth rates are calculated as discrete-time growth rates using average size in the denominator. If $E_{i,t}$ is employment in year *t* of firm *i*, then its growth rate in period *t-1 - t* is $g_{i,t} = \left(E_{i,t} - E_{i,t-1}\right)/X_{i,t}$, where $X_{i,t} = \left(E_{i,t} + E_{i,t-1}\right)/2$. The averaging of the denominator results in growth rates ranging from -2.0 for exits to +2.0 for entrants, reflecting job creation and destruction symmetrically.[16] The mean growth (exit) rates we estimate for various aggregations of firms are weighted sums of firm-level growth (exit) rates, where the weights are based on the average employment size.

At each age, firms are grouped into five size classes: firms with one employee, 2 to 4 employees, 5 to 9 employees, 10 to 19 employees, and 20 employees or more. Exiters are a assigned to the size class of last year employment, but survivors are classified into size classes based on average employment in year *t-1* and *t*. Due to both the average size classes and imputed employment of firms involved in a transfer, the size variable is not an integer. Firms are therefore classified into the following continuous size intervals:]0,2[, [2,5[, [5,9[, [10,20[, [20,∝[.[17] The average size classification is used to mitigate the undesired statistical effects from using a conventional base-year classification (employment in *t-1*) leading in itself to an inverse relation between size and growth. First, if firms are classified in the size class of *t-1*, random fluctuations in size because of measurement error or transitory variance yield a regression bias in the relation between size and growth (Friedman, 1992). Davis, Haltiwanger and Schuh (1996b) illustrate that when using a base-year size classification, regression-to-the-mean effects alone yield an inverse relationship between size and growth.

---

[16] Conventional growth rates range between -1.0 and +∝

[17] For example, the smallest size class]0,2[in period *t-1 - t* includes: exiters with one employee in *t*; stable survivors with one employee in *t-1* and *t*; growing survivors from 1 to 2 employees; declining survivors from 2 to 1 employee.

The average size classification is relatively robust to this effect. A second source of bias when using base-year size classes stems from the fact that, in the subset of survivors, firm employment is bounded below by one, and so the lower tail of possible rates of decline is truncated for smaller firms. This is the most clear for one-employee firms, of which survivors cannot have negative growth rates. Therefore, the base year classification would yield an inverse relation between growth and size in the subset of surviving firms, even if fluctuations in size would be independent of size (Baldwin and Picot, 1995). This bias is avoided in the average size classification, because the range of possible growth rates within each size class is symmetrically distributed around zero.

In a population of where firms fluctuate around a stable long-run size, the average size classification would yield unbiased results. In a population, however, where growth rates are, on average, positive, it in turn introduces a positive relation between growth and size for firms crossing size class borders. In section IV f, we show that our findings are robust to average size classes by presenting the results using two alternative classifications that approximate a continuous size-growth relationship: i.e. dynamic size classes in which a firm's employment growth or loss is allocated to each respective size class in which the growth or loss occurs, and results based on average of growth rates regressed on the employment levels in t-1 and t.

We do not include more detailed size classes above 20 employees by lack of sufficient observations to obtain significant results. This is first due to our focus on the population of *de novo* entrants, of which only a small number attain a large size in the first five years after entry; moreover, the full interactions of age-size classes with industry in our empirical specification (see below), do not allow us to obtain significant results in more detailed size classes.

b. Empirical specification

The estimations are based on firm-level regressions with the growth or exit rate as dependent variable and age and size class as key explanatory variables.[18]

We estimate a saturated dummy regression model to characterize the exit and growth patterns and plot the coefficient estimates on graphs. The standard deviations are included in the tables. The model includes a separate indicator for all possible values taken on by the discrete explanatory variables: besides the interaction terms of the key explanatory variables, age and size class, we include industry dummies, the interactions of age-size and industry, and additive year dummies. As Angrist and Pischke (2009) highlight, saturated regression models fit the conditional expectation function perfectly, regardless of the distribution of the dependent variable, because it is a linear function of the dummy regressors used to saturate. We use employment-weighted least square regression such that our estimates represent net employment creation (or destruction) of a given class of firms.[19] In the growth specification, the dependent variable is the employment weighted growth rate of firm *i* in period *t-1 - t*: $g_{it}^{w} = (X_{it}/\sum_i X_{it})\, g_{it}$; in the exit specification, it is the employment weighted exit rate of firm *i* period *t-1 - t*: $e_{it}^{w} = (X_{it}/\sum_i X_{it})\, e_{it}$, where e_{it} is a dummy for exit period *t-1 - t.*

[18] The regressions include 365 000 firm-year observations of de novo and of 44 000 de alio entrants.

[19] Survivors with zero employment in *t-1* or *t* ('dormant' firms) have the extreme values of growth rates -2.0 or +2.0. They are treated as outliers and removed from the regressions in the periods concerned.

For each dependent variable $y_{it}^{w} = (g_{it}^{w}, e_{it}^{w})$, the regression model is written as:

$$y_{it}^{w} = \sum_{j=1}^{5}\sum_{k=1}^{5}\alpha_{jk}(age_{jit}size_{kit}) + \sum_{d=1}^{n}\beta_d D_{di} + \varepsilon_{it}$$

where $age_{jit}size_{kit}$ are interaction terms of the age dummies indicating age 1 to 5, and the size dummies distinguishing the 5 size classes;

D_{di} are industry and year dummies, and interactions with the age-size classes. They include
- 5 normalized industry dummies: $\widetilde{ind_{il}} = ind_{il} - ind_{i6}$, where ind_{il} and ind_{i6} are dummies for firm *i* in industry *l* and the reference industry manufacturing (the classification into 6 industry groups is listed in Table A.2 in the appendix);
- all interactions of age-size and (normalized) industry, leaving out one reference size class;
- normalized year dummies: $\widetilde{y_{it}} = y_{it} - y_{i12}$, where y_{it} and y_{i12} are dummies for firm i in year t and 2012;

ε_{it} is a random error term.

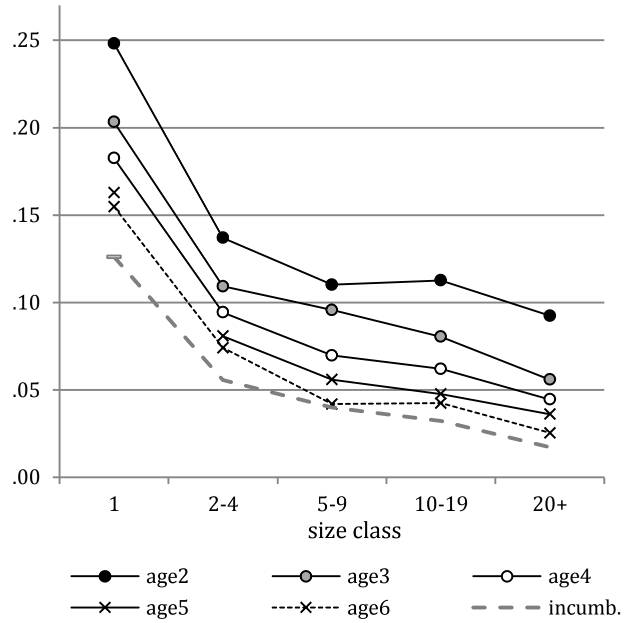
The coefficients of interest are α_{jk}. In the growth specifications, they represent the net employment growth rate of a given age-size class of firms, and in the exit specification, they represent the job destruction rate due to exit.

c. Exit rates

Graph 3 plots the age-size coefficients for the exit rates. They represent job destruction rates caused by exit for a given age-size class of de novo entrants. The exact coefficient estimates and standard errors are reported in Table A.4 in the appendix.

Many studies have found an inverse relationship between the exit rate and age, even when size is held constant; empirical evidence also suggests an inverse relationship between the exit rate and size within distinct age groups. These patterns are strongly evident for *de novo* entrants at age 2 to 6.

Graph 3. Exit rates of *de novo* entrants (employment weighted)



Note: Incumbents are firms with traceable age older than 6

First, within every size class of firms, job destruction rates due to exit decrease with age, indicating less failure as firms age. This can be seen from the vertical downward shift of the plotted coefficients. Exit rates are especially high in the first year of existence and strongly decline with age. Five years after entry, exit rates have been strongly reduced within every size class, and have come close to those of older firms.

Second, within every age group, exit rates are inversely related to firm size. Exit rates are especially high among one-employee firms, and decline with increases in firm size. The downward trend is generally monotonic except in the largest size classes at age 2. The distance between the exit rates of different size classes is similar across age groups of young firms: exit rates of one-employee firms are about 10 percentage points higher than in size class 2-4, which in turn exceed those of the largest size classes by about 5 percentage points. Similar differences are found across size classes of older firms. This suggests that the effect of size on the probability of exit remains equally strong in the five years after entry, as it is at older age.

The results are consistent with a model of passive learning which assumes that firms only learn about their efficiency level with the passage of time (Jovanovic, 1982). While the efficient survive and the inefficient fail, the average efficiency of the survivors improves from period to period, inducing less failure as the cohort ages. Our results show that this reduction in failure rates is concentrated in the very first years after entry. Within a given age cohort, larger firms are the ones that have received favorable cost information in previous periods, and subsequent information is less likely to be unfavorable enough to induce exit. Firms that have grown more slowly, by contrast, are exactly those with higher expected costs and so have lower value of staying in the industry. Or, at the point of failure, firms are smaller than the surviving members of their cohort.

De novo versus de alio entrants - exits versus transfers

Exit rates of young firms based on a straightforward analysis of administrative data are biased by *de alio* entrants and *transfers* after entry. The biases these two types of administrative events introduce are yet in the opposite direction. Table A.7 in the Appendix presents the coefficient estimates based on different treatments of the sample of entrants. The first three columns show exit rates of all administrative entrants and the two subsets, *de alio* and *de novo* entrants, based on a plain administrative identification of exits, i.e. a firm is considered as an exit when its business number disappears from the data set. The last column repeats the final coefficient estimates for *de novo* entrants we use in this paper: i.e. firms that continue activities with a different business number are identified as *transfers* and removed from the population of exits.

De alio entrants, defined as new business numbers that are continuations of established firms, can be expected to show exit patterns that are more similar to those of incumbents than of *de novo* entrants. Indeed, within each size class, exit rates of *de alio* entrants (column 2) are much lower than those of *de novo* entrants (column 3). They also show little, or sometimes random, variation by age, which follows from the fact that the count of years after entry does not correspond to the real firm age. As a result, mixing up *de alio* and *de novo* entrants partly hides the distinct exit patterns of young firms, as can be seen from a comparison of columns 1 (all administrative entrants) and column 3 (*de novo* entrants): it first leads to an underestimation of real exit rates of young firms, especially in larger size classes, where *de alio* entrants are overrepresented; and second, it yields an underestimation of the inverse relationship between exit and age.

Singling out the population *de novo* entrants does not lead in itself to more satisfactory results: entrants may be miscoded as exits if they are involved in a *transfer* after entry. This in turn introduces an upward bias in the exit rates. Graph 2 showed that larger firms are again overrepresented in this kind of transformations of firm identifiers. If *transfers* are accounted for, by linking subsequent identifiers of the same firm, exit rates of *de novo* entrants again decrease, as is clear from comparing column 3 and 4. The final effect on the estimated exit rates is mixed: at young ages, exit rates of *de novo* entrants corrected for *transfers* (column 4) are generally higher than exit rates in the untreated sample of administrative entrants (column 1), and lower in older cohorts. In summary, when focusing on *de novo* entrants that truly *exit*, the inverse relation between age and exit is more strongly manifest, both within size classes and overall.

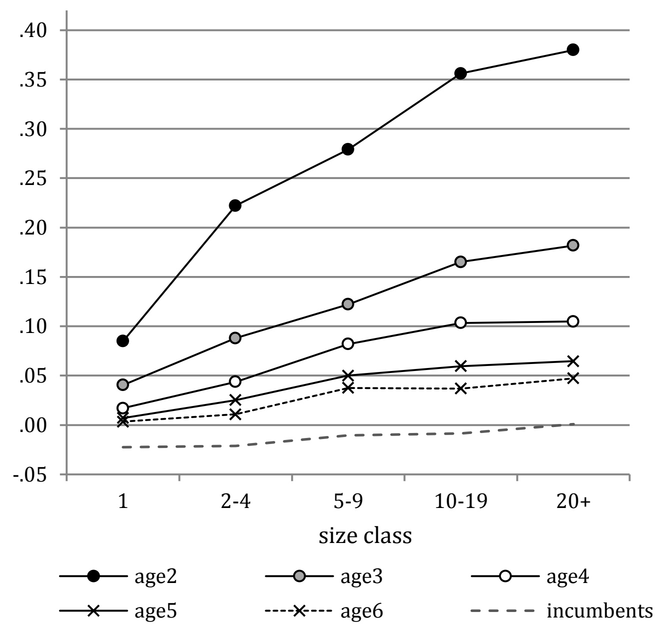
> The narrow focus on *de novo* entrants, after removing entrants that started *de alio*, and on real *exits* versus *transfers* of activities, yields results that are not essentially different from a straightforward analysis of administrative data. The reason is that *de alio* entrants and exits through *transfers* mostly occur jointly, in the same size classes, and introduce opposite biases in the exit rates. This could explain the strong similarities found in many empirical studies on the relation between exit rates, size and age: whatever the incidence of changes in firm identifiers in the underlying data set, the opposite biases introduced by *de alio* entrants and *transfers* keep each other in balance and yield exit patterns that are relatively robust to miscodings of administrative entry and exit. As we will see below, this is not the case for growth rates, where the sum of the two effects is not neutral.

d. Growth conditional on survival

Graph 4 presents the regression coefficients for the employment weighted growth rates of *de novo* entrants that survive in a given period. They represent net employment growth rates of survivors within a given age-size class of firms. The point estimates and standard errors are reported in Table A.5. in the appendix.

The estimated coefficients presented show strongly regular patterns in the relationship between growth rates of surviving young firms and the firm age and size class. First, in the early years after entry, surviving young firms of all sizes exhibit high growth rates which rapidly decline with increases in age. The decrease in growth rates with age is monotonic within every size class, as can be read from the downward shift of the plotted coefficients. The strongest reduction occurs in the first four years after entry. At age 6, growth rates of surviving young firms are still positive and about 4 percentage points higher than among more mature firms of the same size, which suggests that the firm distribution will further to shift to the right.

Graph 4. Growth rates of surviving de novo entrants (employment weighted)



Note: Incumbents are firms with traceable age older than 6

The inverse relation between growth rates and age of surviving firms is consistent with the passive learning model (Jovanovic, 1982) which assumes that a firm typically enters at a small size and, as it learns about its true costs, adjust its size to the expected profit maximizing level. Firms that discover they are more efficient, grow and survive, while the inefficient will decline and eventually fail. This induces sizes of firms of the same cohort to diverge. When estimates about cost efficiency become more precise as more evidence comes in with the passage of time, there is less need for further size adjustments, and growth rates decrease.

Second, the age-size coefficients in Graph 4 show that among young firms of the same age, large firms have higher growth rates than smaller ones. The positive relation between growth and size is especially clear in the very first years after entry, and moves towards more proportional growth rates as a cohort matures. In other words, the dispersion in growth rates across size classes is gradually reduced with increases age. At age 6, differences in growth rates by size are mostly statistically insignificant at the 5 per cent level. The results suggest that Gibrat's law is strongly violated for surviving young firms of the same age, yet as a *positive* relation between size and growth.

A number of important studies have found that smaller firms have higher growth rates than larger ones (Hall, 1987; Evans, 1987; DRS, 1989). If, however, as suggested by Jovanovic (1982), the inverse relationship between growth and size primarily results from the fact that small firms are younger and young firms have higher growth rates, the same relationship does not need to hold within specific age groups. Graph A.5 in the Appendix indeed shows that in our total sample of active firms between 2003 and 2012, a positive relation between growth and size, both within young and older cohorts, coexists with a negative size-growth relationship unconditional on age.

Although empirical studies are not univocal, it is often presumed that the inverse relation between growth rates and size should also be found when holding age fixed (see Caves, 1998). Evans (1987) and DRS (1989), indeed find that growth rates of surviving firms are inversely related to size conditional on age.[20] [21] Haltiwanger et al. (2013) clearly illustrate that higher growth rates of small firms are entirely attributable to young firms being small. Once controlling for age, they find no (or a mild positive) relation between size and growth rates of surviving firms.[22] While these studies examine the size-growth relationship in the population of firms of all ages, we complement them by a narrow focus on *firms of the same age in the first years after entry*. As our findings suggest, the size-growth relationship in the early years after entry is entirely different from that at later age.

The positive relation between growth and size in the early years after entry is consistent with a model of passive learning where all firms enter at a similar small size, and sizes diverge as some firms find that they are more efficient and grow, while the less efficient ones stay small.

If, as suggested by Mata et al. (1995), adjustment costs prevent young firms to fully adjust their size instantaneously in response to an observed cost efficiency level, then current size is likely to be different from desired size, and thus, for expanding firms, current size will understate the desired size. In other words, if expansion can only be made gradually, larger firms, which are the ones that have received favorable cost information in the previous periods, will also be likely to be growing in the future, and a positive correlation between size and growth will be observed. Firms that have stayed small, by contrast, are the ones that have received less favorable cost information; they will be more likely to decide upon lower growth rates or even withdraw from the industry. As revisions of estimated efficiency become smaller for all firms of the same cohort with the passage of time, and firms approach their optimal scale, the differences between growth rates across size classes decreases and firms move towards a more proportional growth distribution. The estimated age-size coefficients in Graph 4 exactly reflect this early adjustment process.

The positive relation between growth and size in the first years after entry will only be observed if all firms enter at a small scale which is suboptimal for the more efficient ones. In the theoretical setting of Jovanovic (1982), size is even identical for all entrants because each firm starts with the same prior beliefs about its efficiency. The population of *de novo* entrants indeed approximates this feature: 94 per cent enters with less than 5 employees and entry sizes larger than 50 employees are extremely rare (0.05 per cent). In the following paragraphs we show that if *de alio* entrants, which typically enter at a large size, are not adequately

[20] The DRS (1989) results need not necessarily to contradict our findings since they are based on broad age cohorts (five-year classes), excluding firms with less than 5 employees, in a sample of manufacturing firms. When we imitate this setting by restricting our sample to manufacturing firms with more than 5 employees, growth rates of surviving de novo entrants, averaged over age 2 to 6, show a similar a negative relation with size.

[21] Lotti, Santarelli, Vivarelli (2001).

[22] Some other studies estimate the effect of initial size on post-entry growth rates (Audretsch, Santarelli and Vivarelli, 1999). Here as well, an inverse relationship between size and growth is found.

filtered out, the positive relation between growth and size is entirely obscured. However, even in the population of *de novo* entrants, we do observe different entry sizes, which, moreover, have predictive power for growth rates already in the first year after entry: firms that enter with a larger number of employees and survive, show higher net growth rates at age 2 than the ones that entered small. This observation is in line with our remark earlier that the employer entry size should be interpreted as the result of a first selection process between the unobserved moment the firm is created (age 0) and the first observation of positive employment (age 1). If larger entrants are indeed considered as the more efficient survivors of this first selection process, the higher growth rates they exhibit already in the first year after entry are fully consistent with the passive learning model.

De novo versus de alio entrants

The positive relation between growth rates of surviving firms and size is entirely absent in the sample of administrative entrants (Graph A.6 panel a.). Instead, the estimated age-size coefficients in this sample suggest an inverse U-shaped relation between growth and size at age 2, and growth rates that are proportional to firm size at age 3 to 6. This result seems to be more in line with other empirical studies. The absence of increasing growth rates with size, however, is entirely attributed to the presence of *de alio* entrants in larger size classes, where they introduce a strong downward bias in the growth rates. This inverts the positive relation found among *de novo* entrants into a negative or a proportional one, depending on the age. Graph A.6 and Table A.8 in the Appendix show the bias *de alio* entrants introduce in the relationship between growth rates and size of young firms.

De alio entrants, which are transformations of established firms, exhibit low growth rates which do neither systematically depend on size nor on age (Graph A.6 panel b.). Growth rates do not depend on age because *de alio* entrants are a set of firms of mixed, undefined maturity. Growth rates are much lower than those of *de novo* entrants because *de alio* entrants are older than their administratively recorded age. It are firms in a more advanced stage of the selection process where there is less need for size adjustments, which explains why growth rates are close to those of established firms of older age.

In summary, both exit and growth rates illustrate that *de alio* entrants lack the typical post-entry dynamics of young firms. They instead exhibit characteristics of more mature firms, simply because it is the population they are part of. Because *de alio* entrants are overrepresented in larger size classes (see Table 1), the bias they introduce is disproportionally found there. If they are not adequately filtered out from the population of administrative entrants, *de alio* entrants entirely obscure the positive size-growth relationship firms exhibit at early ages. Therefore, an analysis of administrative data which fails to focus on *de novo* entrants only, would incorrectly support the belief that the proportional or decreasing growth rates with size, as found in the entire population of firms of all ages, is also a feature of young firms in the early years after entry.

For completeness we remark that, in contrast to the exit analysis, correcting for firms that are involved in a *transfer* (change identification number) does not yield significantly different results (Graph A.6 panel c. and d.). The reason is that when controlling for *transfers*, we primarily include an additional set of firms to the population of survivors which would otherwise be misclassified as exits. This set, however, exhibits *imputed* growth rates which are highly similar to growth rates of other de novo entrants of the same age and size. Therefore, estimated coefficients are not significantly affected, only standard errors after correcting for *transfers* are somewhat smaller.

e. Growth all entrants

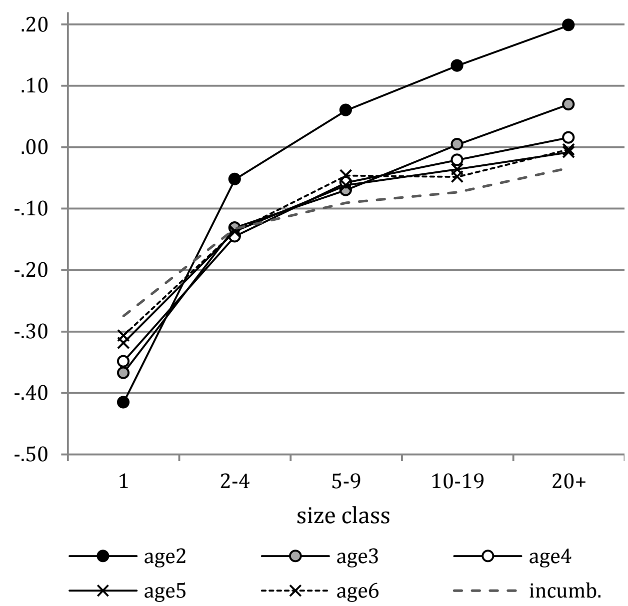
Entrants contribute positively to aggregate employment by the jobs they create in the year of entry. Whether initial employment each cohort creates rises or falls in the years after entry, depends on the combined effect of exit and growth of survivors, and the changes in the size distribution over time. In this short section we look at the net growth rates of all *de novo* entrants within specific age-size classes. In the next section, we discuss the aggregate job creation of a given cohort of entrants, and the contribution of young firms to total employment.

Net employment growth rates of all firms of given age-size class are equal to growth rates of surviving firms, weighted by the probability of survival, minus job destruction rates of exiters, weighted by the probability of exit. Therefore, if young firms of a given size have both higher exit rates than older firms and higher growth rates when they survive, it is undefined whether net job creation of all firms within size classes will increase or decrease with age. Graph 5 plots the estimated age-size coefficients of all *de novo* entrants at age 2 to 6.

Although exit rates are high in the first year after entry, the higher growth rates of survivors sufficiently compensate it to yield higher net job creation rates for older firms of the same size. Only among one-employee firms, job destruction due to exit overwhelms job creation of survivors, and net employment growth is far below that of older firms of the same size. The inverse relation between growth and age conditional on size (the downward shift of the plotted coefficients) is yet absent from age 3 onwards, when exit rates are still high and growth rates of surviving firms have already strongly declined. Only the largest size classes exhibit some decline in growth rates form age 3 onwards.

By contrast, the positive relation between growth and size of all firms of the same age is stronger than it is for surviving firms only, because larger size classes are characterized by both lower exit rates and higher growth rates among surviving firms. As a result, within each age class, Gibrat's Law is more severely violated for all firms than it is for surviving firms.

Graph 5. Growth rates of all de novo entrants (employment weighted)



Note: Incumbents are firms with traceable age older than 6

Growth rates do not remain positive in all size classes. More specifically in the smallest size classes, net growth rates turn negative already at early ages. One is curious about the effect of these negative growth rates on aggregate employment growth of young firms . If the share of small firms, which we showed is extremely large at entry, remains sufficiently large, total employment created at entry will fall as cohorts age. If by contrast, the size distribution moves sufficiently rapidly to the right, the positive growth rates of larger firms could yield a net increase of aggregate employment in the years after entry. This is discussed in section V.

For comparison, we include the estimated growth rates based on the untreated sample of administrative entrants in Graph A.7 in the Appendix. Due to the presence of *de alio* entrants and transfers, which introduce the biases in the exit and growth rates we discussed, the positive relation between growth rates and the firm is absent in size classes with more than 5 employees.

f. Robustness checks

The presented patterns are robust to alternative measurements and different subpopulations. Since the main contribution of this paper for the analysis of post-entry dynamics is the relation between net growth rates of survivors and firm age and size, we conduct robustness analysis for this result. We first present a robustness checks for the method of imputed employment and next look at the exit and growth patterns in two periods (before and after the recession) and in different industries. The final paragraphs present results for alternative size classifications.

Controlling for imputed employment

For de novo entrants that are involved in a transfer of activities after entry (ID-change, merger, split-up), employment has been imputed as explained in section III. We check whether our results are robust to this treatment by excluding firms involved in a transfer from the estimations. Graph A.8 and Table A.9 show that growth rates of survivors are nearly identical to those in the full population of de novo entrants. Standard errors are slightly larger for large size classes due to the smaller sample.

Pre- and post-recession entrants

The estimated growth rates discussed in this paper are averaged across cohorts of firms which entered between 2003 and 2011. This period includes the recession of 2008-09 during which aggregate employment dropped by 2.5 per cent and the number of de novo entrants fell by 12 per cent. Both employment and entry only slightly recovered in the following years and have not returned to the pre-crisis level by the end of the observation period. The empirical specification that is estimated in this paper includes normalized year dummies to control for business cycle effects, yet post-entry growth patterns may differ for firms that entered in the period of high economic growth before 2008, and the ones that entered during the recession or in the period of slow recovery afterwards. Graph A.9 and Table A.10 present the estimated coefficients for these two sets of de novo entrants. The growth rates in the pre- and post-recession cohorts follow the same regularities by age and size as found in the full-period sample.

Industry patterns

The empirical specification used to estimate exit and growth rates includes normalized dummies and their interactions with the age-size dummies for six industries: Manufacturing, Construction, Trade, Accommodation and Food services, Business services, and Mixed household and business services. This implies that the age-size coefficients represent average exit or growth rates of a given class of firms, where each of the six industries are given equal weight. We test whether the observed patterns can also be found in the individual industries. Graph A.10 presents the coefficient estimates for growth rates, based on the same regression model as explained in section IV.b, leaving out the industry dummies and their interactions with age-size.

Growth rates of surviving de novo entrants in separate industries mainly vary with age and size according to the same regularities as found in the total private sector. In all industries, growth rates are high in the first year after entry and strongly decrease with age within each size class. Five years after entry, firms exhibit still positive growth rates and exceed the about zero growth rates of older incumbents, but the difference is reduced to a few percentage points. Only in Accommodation and food services, where average firm size is small there is little room for size diversification after entry since, the higher post-entry growth rates are limited to the first year after entry, and thereafter approach those of older incumbents.

All industries, except Accommodation and food services, exhibit a clear positive relation between growth and size within each age class as found in the total private sector. Growth rates of surviving firms strongly increase with size in first years after entry, and move towards a more proportional distribution as cohorts age. In most industries, however, growth rates in the two largest size classes are more or less similar, and in Manufacturing they are even lower for firms with more than 20 employees. Only in Mixed business and household services, growth rates at each age monotonically increase with size, also in the largest size class.

These patterns seems to be uncorrelated with aggregate industry employment growth and industry-specific firm size: Mixed services has been declining over the observation period, just like Manufacturing, while Construction and Business services are an expanding sectors; both in Mixed services and Manufacturing, average firms size is large, while it is small in Construction and Trade.

Alternative size classifications

As a robustness check of the average size classification used throughout this paper, we present the results for growth rates of surviving de novo entrants using alternative size classifications. We first discuss classifications through which an approximation of a continuous growth-size relation can be obtained: a dynamic size classification, and results based on the average of growth rates regressed on the employment levels in *t-1* and *t*.

In a dynamic size classification, a firm's employment growth between the annual point-in-time observations of year *t-1* and *t* is allocated to each respective size class in which the growth occurs. Firms are initially assigned to a size class based on their employment in *t-1*, but are re-assigned to a new size class when a size-class threshold is crossed. Dynamic-sizing assumes linear employment growth from one period to the next and would equivalent to measuring continuous employment growth, for example on a daily basis, if employment change occurred linearly within the year. Dynamic sizing is used by the Bureau of Labor Statistics for the presentation annual and quarterly Business Employment Dynamics, where it has been developed to avoid regression biases from a base-year classification (Butani, Clayton, Kapani, Spletzer, Talan, and Werking, 2006). For the estimations presented in Graph A.11 Panel a. in the Appendix, we have chosen the size class thresholds such to allow symmetric and equal ranges of growth rates between -0.67 and +0.67 within each class.[23] This way we avoid any bias that might result from size classes allowing different ranges of growth and decline. The coefficient estimates presented in Graph A.11 Panel a. show the same relation between growth rates of young firms and the firm age and size as found when using average size classes. Although the differences in growth rates are less pronounced in the dynamic size classification, employment growth rates decline with age when holding size fixed, and increase in size within every age class. The positive relation between growth and size is the strongest at age 2, and moves towards a more proportional distribution with increases in firm age.

A second way to approximate estimations of a continuous growth size relation is to average employment growth rates evaluated first at the firm level size in t-1 and next in t. Employment growth rates are first estimated using base year employment in t-1 in the denominator: $g_{i,t}^{b} = \left(E_{i,t} - E_{i,t-1}\right)/E_{i,t-1}$, and a second time with end year size t in the denominator: $g_{i,t}^{e} = \left(E_{i,t} - E_{i,t-1}\right)/E_{i,t}$. The two growth rates receive an equal weight of 0.5 and are allocated to the firm size class in *t-1* and *t* respectively. The resulting growth rates are equal to employment growth evaluated ate the mean of employment in t-1 and t: $g_{i,t} = \left(E_{i,t} - E_{i,t-1}\right)/\left(\left(E_{i,t} + E_{i,t-1}\right)/2\right)$. Graph A.11 Panel b. in the Appendix shows that the age-size coefficients we obtain with this estimation method are similar to the coefficient estimates when using the average size classification.

For comparison, we present the coefficient estimates that would be obtained when using a base year size classification, i.e. firm level growth rates are allocated to the size class defined by the employment level in t-1.

---

[23] When a firm moves from one size class to another by either expanding or contracting, the single job that moves the firm across the threshold is allocated to the smallest of the two size classes.

Graph A.12 in the Appendix shows that the estimated relationship between growth rates of survivors and firm size shows little similarity to the results we found when using the average size or alternative classifications: one-employee firms have extremely high growth rates compared to other size classes, and growth rates are more or less proportional across other size classes, except for age 2 where they are declining in the two largest size classes. As discussed before, a base year classification yields an inverse relation between growth and size, even if growth rates would be independent of firm size. The first reason is that in the population of surviving firms, employment is bounded below by 1, and therefore, the possible range of growth rates in smaller size classes is more left-truncated than in larger sizes. This statistical property is most obvious for one-employee firms, which cannot have negative growth rates when they survive. The high growth rates of one-employee firms compared to other size classes are thus an undesired statistical property of the base year classification, which provide little information about the relationship between firm growth and size in the population of surviving firms. A second source of bias in the base year classification stems from regression-to-the mean effects induced by random or transitory variance in employment. If this leads to firms crossing size thresholds twice in one direction and the other, growth rates will be allocated to the smaller size class in the period of employment increase and to the larger one in the period of decline, yielding an inverse relation between growth and size. The possible effect of this bias is simulated in Graph xx in the Appendix. Panel a. presents net employment growth rates of all de novo entrants at age 3 to 6, classified into the base year size class, and panel b. repeats the same estimations based on a population from which we excluded period-observations of firms that change size class for only one period, reverting back to the previous class in the next period. While the size-growth relation suggested by panel a. is undefined, the coefficient estimates from panel b. show a clear positive relationship.

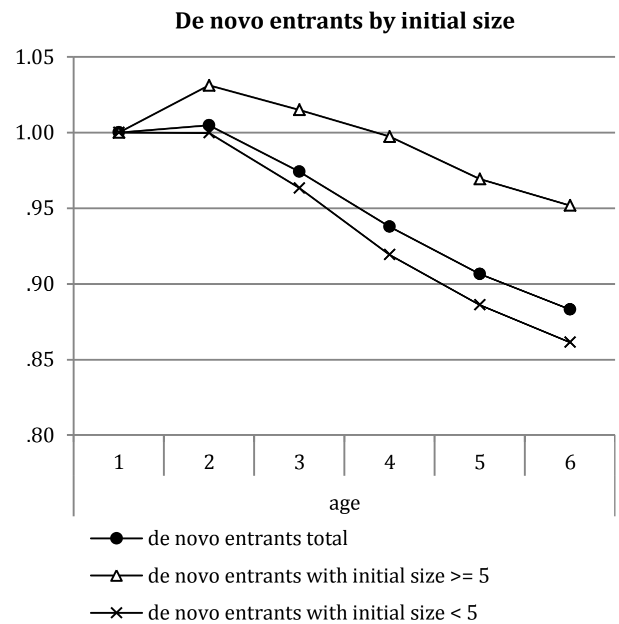
V. Job creation by entrants

A clean identification of entry and exit is informative about job creation by entrants. Empirical studies usually find entrants' shares in total employment below 5 per cent, and cohort's initial employment declining over time (Dunne, Roberts and Samuelson, 1989b; Boeri and Cramer, 1992; Mata et al., 1995). This section combines all three items discussed earlier: the initial size distribution, and entry and growth rates by firm age and size.

In Table 2 we showed that jobs created by *de novo* entrants represent only a very modest share of total employment (1,5%). How this initial employment share will change after entry depends on the net result of job loss due to exit and job growth of surviving firms. In the previous section we discussed how individual firm dynamics differ by size and change with age. We found that net job creation, after subtracting job loss due to exit, remained positive in larger size classes in the five years after entry, but turned negative for the smallest size classes already at early ages. The changing size distribution of younger firms will thus eventually determine whether an entrant cohort's initial employment share will rise or fall over time. If it rapidly moves to the right, the high growth rates of larger firms could compensate for the job loss in smaller sizes and the cohort's employment would increase. If, by contrast, the share of small firms remains sufficiently large, the cohort's employment will decline.

Graph A.13 in the Appendix shows the shift in the employment distribution of *de novo* entrants. Similar to Cabral and Mata (2003), we find that the initially right-skewed distribution moves towards the right as cohorts age. The most important change occurs in the smallest size class. Employment shares of one-employee firms fall from 0.35 to 0.10 in the first five years after entry. The share of size class 2 to 4 also decreases but much less dramatically (from 0.36 to 0.29). Employment shares of all other size classes increase. By about 8 percentage points in middle size classes, and only moderately for size classes with more than 50 employees. In other words, the share of the smallest size classes remains important at all ages, and will strongly counter the positive job creation realized by larger young firms.

Graph 6. Predicted employment growth of young firms after entry (index entry year age 1 = 1)



Graph 6 shows predicted employment growth of young firms in the five years after entry. Initial employment of an entrant cohort is indexed to 1 in the year of entry (age 1) and subsequent employment growth is based on coefficient estimates of a similar regression model as explained in section IV.b. The model now only includes the age dummies as main explanatory variables and, as before, normalized year dummies, normalized industry dummies and their interactions with age. We also show the results of separate estimations for two subsets of de novo entrants: firms that enter with less than 5 employees, and the ones that have at least 5 employees in the year of entry.

Similar to Boeri and Cramer (1992), we find that a cohort's initial employment remains stable in the first year after entry and then decreases monotonically with age. Five years after entry, employment has dropped to 88 per cent of its initial value. This means that the growth rates of surviving firms are insufficiently high to compensate the job loss due to firm failures. Being very small at entry seems to have some predictive power for employment growth. Total employment of the large subset of firms that start up with less than 5 employees (70 per cent of all de novo entrants) more rapidly declines than that of larger entrants and strands at 86 per cent of its initial value five years after entry. Employment of larger entrants, by contrast, increases in the first year and more slowly decreases thereafter. It has dropped to 95 per cent of its initial value five years after entry. This result is in line with the interpretation of initial size suggested before: if employer size at entry is the result of a selection process between the unobserved point in time the firm is created and the first observation of positive employment, larger entrants should be considered as the more efficient survivors of this first selection process.[24] They are exactly the ones which will have lower post-entry exit rates and higher growth rates, which leads to the result that they better succeed to maintain the employment level created at entry, than small entrants do.

De novo versus administrative entrants

In a straightforward analysis of administrative data, the effect of initial size would be reversed. This is presented in Graph A.14 in the Appendix. Total initial employment of the set of firms starting with at least 5 employees starts to decline immediately after entry, while for firms starting with less than five employees, it first rises and declines only from year two onwards. Five years after entry, the subset of small entrants has preserved 85 per cent of its initial employment level, while it is only 81 per cent in the set of larger entrants. This opposite result, where the aggregate set of small entrants shows better post-entry employment growth than the set of larger ones, is entirely driven by changes in firm identification numbers which lead to miscodings of entry, exits and true firm-level employment growth. It incorrectly supports the perception that initial size is negatively related to post-entry performance.

The different effect of initial size in our carefully constructed sample of de novo entrants and a naive reading of administrative data is mainly explained by two effects. First, de alio entrants introduce a downward bias in employment growth of young firms, especially in entry size classes above 10 employees where they represent more than 50 per cent of administratively recorded new firms (Table 2). This leads to an underestimation of aggregate employment growth of the subset of firms with a larger initial size. It also leads to an underestimation of aggregate employment growth of the total cohort of entrants, because de alio entrants represent 44 per cent of employment of administratively recorded entrants. Second, the better performance of very small entrants immediately after entry is explained by transfers: firms that transfer (part of) their activities to a new firm identification number, for example in case of a merger or split-up, often first put one or two employees in place to administratively prepare the transformation; the new firm is then classified into the smallest size classes at entry. Once the transfer is actually carried out and a cluster of employees is transferred from the old to the new firm identification number, a jump in employment is recorded which typically amounts to many times its initial value. Though small in number, this type of transfers makes the difference between an administrative recorded increase in aggregate employment at age 1, and a decrease when transfers are corrected for.

By industry

We find declining initial employment with age in all industries except in Manufacturing, where employment of entrants increases with more than 10 per cent in the first year and only slightly decreases afterwards, and in Business services, where it increases to 112 per cent of its intimal value in the two years after entry and remains stable afterwards (Graph A.15). In Accommodation and food services, initial employment has dropped most dramatically to only 54 per cent after five years, while it falls to about 90 per cent of its initial value in the three other industries. Very small entrants (less than 5 employees) show lower aggregate employment growth than larger entrants in all industries except Manufacturing and Construction, where the effect of small initial size is the opposite.

[24] Mata (1995): which is also quite reasonable, since this is exactly the type of entrant to whom the marginal benefits of learning may be greater.

VI. Conclusions

To be completed

References

Abowd, J.M. and Vilhuber, L. (2005) The sensitivity of economic statistics to coding errors in personal identifiers. *Journal of Business and Economic Statistics*, 23:2, 133-152.

Albaek, K. and Sorensen, B.E. (1998) Worker flows and job flows in Danish manufacturing, 1980-91. *The Economic Journal*, 108, 1750-1771.

Angrist J.D. and Pischke J.S. (2009) *Mostly harmless econometrics. An empiricist's companion*. Princeton and Oxford: Princeton University Press.

Audretsch, D.B., Santarelli, E., Vivarelli M. (1999) Start-up size and industrial dynamics: some evidence from Italian manufacturing. *International Journal of Industrial Organization*, 17, 965–983.

Baldwin, J.R. and Picot, G. (1995) Employment Generation by Small Producers in the Canadian Manufacturing Sector. *Small Business Economics*, 7:4, 317-331.

Baldwin, J.R., Beckstead, D. and Girard, A. (2002) The Importance of Entry to Canadian Manufacturing. OECD STI Working Paper No. 189. Paris: OECD.

Baldwin, J.R. and Gorecki, P.K. (1987) Plant Creation versus Plant Acquisition. The Entry Process in Canadian Manufacturing. *International Journal of Industrial Organization,* 5,27-41.

Bartelsman, E., Scarpetta, S. and Schivardi, F. (2005) Comparative analysis of firm demographics and survival: evidence from micro-level sources in OECD countries. *Industrial and Corporate Change*, 14:3, 365-391.

Bartelsman, E., Haltiwanger, J. and Scarpetta, S. (2009) Measuring and Analyzing Cross-country Differences in Firm Dynamics. In Producer Dynamics: New Evidence from Micro Data (eds. Dunne, Jensen and Roberts), pp. 15-76. University of Chicago Press.

Benedetto, G., Haltiwanger, J.C., Lane, J. and McKinney, K. (2007) Using Worker Flows to Measure Firm Dynamics. *Journal of Business and Economic Statistics*, 25:3, 299-313.

Boeri, T. and Cramer, U. (1992) Employment growth, incumbents and entrants: Evidence from Germany, *International Journal of Industrial Organization*, 10:4, 545-565.

Butani, S.J., Clayton, R.L., Kapani, V., Spletzer, J., Talan, D.M., and Werking, G.S. (2006) Business Employment Dynamics: Tabulations by Employer Size. *Monthly Labor Review*, February 2006.

Cabral L.M.B. and Mata J. (2003) On the Evolution of the Firm Size Distribution: Facts and Theory. *The American Economic Review*, 93:4, 1075-1090.

Campbell J.R. (1998) Entry, Exit, Embodied Technology, and Business Cycles. *Review of Economic Dynamics*, 1, 371-408.

Caves, R.E. (1998) Industrial Organization and New Findings on the Turnover and Mobility of Firms. *Journal of Economic Literature*, 36:4, 1947-1982.

Davis, S.J., Haltiwanger, J.C. and Schuh, S. (1996a) *Job creation and destruction*. Cambridge/London: MIT Press.

Davis, S.J., Haltiwanger, J.C. and Schuh, S. (1996b) Small business and job creation: Dissecting the myth and reassessing the facts. *Small Business Economics*, 8:4, 297-315.

Dunne, T., Roberts, M.J. and Samuelson, L. (1989a) The Growth and Failure of U.S. Manufacturing Plants. *The Quarterly Journal of Economics*, 104:4, 671-698.

Dunne, T., Roberts, M.J. and Samuelson, L. (1989b) Plant Turnover and Gross Employment Flows in the U.S. Manufacturing Sector. Journal of Labor Economics, 7:1, 48-71.

Eriksson, T. and Kuhn, J.M. (2006) Firm spin-offs in Denmark 1981–2000. Patterns of entry and exit. *International Journal of Industrial Organization*, 24, 1021-1040.

Eurostat / OECD (2007) Eurostat-OECD Manual on Business Demography Statistics. http://www.oecd.org/dataoecd/8/8/39974460.pdf

Eurostat (2003) Business register: Recommendations manual. Luxembourg: Office for Official Publications of the European Communities. http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-BG-03-001-__-N-EN.pdf

Evans, D.S. (1987) The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries. *The Journal of Industrial Economics*, 35:4, 567-581.

Friedman, M. (1992) Do Old Fallacies Ever Die? *Journal of Economic Literature*, 30:4, 2129-2132.

Geurts, K. and Vets, P. (2013) Employee flows to study firm and employment dynamics. KU Leuven: DynaM Working Paper.

Hall, B.H. (1987) The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector. *The Journal of Industrial Economics,* 35:4, 583-606.

Haltiwanger, J., Jarmin, R.S. and Miranda, J. (2013) Who creates jobs? Small versus large versus young. *The Review of Economics and Statistics*, 95:2, 347-361.

Hethey T. & Schmieder J. (2013), Does the use of worker flows improve the analysis of establishment turnover? Evidence from German administrative data. NBER Working Paper 19730.

Jovanovic, B. (1982) Selection and the Evolution of Industry. *Econometrica*, 50:3, 649-670.

Korkeamäki, O. and Kyyrä, T. (2000) Integrated panel of Finnish companies and workers. VATT Discussion Paper 226. Helsinki: Government Institute for Economic Research.

Lotti, F., Santarelli, E. and Vivarelli, M. (2001) The relationship between size and growth: the case of Italian newborn firms. Applied Economics Letters, 8:7, 451-454.

Mata, J., Portugal, P. and Guimaraes, P. (1995) The survival of new plants: Start-up conditions and post-entry evolution. *International Journal of Industrial Organisation*, 13:4, 459-81.

Muendler, M.-A., Rauch, J.E. and Tocoian, O. (2012) Employee spinoffs and other entrants: Stylized facts from Brazil. *International Journal of Industrial Organization*, 30, 447–458.

Mikkelson, G., Unger, L.I. and LeBel D. (2006) Identifying and accounting for mergers and acquisitions in measuring employment. BLS Statistical Survey Papers 2006. Washington: Bureau of Labor Statistics.

Persson, H. (1999) Job flows and worker flows in Sweden 1986-1995. In *Essays on Labour Demand and Career Mobility*, pp. 1-84. Stockholm: Swedish Institute for Social Research.

Rollin A.-M. (2013) Developing a Longitudinal Structure for the National Accounts Longitudinal Microdata File (NALMF). Proceedings of Statistics Canada Symposium 2013.

Spletzer, J.R. (2000) The contribution of establishment births and deaths to employment growth. *Journal of Business and Economic Statistics*, 18:1, 113-126.

Sutton, J. (1997) Gibrat's Legacy. *Journal of Economic Literature*, 35:1, 40-59.

Vilhuber, L. (2008) Adjusting Imperfect Data: Overview and Case Studies. In The Structure of Wages: An International Comparison (eds. Lazear and Shaw), pp. 59-80. University of Chicago Press.

Appendix

To be completed.