Firm creation and post-entry dynamics of de novo entrants*

Karen Geurts  Johannes Van Biesebroeck§

January 2016

Abstract
We show that within the same age cohort, growth rates of young firms are strongly increasing in firm size. This robust empirical pattern is confined to the initial years after entry; growth rates become independent of size as a cohort matures, as predicted by Gibrat’s law of proportionate growth. Both the initial pattern and the subsequent convergence are consistent with the framework of the passive learning model if young firms adjust their size only slowly to new information, for example due to financing or hiring frictions. Importantly, we focus our analysis on firms that enter de novo. We distinguish them from pre-existing firms that merely re-enter an administrative dataset, for example following a restructuring or merger. The extremely narrow size distribution that we observe for de novo entrants provides further support for the passive learning model.

Key words: firm dynamics; passive learning model; growth

JEL: L11; L25; M13

* The authors are grateful to the Belgian National Social Security Office (NSSO), in particular to Peter Vets, for providing the data and for helping to develop the employee flow record linking method; and to Statistics Belgium, in particular to Youri Baeyens and Antonio Fiordaliso, for linking the NSSO data to the Belgian Business Register and for providing firm record linkages created for the Eurostat Structural Business Statistics. Financial support from ERC grant No. 241127 and KU Leuven Program Financing is gratefully acknowledged.

§ Geurts: University of Leuven (KU Leuven), Karen.Geurts@kuleuven.be; Van Biesebroeck : University of Leuven (KU Leuven) and CEPR, Jo.VanBiesebroeck@kuleuven.be.
1 Motivation

New firms entering the economy are generally both numerous and small. As an entry cohort matures, the average firm size increases and the size distribution, being initially highly right-skewed, shifts to the right. Empirical studies have consistently documented how selection leads to a rapid increase in concentration in a given cohort: many young firms fail shortly after entry and firms that expand have a higher probability of survival than firms that stay small (Evans 1987a; Dunne, Roberts and Samuelson 1989; Mata, Portugal and Guimaraes 1995). The passive learning model of Jovanovic (1982) has been widely used to rationalize these post-entry patterns. It assumes that firms enter with an innate productivity they do not know themselves at entry but gradually discover by operating in the market. Firms that learn they are more efficient grow and survive, while the inefficient exit.

Less consensus exists how growth patterns of young surviving firms contribute to the tendency towards increased concentration in a given cohort. Empirical studies typically find that growth rates are very high in the first years after entry and rapidly decrease with age, another regularity in line with model of Jovanovic (Evans 1987a; Haltiwanger, Jarmin and Miranda 2013; Mata and Portugal 2004). But it is unclear whether within a cohort smaller firms grow faster and to some extent catch up in size, or whether larger firms have higher growth rates. The first pattern would imply a negative size-growth relationship and slow down concentration, while the second pattern would accelerate the trend towards increased concentration (Dunne and Hughes, 1994).

Knowing the form of this relationship is important for two reasons. First, theoretical models of firm dynamics often assume or imply a specific relation between growth and size. Second, policy measures to support entrepreneurship and growth often discriminate between firms of different size.

The few studies that have examined the relationship between growth and size of young survivors conditional on age, both measured in terms of employment, report contrasting findings. Evans (1987a), Lotti, Santarelli and Vivarelli (2003) and Mata (1994) find a negative relationship, but Haltiwanger et al. (2013) conclude that there is no systematic relationship between firm size and growth. Furthermore, when using their preferred methodology, Haltiwanger et al. (2013) find that the growth-size relationship within a given age cohort is positive, both for young and older firms. The model of Jovanovic provides little guidance either. In the general version, the relationship is undetermined. Only under specific assumptions does the model predict growth among firms of the same age to be independent of size.

We use data for the universe of Belgian employer firms over a ten-year period and find that the size-growth relationship of young, surviving firms of the same age is strongly and robustly positive. We show, however, that this relationship is confined to the very first years of operation. When entrants mature, the empirical pattern converges to growth rates that are

---

1 The general results presented in Haltiwanger et al. (2013) are shown in greater detail for young firms in Decker, et al. (2014), they confirm the positive relationship.
more or less proportionate to size. This convergence confirms previous studies showing that Gibrat’s law tends to hold for older and larger firms (Mansfield 1962; Hall, 1987; Geroski 1995). A positive size-growth pattern among older firms, as in Haltiwanger et al. (2013), cannot be a steady state as the firm size distribution would become degenerate.

Both the initial deviation and the subsequent convergence are consistent with an augmented passive learning model. The initial pattern can be rationalized within the Jovanovic (1982) framework if one takes into account that young firms adjust their size only gradually to new information and not instantaneously as is assumed in the stylized setting of the model. For example, financing or hiring constraints may prevent young firms from expanding immediately to their desired size. Recent evidence indeed suggests that young firms face more severe financial constraints than older firms (Cabral and Mata, 2003; Beck et al., 2006), and that fast-growing firms experience the greatest constraints to growth (Brown, Earle and Morgulis, 2015). A similar delay before weak performers exit the industry will also reduce the growth rate of small firms and contribute to the observed positive relationship between growth and size. As firms mature and gradually learn their true efficiency, additional information becomes less informative and they converge to their steady state size.

Two measurement problems we explicitly address are worth highlighting as they illustrate the empirical pitfalls in estimating the relationship between growth and size for young firms. First, the estimated pattern is sensitive to accurate identification of truly new firms, which we call de novo entrants, and their post-entry histories in the data. Second, the potential negative bias in the relationship, induced by sample selection and regression-to-the-mean, is exacerbated in a sample that consists of very small firms. We now discuss these two aspects in more detail.

Firm-level administrative data are currently the main source for empirical analysis on firm dynamics. The identification of individual firm histories in the data is, however, hampered by the fact that some firms may change ID code or restructure. It is widely recognized that such events lead to spurious measurements of entry and exit, and to overestimations of firm and employment dynamics (Haltiwanger et al., 2013; Geurts, 2016). The bias they introduce in post-entry growth patterns has, however, received less attention. We make use of two state of the art record linking methods to solve these problems. They enable us to trace the complete histories of de novo entrants, from the moment they start operating till they cease activities, i.e. true economic exit. We show that failing to identify even a small amount of spurious entrants has major implications for the estimation of post-entry growth patterns.

Our exclusive focus on de novo entrants reveals that the firm size distribution at entry is confined to a much narrower range of small size classes than found in many previous studies. This empirical observation is very much in line with the passive learning model which predicts that firms, lacking prior information about their efficiency, all enter at the same size. Studies that cover a sample with a broad range of firm sizes already at entry must be investigating growth in a different population than de novo entrants.

---

2 Abbring and Campbell (2005) show that many poorly performing firms stick around while making losses as they are committed to a year’s lease on their premises.
It is well-known that two statistical problems may bias the relationship between size and growth for surviving firms. Regression-to-the-mean as well as sample selection may spuriously induce a negative relationship if firm size is measured in the base year, i.e. at the start of the period over which growth rates are calculated (Hall, 1987). Although these problems may be less important for larger firms, the statistical side-effects of the base-year size classification are greatly exacerbated in a sample of small firms, as is the case in our sample of de novo entrants. We therefore need to directly address these measurement issues. To avoid bias in the size-growth relationship, we use three alternative firm-size classifications that approximate a continuous size-growth relationship. We find a robust positive size-growth relationship for each of the alternatives.

The remainder of the paper is organized as follows. Section 2 starts with a brief overview of stylized facts on entry and post-entry dynamics. It also reviews predictions of the model of Jovanovic (1982) and discusses how they are affected by delayed adjustment. Section 3 presents the dataset and our strategy to identify de novo entrants and their post-entry histories. In Section 4, the empirical model and the size measurement issues are discussed. The results are presented in Section 5, first showing that some well-established facts about firm entry and exit are replicated in the Belgian dataset, and then showing post-entry growth patterns by age and size. Section 6 concludes.

2 Facts and theory

2.1 Some stylized facts

Empirical studies for various countries have found entry rates of new firms in manufacturing and services to vary between 5 and 15 percent per year. Most entrants tend to be much smaller than the average incumbent, such that the employment share of new entrants is generally far less than 5 percent of the workforce (Siegfried and Evans 1994; Geroski, 1995; Caves, 1998). As a cohort matures, average firm size increases and the number of firms falls. This tendency towards increased concentration in a given age cohort is very strong in the first years after entry. A typical pattern is that 5 to 10 years after entry, average firm size has doubled, but only half of an entry cohort survives.3 Cabral and Mata (2003) showed firm size to be highly right-skewed at entry and shift towards a more symmetric distribution over time. The long-run cohort’s size distribution remains, however, right-skewed, without convergence to a common size (Konings, 1995).

The rapid increase in concentration among an entry cohort is explained by specific post-entry dynamics showing systematic differences between young firms and incumbents. A first difference is a selection process that reduces the number of smaller firms in a cohort. Many empirical studies have shown that young firms exhibit high failure rates immediately after

---

3 See for example Dunne, Roberts and Samuelson (1988) for the U.S., Wagner (1994) and Boeri and Cramer (1992) for Germany; Mata et al. (1995) for Portugal.
entry. Two patterns are highly robust: (i) exit rates are decreasing in firm size and (ii) survival rates increase as firms mature.4

Another well-established fact is that young surviving firms exhibit remarkably high growth rates which decline with age.5 Variation in growth rates among surviving firms can contribute to increased concentration if larger (young) firms tend to grow faster than smaller ones, or if growth rates exhibit positive serial correlation (Dunne and Hughes, 1994). The existing evidence on which pattern prevails in the early post-entry process has been inconclusive.

Several studies lump all firms below a certain age in one cohort and verify whether growth rates conditional on survival increase or decrease with firm size among these young firms. Dunne et al. (1989) and Almus and Nerlinger (2000) find, for the manufacturing sectors of the U.S. and Germany, that smaller plants or firms grow faster than larger ones. Wagner (1994) also studies German manufacturing firms, but finds growth rates to be independent of size.6 As these patterns include an age effect within the broader cohort—and we know that younger firms tend to be smaller and growing faster—they provide imperfect evidence on the size-growth relationship among firms of the same age.7

The few studies that have investigated post-entry growth conditional on age obtain contrasting results. Evans (1987a) and Lotti et al. (2003) report an inverse relationship between growth and size given age for surviving young firms in the first six years after entry. They find this pattern diminishes with age and converges towards growth that is proportionate with size, consistent with evidence that suggests Gibrat’s law holds in a sample of older firms or among firms that have exhausted scale economies (Mansfield 1962; Hall, 1987; Geroski 1995). Mata (1994) finds a weak negative relationship that is insignificant at the 1 percent level.8 Haltiwanger et al. (2013) report a negative as well as a positive pattern, depending on the size classification method. When using their preferred methodology, they find larger firms to grow more rapidly than smaller ones among young survivors of the same age. Moreover, their results show no convergence towards proportionate growth for older firms. The contrasting results may be partly explained by differences in measurement methods and industry scope, as we discuss in more detail below.9 We will also show that an accurate identification de novo entrants matters greatly.

---

4 See for example Evans (1987a) and Dunne et al. (1989) for U.S. manufacturing plants, Haltiwanger et al. (2012) for U.S. manufacturing and services; Mata et al. (1995) for Portugal.

5 See the same studies for the U.S.; Mata and Portugal (2004) for Portugal.

6 The three studies group together all firms younger than, respectively, 5, 6, or 10 years. Audretsch, Santarelli and Vivarelli (1999) also investigate the relation between growth and size of young Italian manufacturing firms, but they estimate growth rates relative to size at entry.

7 A discussion of this age composition effect is provided in Section 5.3.

8 Pooling young firms up to age 4 into one age class, Mata (1994) finds a stronger negative relationship. As noted before, this result is likely to reflect an age composition effect of small, fast-growing firms being younger.

9 Evans (1987a) and Haltiwanger et al. (2013) report results for U.S. firms. Lotti et al. (2003) cover Italian firms and Mata (1994) Portuguese firms. Haltiwanger et al. (2013) classify firms by average size in t-1 and t and include both manufacturing and services, while the other studies use a base-year size classification and are limited to manufacturing firms.
Note that some studies have used firms as their unit of analysis while others used plants or establishments. In our analysis, we are not interested in country comparisons of performance, but rather try to uncover general patterns of firm behavior. The unit of analysis most closely related to the theoretical notion of new firm creation is the firm and that is the unit of observation we will work with. As the vast amount of new entrants tend to have only a single plant or establishment, this definition covers a subset of the entrants that plant-level studies would identify.

2.2 Theoretical framework

**How do firms enter?**

The passive learning model of Jovanovic (1982) implies a particular process of firm dynamics by age and size and has often been used to rationalize exit and growth patterns of entrants. The key assumption is that firms enter without knowing their own innate productivity. Prior to entry, they receive, but do not observe, a random draw from the productivity distribution in the industry. Since entrants know the population distribution, they have the same prior beliefs and all enter at the same size. Each period they update their prior distribution over their own productivity level using Bayes’ law as evidence on profitability is realized. Firm sizes diverge as the cohort matures even though the underlying firm-specific productivity level is constant.

This modeling approach contrasts with Lucas (1978) which features a dispersion of managerial skill in the population. High-skill individuals self-select into entrepreneurship, rather than becoming an employee, and they choose their firm size optimally upon entry. It also contrasts with the model of Hopenhayn (1992) where firms similarly receive a random draw from a known productivity distribution, but they observe this realization after paying a fixed entry cost and before hiring any production factors. If they enter, they immediately do so at the “right” size.

The first implication of the Jovanovic model rarely holds in large-scale datasets used to investigate firm dynamics. Firms are predicted to all enter at the same or similar scale, while actual entrants typically span a broad range of firm sizes. Deviations might simply be due to the stylized assumptions of the model, but two measurement issues help explain the discrepancy between the prediction and stylized facts. First, variation in entry size can reflect choices between the moment a firm is established and the first time it is observed in the dataset. In our administrative database of Belgian employers, new firms are observed in the first year they record positive employment on June 30. On that day, some firms have already been in existence, either without employees for an unknown period, or with employees for up to 12 months. They have had the chance to learn about their innate productivity and choose different growth rates, or even exit, in response. The observed entry size distribution should thus (at least partly) be regarded as the outcome of an initial selection and size adjustment process.

---

10 Models of entrepreneurial entry with financing constraints, such as Evans and Jovanovic (1989) and Cabral and Mata (2003), also predict that the size distribution of entrants will cover a narrow range.
Second, and more importantly, the group of entrants in administrative datasets typically includes some pre-existing firms that re-enter the dataset after a legal or ownership restructuring or enter with a new subsidiary. Examples include divestitures, control changes, legal restructuring for tax or liability reason, incumbents entering a new industry or starting activities in a new region, etc. These other modes of entry are certainly economically relevant, but we do not expect post-entry dynamics of these firms to conform to the predictions of the passive learning model. We label them as spurious entrants, as opposed to de novo entrants which we study in this paper.\(^{11}\) Several studies have demonstrated that entry by established firms fundamentally differs from de novo entry (Dunne et al. 1988; Baldwin and Gorecki 1987; Konings et al., 1996; Bilsen and Konings, 1998; Mata and Portugal 2004). These firms already have a better idea of their own productivity. They tend to enter with a larger size, are less likely to fail, and exhibit less dynamic growth patterns. They are an interesting group of firms to study, as these changes could very well be systematically related to past or future performance, but here we choose to focus on de novo entrants.

**How do firms grow after entry?**

Many heterogeneous firm models do not incorporate firm-specific stochastic elements that give rise to systematic heterogeneity in growth rates. In the model of Hopenhayn (1992), firms enter immediately at their optimal size and later adjustments in firm size are responses to random productivity shocks firms have no control over. Abbring and Campbell (2004) add persistence in post-entry shocks to the model which leads to serial correlation in growth rates and eventually to a positive size-growth relationship.

The passive learning model of Jovanovic (1982) is one exception.\(^{12}\) Firms only discover their own innate efficiency level from operating in the market. Initially, they have the same beliefs about this and they all enter at the same size. Realized profits depend on their actual underlying efficiency and idiosyncratic cost shocks and they use Bayes’ rule to update their beliefs and expand or contract into their correct size. Firms that discover they are more efficient, grow and survive, while the inefficient shrink and exit. As time passes, firm sizes within an entering cohort diverge and become strictly increasing in firms’ estimate of their own efficiency. As firms mature and gradually learn their true efficiency, additional information becomes less informative and they converge to their steady state size.\(^{13}\)

This model generates several testable predictions about exit and growth patterns in relation to the firm’s age and size. First, the noisy selection process implies an inverse relationship

\(^{11}\) As shown in Section 3., the vast majority of spurious entrants we distinguish from de novo entrants are simply incumbents that continue the same activities with a new identification code after an administrative or legal change.

\(^{12}\) The active learning model of Ericson and Pakes (1995) is another exception. In their model, growth is a function of firms’ actions as they can make investments to raise productivity. As the link between investment and productivity is stochastic, even identical investments can generate different outcomes.

\(^{13}\) Further growth is driven solely by business cycle shocks affecting all firms similarly. In the model of Hopenhayn (1992), even mature firms experience random productivity shocks that induce random growth rates in steady state, but these are unrelated to firm size.
between exit and size given age and between exit and age. Unsuccessful firms stay small, they might even contract, and eventually choose to exit. Larger firms are those that received favorable cost information in previous periods and have expanded. While initial profit realizations provide new entrants with a lot of information on their ability, subsequent information becomes gradually less informative and is less likely to induce exit.

Second, the model implies that conditional on survival younger firms have higher and more variable growth rates than older firms. They are still highly uncertain about their own quality and respond to market success by expanding. As the weakest firms exit, average efficiency among surviving firms improves from period to period which is reflected in higher average firm sizes. As firms mature, revisions of estimated efficiency become smaller. Firms eventually approach their optimal scale and the variance of growth rates converges to zero.

Third, because smaller firms are on average younger, the model also predicts an inverse relationship between growth rates and size in a cross-section of firms that encompasses a range of cohorts. Several empirical studies find evidence for this inverse relationship and Jovanovic (1982) cites it as a key motivation for the model. However, without additional assumptions, the model does not imply any systematic relationship between growth rates and size conditional on age. Assuming a Cobb-Douglas cost function leads to a prediction that growth rates are independent of firm size among firms of the same age cohort, consistent with Gibrat’s law. At each point, a firm’s size reflects its best estimate of its efficiency. With this cost assumption, adjustment is complete and subsequent adjustments depend only on future information which is by definition random.

In the stylized framework of the Jovanovic model, a firm’s current size only reflects its past growth history. The model assumes instantaneous adjustment to new information, but in reality, frictions might distort this process. Hsieh and Klenow (2009) show for several countries that deviations between factor prices and marginal productivities and between observed and optimal output levels are widespread. As these deviations partially reflect the dynamic adjustment of quasi-fixed production factors (Asker, Collard-Wexler and De Loecker 2014), it is likely that younger, less established firms face greater external frictions. For the prediction of the size-growth relationship conditional on age, it matters greatly whether they already affect firm size at the moment of entry or whether they mainly influence adjustments in firm size following entry.

A prominent example of the first situation is the model of Evans and Jovanovic (1989) where entrepreneurs face liquidity constraints. Heterogeneity among firm size at startup reflects that the smallest entrants faced the strongest financial constraint. They need to finance their expansion from realized profits. If the friction is not perfectly correlated with ability, they will also have the highest growth potential and we should observe a negative relation between initial size and subsequent growth, as in Audretsch et al. (1999). An alternative mechanism that generates the same prediction is developed by Cabral (1995). If production capacity requires substantial sunk costs that are foregone when firms exit, smaller

14 Dunne et al. (1989) argues that efficiency levels, and thus firm sizes, are bounded from above. This leads to a negative relationship as there is less room for further increases for larger firms.

8
firms are more likely to exit and they will choose to invest gradually and enter at even smaller scale.

In the second situation, entry size is not distorted by frictions. Yet following entry, some firms cannot immediately adjust to their desired size when they revise their estimate of their innate efficiency. Credit, hiring, or regulatory constraints can limit growth in the first years. For some expanding firms, current size will be below desired size and they will need several years to incorporate positive information into their size. For some years, their size and growth rate both reflect underlying firm quality. Until adjustment is complete and desired size catches up with actual size, it leads to higher growth rates for larger firms. Delayed adjustment of firm size introduces a positive correlation between past and current growth, and thus between firm size and growth.\(^{15}\)

Delayed adjustment can have many reasons. It can be externally imposed, for example credit constrained firms may need to finance investments from retained earnings. A vast literature documents the excessive sensitivity of many firms’ investments to free cash flow (Fazzari, Hubbard and Petersen 1988; Evans and Leighton 1989). Cabral and Mata (2003) and Beck et al. (2006) find that young firms face more severe financial constraints than older firms, while Brown, Earle and Morgulis (2015) show that fast-growing firms experience the greatest constraints to growth. Search frictions to hire specialized staff in thin labor markets or zoning regulations are other external frictions that can delay adjustment to positive shocks. Risk aversion will exacerbate the pattern of gradual adjustment. While larger firms might be risk-neutral, individual entrepreneurs are likely to be somewhat risk averse (Brockhaus, 1980). Especially in the face of irreversible investments and sunk costs, firms will not incorporate all positive information immediately in their size. Past growth will result in a somewhat higher size, but also be followed by future growth.

Delayed exit further contributes to a positive relationship between growth and size. The option value associated with the sunk entry costs may provide an incentive for some loss-making firms to continue operations before eventually deciding to withdraw from the market. In many administrative firm-level datasets it is common to observe firms with no employment and no or minimal sales for several years. If fixed costs are low relative to sunk entry costs, small firms might simply hang around for the business cycle to improve rather than exit.

3 Data

The analysis is based on the register of Belgian employers maintained by the National Social Security Office (NSSO). It includes all private firms with at least one employee and covers the period from 2003 to 2012. In an average year, the sample includes 178,000 firms and 2,070,000 employees.

We identify de novo entrants in the data by making use of two state-of-the-art record linking methods. The methods are further used to trace the complete histories of firms from

\(^{15}\) In a Markov Perfect equilibrium, the value of current state variables are sufficient statistics for the entire firm history (Ericsson and Pakes 1995). With adjustment frictions this is not necessarily the case anymore.
the moment they start operating till they cease activities, i.e. true economic exit. For those firms that change ID code or restructure, we impute employment measures up to the sixth year of existence. To our knowledge, we are the first to use this approach to obtain consistent post-entry firm histories. The details of our methodological approach are explained in Appendix A. Below, we provide a summary and show that the size range of de novo entrants dramatically differs from the size range at entry suggested by the raw dataset. This has major implications for the post-entry size-growth relationship.

It is widely recognized that administrative firm-level data suffer from missing links in individual firm histories, which hinders the straightforward identification of firm dynamics. Firms may change ID code due to mergers, takeovers, split-offs, ownership changes or for tax optimization purposes. These events generate various biases in empirical measures, such as spurious measurements of entry and exit, misclassifications of firm growth across age and size classes, and overestimations of job and firm turnover (Haltiwanger et al. 2013; Geurts, 2016). To solve these problems, we use two record linking methods cumulating the linkages we identify.

The first consists of a set of traditional record linking techniques developed by Statistics Belgium in line with the OECD-Eurostat recommendations on constructing longitudinal business data (Eurostat-OECD 2007). The method relies on probability-based matching and the use of supplementary data sources with information on firm continuity. The second linking method is based on an employee-flow approach. It follows one of the key production factors of the firm, the stock of employees, to identify changes in ID codes and firm structure. Continuity of the firm’s workforce is thus used to identify firms that operate continuously.

The established linkages are first used to identify continuing firms that are misclassified as exits and entrants in consecutive years. They are labeled as ‘spurious’ exits and entrants as opposed to true exits and de novo entrants. It is especially important to recognize that spurious entrants are pre-existing firms that are likely to exhibit characteristics similar to other incumbents. If they are mixed up with de novo entrants, the typical size and growth patterns of young firms will be biased towards those of incumbents. Panel (b) of Table A.1 in the Appendix shows that 78 percent of the spurious entrants we identify are simply incumbents that continue the same activities with a new identification code after a purely administrative or legal change. Another 18 percent are split-offs of another firm.

Next, for de novo entrants that are involved in an ID change or restructuring in the years following entry, employment is imputed up to the sixth year after entry. For one-to-one ID changes, which represent the vast majority of events, this simply means replacing the new by the old ID code. For more complex events, we assume the same growth rate for each firm involved in the event. This imputation method preserves the firm size distribution in the sample and allows a more accurate estimate of post-entry employment patterns by size.

Table 1 shows that the two linkage methods are complementary for the accurate identification of de novo entry across different size classes of firms. The first row reports the average annual number of entrants as observed in the raw administrative data. The next rows present the fraction of these firms that are identified as either de novo or spurious entrants.
Table 1. Share of *de novo* and spurious entrants in all administratively recorded entrants

<table>
<thead>
<tr>
<th>Number of firms</th>
<th>Total</th>
<th>1-4</th>
<th>5-9</th>
<th>10-19</th>
<th>20-49</th>
<th>50-49</th>
<th>100+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of firms</td>
<td>17,283</td>
<td>15,368</td>
<td>1,209</td>
<td>446</td>
<td>190</td>
<td>39</td>
<td>32</td>
</tr>
<tr>
<td>Share of de novo entrants</td>
<td>0.91</td>
<td>0.95</td>
<td>0.64</td>
<td>0.41</td>
<td>0.26</td>
<td>0.17</td>
<td>0.03</td>
</tr>
<tr>
<td>Share of spurious entrants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (both methods combined)</td>
<td>0.09</td>
<td>0.05</td>
<td>0.36</td>
<td>0.59</td>
<td>0.74</td>
<td>0.83</td>
<td>0.97</td>
</tr>
<tr>
<td>1. Identified by traditional method</td>
<td>0.06</td>
<td>0.05</td>
<td>0.12</td>
<td>0.16</td>
<td>0.21</td>
<td>0.32</td>
<td>0.44</td>
</tr>
<tr>
<td>2. Identified by employee-flow method</td>
<td>0.05</td>
<td>-</td>
<td>0.32</td>
<td>0.57</td>
<td>0.72</td>
<td>0.82</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Note: Average of annual shares over the 2003-2012 period. Firm size classes are based on employment.

Spurious entrants only represent 9 percent of the total, but this low fraction does not mean it is an unimportant group. The probability that a new ID code corresponds to spurious entry increases dramatically with size. They account for more than one third of administrative entrants with 5 to 9 employees and even two thirds of those with 10 or more employees. *De novo* entrants with more than 50 employees are extremely rare. As a result, the size-distribution of *de novo* entrants is more strongly right-skewed than in the unedited data and the presence of spurious entrants would introduce a bias in post-entry patterns by size. Table 1 further shows the complementarity of the two linkage methods. The traditional method is needed especially in the size class below five employees, where employee-flow links are absent by construction. Yet the employee-flow method is essential in larger size classes, where it identifies two to three times more spurious entrants than the traditional method.

4 Empirical model

We characterize survival and growth patterns for young firms by age and size using the employment history of *de novo* entrants up to the moment of true economic exit. As shown in Dunne et al. (1989), the mean growth rate of a class of firms can be decomposed into the growth rate of survivors weighted by the probability of survival, minus the probability of exit. The two equations, using the firm-level growth rate and the exit dummy as dependent variables, are estimated separately.

Employment is measured as the number of employees registered on June 30. The set of entrants in year \( t \) includes all firms that started as an employer after June 30 of year \( t-1 \) and survive until June 30 of year \( t \). It conditions on surviving a first selection process, from a firm’s establishment, the unknown point in time of age 0, to the first recorded instance of positive employment, denoted as age 1. Exits in observation period \( t-1 \) to \( t \) are firms for which \( t-1 \) is the last year of positive employment. Firms that change ID code or firm structure are not considered as exits. Their growth path following the event is based on imputed employment. The years between entry and exit, firms are denoted as survivors.\(^{16}\)

\(^{16}\) Some survivors have zero employment in a given year (‘dormant’ firms). They are treated as outliers and omitted from the regressions in the periods concerned.
Following Davis, Haltiwanger and Schuh (1996a), firm-level growth rates are calculated as discrete-time employment changes relative to the average of employment in year \(t-1\) and year \(t\). Denoting employment of firm \(i\) in year \(t\) as \(E_{it}\), the growth rate over the preceding year equals \(g_{it} = (E_{it} - E_{it-1})/E_{it}\), with \(E_{it} = (E_{it} + E_{it-1})/2\). These growth rates range from -2 for exits to +2 for entrants, show job creation and destruction symmetrically and are bounded away from infinity.\(^{17}\) Regressions use employment weights such that the coefficient estimates are readily interpreted as aggregate employment changes for a class of firms. Specifically, the mean estimated growth rate represents the rate of net employment creation in a given age-size class of firms, and the exit rate represents the job destruction rate.

At each age, firms are grouped into six size classes, based on the number of employees and defined on a logarithmic scale: \([0,2]\), \([2,4]\), \([4,8]\), \([8,16]\), \([16,32]\), and \([32,\infty]\).\(^{18}\) All observations with more than 32 employees are in the same size class because few \emph{de novo} entrants reach this size within the first five years of existence. Exits are assigned to the size class of employment in their last year.

To document patterns of firm dynamics, we regress the dependent variables on age and size classes using a saturated dummy regression model. It includes separate indicators for all possible values taken by the two discrete explanatory variables and their interactions. This approach follows Haltiwanger et al. (2013) and has two advantages over other estimation methods used to examine the relationship between growth and size. First, as emphasized by Angrist and Pischke (2009), a saturated regression model fits the conditional expectation function perfectly, regardless of the distribution of the dependent variable. Moreover, no particular shape of the size-growth relationship has to be imposed. Second, the estimates are robust to heteroscedasticity, a recurrent problem in empirical studies of the size-growth relationship.\(^{19}\)

For each of the two dependent variables, \(y_{it} = \{g_{it}, e_{it}\}\), firm-level employment growth and the exit dummy, the following regression model is estimated:

\[
y_{it} = \sum_{j=2}^{6} \sum_{k=1}^{6} (\alpha_{jk} + \beta_{jk} D_{it}^d) 1[age_{it} = j] 1[size_{it} = k] + \sum_d \gamma_d D_{it}^d + \gamma_t + \varepsilon_{it}
\]

\(^{17}\) This growth rate is close to the more commonly used logarithmic growth rate \(g_{it} = \ln(E_{it}/E_{it-1})\), especially for values between -1 and +1. Both measures show expansion and contraction symmetrically, whereas the growth rate relative to base-year employment \(t-1\) ranges from -1 to infinity. Symmetry is a crucial feature for estimating mean growth rates of young firms, as their employment fluctuates widely. A further advantage of our growth rate is that using the corresponding employment weights, \(\bar{E}_{it}\), in the regressions yields coefficient estimates that exactly represent net employment growth of a class of firms. Equivalent weights do not exist for the logarithmic growth rate. In the exit regressions we use \(E_{it-1}\) as employment weights.

\(^{18}\) Due to the use of average employment and imputed employment levels, size is a continuous variable.

\(^{19}\) For a further discussion of the econometric problems see Hall (1987), Evans (1987b), and Dunne et al. (1989). Since we examine how growth rates of survivors depend on the current size of the firm, where both growth and size are updated at each age, we also avoid the sample censoring bias many previous studies had to address (Mansfield, 1962).
where the dummy variable $1[\text{age}_{it} = j]$ takes a value of one if the age of firm $i$ in year $t$ equals $j$ and similarly for the size category dummies. The six industry dummies $D^A_i$ enter both additively and interacted with the full set of age-size interactions. As we impose that $\Sigma_t \beta_{ijk} = 0$, the average effect of age and size on growth and exit is captured by the uninteracted $\alpha_{jk}$ coefficients, while the $\beta_{ijk}$ coefficients allow for industry heterogeneity. The additive year dummies control for business cycle effects.

**Size classification of surviving firms**

We approximate a continuous size-growth relationship using three alternative approaches to allocate surviving firms in a size category. The objective is to mitigate two statistical side-effects of a conventional base-year classification, which classifies firms by size in $t-1$. First, as discussed extensively in the literature, regression-to-the-mean may spuriously induce a negative relationship between size and growth if firm size is measured at the start of the period over which growth rates are calculated. Even if employment growth is independent of size, random variation due to measurement error or transitory fluctuations will systematically bias growth estimates upwards for firms that are small in $t-1$ (Hall 1987; Friedman 1992; Davis et al. 1996b). Second, employment in the subset of surviving firms is bounded from below by one. Therefore, the lower tail of possible rates of decline is truncated, while the upper tail of growth rates is unaffected. It especially affects smaller firms which will already exit when hit with a moderate negative shock and leads to sample selection bias. It again induces an inverse relation between size and growth if size is determined at the start of the period (Mata 1994; Baldwin and Picot 1995).

Hall (1987) and others have found that these problems have little effect on the size-growth relationship for larger, more established firms. However, they are exacerbated in a population of predominantly small firms, as in our sample of de novo entrants. Single employee firms that survive cannot even have a negative growth rate. Dunne et al. (1989) and Mata (1994) largely circumvent these statistical problems by excluding the smallest firms from their sample. This is not an option for us, given our focus on de novo entrants which are predominantly observed in size classes below 5 employees.\footnote{Among de novo entrants, 94 percent of firms have fewer than 5 employees at age 2 and 82 percent at age 6.} Instead, we use three alternatives to allocate firms in a given size class. The intention is to approximate firm growth in continuous time and we refer to the ‘current’ size of the firm. A more detailed discussion of these methodologies is in Appendix B; here we provide a brief overview.

The first size classification method, and the one we use for our benchmark estimates, allocates employment gains and losses to each of the size classes that the firm passes through as it grows or contracts (Butani et al. 2006). In this ‘dynamic’ size classification, firms are initially assigned to a size class based on employment in $t-1$, but are re-assigned to a new class when they cross a threshold. The growth from $E_{it-1}$ to the threshold is assigned to the initial class and the remaining growth from the threshold to $E_{it}$ is assigned to the next size class. This methodology approximates instantaneous class re-assignment that would be feasible if size and growth were measured in continuous time. As it attributes symmetric
employment changes to the same size classes, it avoids the negative as well as the positive bias in the size-growth relationship that afflict other methodologies.

The second classification method uses each firm twice in the regression, assigning a weight of one half to each observation. One observation uses the firm’s employment level at the beginning of the period—both as a base for the growth rate and to determine the size class. The second observation uses the firm’s employment at the end of the period again for both calculations. This approach was proposed by Prais (1958) to avoid regression-to-the-mean bias and can be motivated similarly as the use of average wage shares in a Solow residual, i.e. as a discrete approximation to the continuous Divisia index of productivity growth (Caves, Christensen and Dievert, 1982).

A last classification method follows Davis et al. (1996a, 1996b) and uses the average of firm size in years t-1 and t as a proxy for the size over the intervening period. It is adopted for comparison with the results reported by Haltiwanger et al. (2013). Baldwin and Picot (1995), however, indicate that this size classification introduces an upward bias between size and growth if there is positive trend growth rate in the population.  

5 Results

In constructing the dataset, we have taken great care to only identify firms as de novo entrants when they start new operations, corresponding to firm creation in Jovanovic (1982). With continuing firms misclassified as entrants or exits filtered out, we find two novel patterns. In particular, we show that de novo entry is confined to a much narrower range of small size classes than usually found and that growth rates for surviving entrants are increasing with firm size. We discuss the two novel results in detail below, but first summarize a few patterns for de novo employer entrants in the Belgian private sector that are consistent with the empirical evidence from other countries, as discussed in Section 2. They suggest that the novel findings are not an artifact of the Belgian dataset. A brief summary of the confirmed patterns is provided below, while Appendix C contains a more detailed discussion.

5.1 Confirmed patterns

In line with results for many other countries, statistics in Table A.2 show that the annual rate of firm entry in Belgium is high (9%), but involves only a small fraction of employment (1.5%). Most entrants are extremely small; average entry size is 1.9 employees, six times smaller than the average size of incumbents. In the years following entry, a large fraction of the entering cohort exits and the average size among survivors increases. Only half of all entrants survive to age 6, at which time the average firm size in the surviving group has almost doubled.

---

21 The weights in the growth regressions follow naturally from the three size classification approaches. They always equal the employment used in the denominator of the growth rate calculation: (i) the truncated average employment within the size class, (ii) $E_{it-1}$ or $E_{it}$, and (iii) $E_{it}$. 
A first mechanism generating this pattern of increased concentration within an entry cohort is selective survival. In line with previous evidence we find high exit rates for young firms which are decreasing in age as well as in size, see panel (a) of Figure A.1. Our results suggest that the selection process of the passive learning model—which predicts market exit of the least efficient and therefore the smallest firms—unfolds quickly in the first years after entry. By age 6, exit rates have approximately halved. A second prediction of the passive learning model is also borne out in the Belgian data. Surviving young firms exhibit high growth rates in the early years after entry, but growth slows down rapidly with age. The average growth rate declines convexly as it converges to a constant steady state – panel (b) of Figure A.1.

As young firms have much higher growth rates and are overrepresented in smaller size classes, the changing composition of the sample leads to a negative relationship between growth and size in a cross-section of firms if we pool all ages. Such a relationship has often been documented in the literature and it is also what we find for Belgium, as shown by the ‘all firms’ line in Figure 1. Growth rates among all firms that survive from year \( t-1 \) to \( t \) decline monotonically with the current size of the firm. It is instructive, however, to separately consider the size-growth relationship for young firms of at most six years old, and that of older firms. The dashed line at the bottom of Figure 1 shows low growth rates for incumbents regardless of firm size. For them, absolute employment growth is proportional to the current size of the firm, confirming Gibrat’s law in our data set. In contrast, growth rates for young firms are not only higher, they clearly increase with size.

**Figure 1. Growth rates of surviving firms by size: young firms versus incumbents**

![Growth rates of surviving firms by size: young firms versus incumbents](image)

Note: Annual averages over the 2003-2012 period. We use the dynamic size classification as benchmark method to construct the X-axis. For young firms the 32-63 size class is really 32+, but very few de novo entrants have more than 63 employees (shown below).
The patterns described so far are in line with results from other studies based on large-scale firm-level datasets, even when no or little attempt has been made to distinguish between what we have labelled *de novo* and spurious entrants. It suggests that most patterns are fairly robust to less accurate identification of truly new and young firms. When calculated using the raw administrative data, we indeed find almost the same results for incumbents and all firms as in Figure 1.22 The positive relationship between growth and size that we observe among young *de novo* firms, however, is not replicated in the raw sample of administrative entrants. Instead, the light gray line in Figure 1 for the unadjusted administrative data suggests that among young firms, small firms have higher growth rates than larger ones. In Section 5.3 below, we show that the difference between *de novo* and administrative entrants is even more pronounced when growth rates are estimated conditional on age, and how spurious entry biases the estimated relationship.

As discussed before, the passive learning model of Jovanovic (1982) has no prediction on the size-growth relationship for young firms. Only with some functional form restrictions does it predict growth to be independent of firm size for all age cohorts. Whatever form the relationship takes, as long as small size classes have relatively more young firms and surviving young firms have higher growth rates—two confirmed predictions of the model—the size-growth relationship is guaranteed to be a negative in the full population of firms. At least if the composition effect is strong enough to overturn the positive relationship for young firms. This is certainly the case in our sample of *de novo* entrants, a finding we turn to first in the next section.

### 5.2 Entry distribution

Although summary statistics based on all administrative entrants or limited to the set of *de novo* entrants look very much alike, a closer examination of both samples reveals some fundamental differences. This is because spurious entrants—pre-existing firms that underwent some reorganization and are misclassified as entrants—introduce incumbent-like features into the population of administrative entrants. As a small group they have little impact on average statistics, but they strongly affect the entry distribution by size or the size-growth pattern, especially if we use weights to reflect the aggregate employment evolution.

The importance of identifying entrants correctly is readily seen from the employment distribution at entry by firm size. Figure 2 shows average annual employment divided into seven size classes on a logarithmic scale. The left panel shows the employment distribution of *de novo* entrants (dark) against that of all administrative entrants (light). It is well-known that new firms predominantly enter in the smaller size classes, but the distribution based on the administrative sample greatly understates this pattern. Employment of *de novo* entrants is almost entirely concentrated in the first three size categories, which account for fully 82% of total job creation of new start-ups. Firms entering with at least 32 employees are exceedingly rare and account for less than 5% of total job creation.

---

22 The results calculated using the raw administrative data are reported in Geurts and Van Biesebroeck (2014).
Figure 2. Employment distribution of entrants

The distribution of spurious entrants—the difference between the two series in the left graph—mirrors this pattern. It is mainly concentrated in the larger size classes. The right panel shows the employment distribution of spurious entrants (dark) relative to that of incumbents (light). The cumulative employment share of the first three size classes is only 13% for spurious entrants, while firms with at least 32 workers employ 58% of the group’s total. The employment distribution of spurious entrants is remarkably similar to that of incumbents. It confirms that spurious entrants are a subset of older firms and suggests that their incidence is unrelated to firm size.

As we have shown, the sample of administrative entrants that uses untreated firm-data mixes two distinct populations of firms. Failing to distinguish between them, as is generally not done, has two implications. First, the size distribution of entrants has a much more dispersed shape than the strong right-skew we observe for de novo start-ups. Second, given that employment by spurious entrants accounts for 44% of the total in the sample of administrative entrants, it gives an inflated impression of the importance of new firms for job creation in official statistics. In an average year, new job creation by all de novo entrants only represents 1.5% of the Belgian private-sector workforce. Using administrative entrants instead would suggest this fraction is 2.7%, 1.8 times higher.

Besides eliminating false entrants with incumbent-like characteristics, our focus on de novo entrants has another important implication. It shrinks the firm sizes that we observe for entrants to a very narrow range. Note that the bottom five size classes, which capture almost all employment of new entrants, are all firms with fewer than 32 employees. This empirical observation is very much in line with the passive learning model, where entrants—having no prior knowledge about their own efficiency—are assumed to all enter at the same size. This
is approximately what we observe, and contrasts with the much wider range observed in most previous studies.

The limited size differences we do observe among de novo entrants are plausibly the result of selection and growth effects occurring between a firm’s startup and the first time we observe it, as new firms only enter the dataset on June 30. Alternatively, they can reflect some prior knowledge that entrants have about their own intrinsic quality even before they enter the market. The narrow range of observed sizes then implies that a lot is still unknown to these firms when they enter.

It can be expected that spurious entrants also exhibit incumbent-like dynamics following entry and that their overrepresentation in large size classes creates a bias in the size-growth and size-exit pattern for entrants. The bias is hardly noticeable in exit probabilities by size, see Figure A.2 in the Appendix, since exit rates are decreasing in size both for young and older firms. The bias is, however, large in growth estimates by size, where young and older firms strongly differ. This is the topic we turn to next.

5.3 Post-entry growth

Most previous studies that empirically examined the relationship between growth and size of young firms have taken for granted firm entry, exit and growth as observed in the data, or applied only a rough correction for spurious entry and exit. It is thus unlikely that reported empirical patterns refer to a well-defined set of truly young firms. Including spurious entrants does not markedly affect many entry and post-entry patterns, as illustrated above. It does, however, bias the size-growth relationship of young firms. Only Haltiwanger et al. (2013) use a dataset which has been edited by advanced record linking methods to distinguish between real and spurious entry and exit. It is therefore not surprising that our results are more in line with that study.

---

23 The problem that large-scale firm-level data suffer from spurious entry and exit due to administrative or legal changes, has been recognized since the nineties. However, only with the recent development of advanced record linkage methods, has the extent of the problem and its profound impact on empirical results become clear. Most previous studies did not or could not address this problem. Evans (1987a, 1987b) uses U.S. data from the Dunn and Bradstreet files which are known to suffer from data problems with respect to young and small firms (Davis et al., 1996). Almus and Nerlinger (2000), Lotti et al. (2003) and Mata (2004) do not report the use of linkage methods to clean the sample from spurious entry and exit. Wagner (1994) recognizes that large entry is unlikely and therefore excludes the largest firms from the entry sample, ignoring that spurious entrants also occur in other size classes. Dunne et al. (1989), using the U.S. Census of Manufacturers, partially correct for ownership changes but not for other administrative changes or changes in firm structure.

24 Unfortunately, most studies do not report the employment distribution at entry, which would be informative about the size range and employment share of larger entrants. Haltiwanger et al. (2013) use traditional record linking methods to eliminate spurious entrants, but additionally rely on physical addresses to more accurately identify entry and exit of multi-establishment firms. It is unclear to what extent their approach identifies all spurious entrants, especially in medium and large size classes where we relied heavily on the employee-flow method. In their sample, larger firms still represent an important share of employment at entry. Firms entering with more than, respectively, 20 or 250 employees represent 50% or 18% of employment at entry. The corresponding shares in our sample are only 7% or less than 1%. At a minimum, it is likely that their entrant population is not limited to de novo entrants as we defined them.
A positive relationship between growth and firm size

We show the size-growth relationship of de novo firms in their first years after entry in Figure 3. We then illustrate the robustness of the pattern in Figure 4 and describe how delayed adjustment can explain it. In Appendix E, we discuss likely reasons why previous studies did not find the same pattern.

Figure 3 plots the coefficients from the employment growth regression of de novo entrants that survive from period t-1 to t. Due to the weighting, they represent the net employment growth rates of the entire group of survivors within each age-size class. The benchmark results use the dynamic size classification to assign firms to a size class, while results using two alternative classification methods follow below. As discussed, each method represents an alternative way to classify firms by current size to approximate a continuous size-growth relationship. For clarity, we do not show confidence bounds but report all coefficient estimates and standard errors in Table A.3 in the Appendix. Coefficients are estimated extremely precisely and almost all successive point estimates are significantly different.

As can be seen from the ordering of the different curves, growth rates decrease with firm age when firm size is held constant, in line with the prediction of the passive learning model. In the first year after entry (age 2), surviving young firms of all sizes exhibit very high growth rates. Thereafter, growth rates decline monotonically with age within every size category. Growth rates fall most strongly between age 2 and age 3, and decline at a decreasing rate when an entry cohort matures. The convergence to the pattern for incumbents (labeled age 6+) has not been completed entirely when entrants reach age 6, i.e. when we have observed them for five years.25

Figure 3. Growth rates of surviving de novo entrants by age and size

Note: Annual averages over the 2003-2012 period

25 Showing the growth rates against age with different curves for each size class highlights the convergence in growth rates to approximately zero growth for incumbents that holds for all firm sizes, see Appendix D.
The more remarkable pattern in Figure 3 is that growth rates are strongly increasing in current size for firms of the same age cohort. Larger firms grow on average more rapidly than smaller firms of the same age.

The positive relationship between growth and size is most pronounced in the first year after entry and gradually weakens with age. Already at age 6, five years after entering the dataset, the relationship has shifted towards growth rates that are almost proportional to the current size of the firm. The point estimates for incumbents suggest that growth rates will continue to decline and eventually converge to growth rates close to zero in all size classes. This contrasts with the exit probabilities, which are inversely related to size even for older cohorts. For the smallest firms, growth has basically stalled after five years while for larger firms growth will remain positive for a few years longer. As a result, the firm distribution will continue to shift to the right as illustrated in Figure A.3 in the Appendix.

A robust relationship

As most de novo entrants start with very few employees, we measure firm growth in the following years over a much narrower range of small size classes than is usually the case in other studies. This heightens the statistical problems associated with the conventional base-year size classification that we discussed earlier. To complement the results based on the dynamic size classification, we show in Figure 4 estimates based on two alternative size classifications. Panel (a) presents results that average over growth rates using the beginning-of-period and end-of-period sizes as base. In panel (b) firms are classified by the average of their size in years $t-1$ and $t$, as in Haltiwanger et al. (2013).

The patterns using both alternative methods are similar to the benchmark results. Growth rates are increasing in firm size within each age class. The strong positive slope in the first few years following entry gradually converges to a virtually flat profile for incumbents. The positive relationship is somewhat more pronounced than in our benchmark results, especially in panel (b), where job gains of fast-growing firms are entirely allocated to the intermediate size class between $t-1$ and $t$. In the dynamic size classification, this growth is allocated to each respective size class the firm passes through.

It is quite remarkable that across the three graphs, there is only a single instance where any of the curves intersect. The patterns we uncover are very smooth and monotonic: growth rates increase with size for each age cohort and decrease with age for each size class. This is even more remarkable given that they have been estimated over the very turbulent 2003-2012 period that includes the Great Recession. The patterns also hold if we limit the sample to firms entering between 2003 and 2007 and follow their growth to at most 2008, the onset of the crisis, or if we limit the sample to firms entering from 2008 onwards.\footnote{Separate results for pre and post-crisis entrants are shown in Geurts and Van Biesebroeck (2014).}
Figure 4. Alternative size classifications: Growth rates of surviving *de novo* entrants

(a) **Average of estimates using firm base size at t-1 and t**

(b) **Average size classification**

Note: Annual averages over the 2003-2012 period

The positive relationship between growth and size of young firms of the same age confirms the results in Haltiwanger et al. (2013) that are obtained using the average size classification. As noted before, that study uses a dataset and size measurement that reduce potential biases when working with small and young firms. An important difference, however, is that the growth rates they report do not evolve to size-invariant growth among older firms, while many studies have found that Gibrat’s law is a good approximation of the size-growth relationship among large and well-established firms (Mansfield 1962; Hall, 1987; Geroski 1995). In addition, a positive size-growth pattern cannot be a steady state as the firm size distribution would collapse.

**What explains the positive relationship?**

We have argued that the positive relation between growth and size among young firms of the same age cohort is not at odds with the predictions of Jovanovic (1982) if one takes into account that young firms exhibit some lag of adjustment to prior information.27

Firms that receive positive information, i.e. learn that they are more efficient than previously realized, will not always adjust completely to this new information right away. Risk aversion might induce them to wait an extra period for the positive information to be

---

27 While the model in general has no prediction for the size-growth relationship conditional on age, under some assumptions—in particular constant returns to scale—growth rates should be size invariant. If firms adjusted to information in year $t$ in line with Bayes’ law, optimally weighing their prior and the new information, the random arrival of new information in $t+1$ would be uncorrelated with firm size.
confirmed or it might take some time for additional capacity to become operational. Financial constraints, hiring frictions, or regulations can impose external barriers that need to be overcome before a firm can expand its operations. Such partially delayed growth will induce a positive size-growth relationship. Some of the positive news leads to instantaneous growth and raises a firm’s current size. The remaining fraction of growth postponed to subsequent years then leads to a positive correlation between growth and size.

A corresponding delay for firms that adjust to negative information will further strengthen the positive correlation. If annually recurring fixed costs of operation are sufficiently low relative to sunk entry costs, firms might delay their eventually withdrawal from the market even as they make losses. In the administrative data set we even observe many firms with no employees for some years. We omitted them from the analysis, but it suggests that merely surviving might not be all that costly. As firms adjust their size downward but postpone exit, it leads to low or negative growth rates for smaller firms.

Figure A.4 in the Appendix provides some evidence for such behavior. In panel (a), firms that are about to exit in the next period exhibit much lower growth rates than firms that will survive. The difference is approximately constant in each of the 5 years following entry. Average growth rates are negative for impending exiters at all ages except age 2, indicating that firms stay small or decline in the year before they exit. The difference in growth rates already appears two years before exit, shown in grey, but is less pronounced. Given that there are many more firms exiting in the smaller size classes, this pre-exit growth difference contributes to the positive size-growth relationship.\textsuperscript{28}

6 Conclusions

In constructing the dataset, we have taken great care to identify a sample of firms that start new operations, corresponding to actual new firm creation. Complementing the traditional firm linkage method with an employee-flow method, we filtered out misclassified, spurious entrants. Given their incumbent-like entry distribution and growth patterns, they bias the patterns of interest. By establishing a more complete set of firm linkages, we also avoid confusing firm restructuring events with economic exits. For the remaining group of \textit{de novo} entrants, we confirm several patterns from the literature. In particular, exit rates are shown to be strongly declining in age and size, while growth rates for survivors decline with age and also with size if we pool across age cohorts.

In addition, we obtain two novel findings. First, we find that firm entry sizes are reduced to a narrow range of small size classes. Second, growth rates of \textit{de novo} entrants are increasing with size in the first years, but quickly converge to proportionate growth as an entry cohort matures. The firm size distribution at entry differs more markedly from that of mature firms than is usually the case, but the positive size-growth pattern accelerates the

\textsuperscript{28} Panel (b) in the same figure shows that delayed exit does not explain the observed positive size-growth relationship entirely. Excluding all \textit{de novo} entrants that exit before age 6 and re-estimating the growth rates still shows a positive relationship in the first years that gradually converges to a size-invariant pattern.
tendency towards increased concentration in an entry cohort and leads to a pronounced right shift in the firm size distribution.

The exit and growth patterns by age and size class are remarkably regular. We have estimated them over an extremely turbulent time period that includes the Great Recession, but all age and size patterns are entirely monotonic. The persistent features of firm dynamics of very young firms seem to dominate cyclical factors.

Our results are consistent with firms having a very imperfect knowledge of their productivity at entry. All patterns are in line with the passive learning model of Jovanovic (1982) where a firm’s underlying efficiency is constant, but is only discovered as a firm operates in the market. If we add delayed adjustment, both in exit and in growth, even the positive size-growth relationship for young firms is consistent with the model.

Note that frictions could even influence firms’ choice of initial entry size. Evans and Jovanovic (1989) provide an alternative model where liquidity constraints force some firms to enter below their desired size and grow into their optimal size afterwards using retained earnings to expand. This would lead to a negative size-growth relationship as constrained, smaller entrants would have a greater upside potential. Both the narrow firm size distribution and the positive size-growth relationship we have documented are more supportive of constraints affecting firms following their entry decision rather than before.

The analysis also has a few policy implications. A recent literature has documented that especially in less developed economies, production factors are often stuck at unproductive firms (Hsieh and Klenow, 2009). This type of misallocation lowers potential output and aggregate productivity. Our evidence suggests that new firms do not know their own likelihood of success very well and it is inevitable that some unproductive entrants end up with too much resources. Policy should accommodate this by making sure that adjustments to firm size after entry are easy to make. At the same time, lowering entry barriers in a situation where adjustment frictions after entry are large is likely to generate bad aggregate outcomes.

We have suggested that delayed adjustment is one mechanism that can explain the observed positive size-growth relationship. In increasingly global markets and with rapid technological advancement, such growth delays can be quite costly. New entrants often have only a narrow window of opportunity to occupy a market niche. If scaling-up in response to positive information happens too slowly, a firm risks coming too late and be shut out of the market by early movers.

Guner et al. (2008) provide evidence that many government policies favor small firms. This is often rationalized on the assumption that small firms are the engine of job creation in the economy. Previous literature has already highlighted that one should not confuse the (conditional) effects of age and size—it tends to be young firms which are vital for job creation. Our current findings cast further doubt on the employment growth potential of small entrants. Among young firms of the same age, those showing up in the dataset with a smaller size also tend to grow more slowly subsequently.

In continuous time, one can think of firm entry as the moment the first employee is hired. Some entrants add additional employees in the next minutes or days, while others take years.
With adjustment frictions, it is likely that a size-pattern established early on will be perpetuated over time. A small size, conditional on age, is indicative of negative news about a firm’s profitability early on. While not all firms can freely choose their size—a large literature documents constraints and frictions that limit a firm’s initial size—our overall patterns suggest that by and large small firms choose to be small. Directing subsidies primarily towards the smallest firms or imposing size restrictions to qualify for government support are policies that should be avoided.
References


Appendix

A. Data

The analysis is based on a firm-level dataset maintained by the National Social Security Office (NSSO) of Belgium. It covers the universe of firms with at least one employee over the period 2003-2012. For comparability with other studies, we restrict the analysis to firms in the private, non-farm sector and also exclude highly subsidized sectors which receive strong support from government programs. In an average year, the sample includes 178,000 firms and 2,070,000 employees. Total employment increased during the sample period by 0.9 percent per year till 2008, dropped by 2.5 percent between 2008 and 2010 and has been more or less stable since.

Large-scale firm-level data collected for administrative or statistical purposes have become the main information source for empirical analysis on firm dynamics. Two advantages are that they provide information about the full distribution of firms, including the smallest, and that individual firm histories can be observed over a long period. A drawback, however, is that changes in ID code or firm structure can mistakenly introduce entry and exit events. This so-called longitudinal linkage problem is widely recognized, but rarely adequately solved. It generates various biases in empirical measures, such as spurious measures of entry and exit rates, misclassification of firm growth across age and size classes, and an overestimation of employment turnover (Haltiwanger et al. 2013). Because the extent of the problem is register-specific, it also hampers comparative analysis.30

The problem is as follows. Between the moment a firm starts operations and exits the market, the unique ID code that identifies it in the dataset may change for various reasons. The administration may assign a new ID code when the ownership or legal form changes or the firm itself may, for tax optimization or liability reasons, close down their legal entity and continue the same activities in a newly registered company. Instead of being observed as one continuing firm, the firm will be observed twice: once as an exit and once as an entrant. This type of exit is unlikely to be preceded by the same firm dynamics that precede economic failures and this type of re-entry in the dataset is clearly different from de novo firm creation (Dunne et al. 1988; Baldwin and Gorecki 1987).

Changes in firm structure as a result of mergers, takeovers or split-offs create additional longitudinal linkage problems that may even involve multiple firms. In addition to spurious exits and entries, they lead to administrative transfers of employees between ID codes that appear in the data as large expansions or contractions of individual firms.

---

29 Table A.4 lists all NACE sectors we include in the analysis and classifies them into six industries. Excluded sectors include “Human health and social work activities,” where most expenditures are publicly financed, and “Subsidized household help,” where service vouchers subsidize 70% of the wage cost.

30 Davis, Haltiwanger and Schuh (1996b) and Baldwin et al. (1992) discuss the linkage problem in detail. Villhuber (2008) and Bartelsman et al. (2009) discuss the implications for cross-country comparisons.

31 See Benedetto et al. (2007) for a discussion of these practices in the U.S.
The straightforward solution is to link across years the ID codes that belong to the same firm, or in the case of restructurings, to parts of the same firm. National statistical agencies traditionally implement a probabilistic record linking method, but Geurts (2016) shows that this method tends to miss many events leading to distorted measures of firm dynamics. We complement the traditional linkage method with a second method, using an employee-flow approach to deal with many forms of restructuring. In addition to repairing broken links, we impute consistent employment measures for young firms that bridge those links, for up to the sixth year of existence. To our knowledge, we are the first to use this approach to obtain consistent post-entry firm histories.

In contrast to many other countries, administratively imposed ID changes are very rare in Belgium. Firms are uniquely identified by the official Belgian enterprise number (CBE number), which each new enterprise receives upon registration and keeps for its entire lifetime, even when the legal status or ownership changes. It makes the NSSO dataset a good starting point for longitudinal firm analysis.32

The first linking method we apply has been developed by Statistics Belgium and implements the OECD-Eurostat recommendations on business demography statistics (Eurostat-OECD 2007). It exploits information on firm continuity from a comprehensive database that combines information from different administrations such as the national register of legal entities, the trade register, VAT declarations, and Social Security reports. In addition, it relies on a probabilistic matching procedure that uses similarities in firm name, address, and industry code to link different ID codes of the same firm across two years.

Our second linking method uses a definition of firm continuity that is based on its workforce. It follows one of the main production factors of the firm, the stock of employees, to trace changes in ID codes and firm structure. This so-called employee-flow approach refines the method pioneered by Baldwin, Dupuy and Penner (1992) for Canada and implemented for the United States by Benedetto, Haltiwanger, Lane and McKinney (2007). It exploits the linked employer-employee information in the NSSO dataset: both firms and employees are identified with a unique ID code. The advantage is that an individual never changes ID and can always be followed. If a firm changes ID code but continues its activities, the stock of employees will largely be the same for the old and the new firm ID. Similarly, when firms merge or split up, this will be reflected in a merge or division of workforces. Continuity of the workforce can thus be used to identify firms that operate continuously but change ID code or firm structure.

In practice, we follow clusters of employees that move simultaneously from one ID code to another between two quarterly observations. A set of decision rules regarding the size of the employee cluster relative to the firms’ total workforce is used to determine whether we should consider the two ID codes as a single, continuing firm. The primary rule, to identify one-to-one ID changes, verifies whether the cluster represents at least 50 percent of the workforce of

---

32 The CBE number only changes when a self-employed transforms its activities into a legal company. Vilhuber (2008) surveys the practices in several countries. Baldwin et al. (1992), Jarmin and Miranda (2002), and Hethey and Schmieder (2013) provide details respectively for Canada, the United States, and Germany.
both the disappearing and the newly appearing ID code. A second rule identifies takeovers, allowing the receiving ID code to exist already, but requiring a cluster of at least 75 percent of the workforce of the initial ID code to move together. A set of additional decision rules is listed in Table A.1 and these capture takeovers, split-offs and other forms of organizational restructurings. The table shows that the first two rules account for 80% of the identified links. If the cluster does not satisfy any of the rules, we leave the administrative data as is. In line with Baldwin et al. (1992) and Benedetto et al. (2007), we only use clusters with at least five employees. For smaller clusters, there is a high probability that an employee flow between two ID codes merely represents individual job changes. Due to the minimum cluster size, the employee-flow method is inappropriate for identifying missing linkages of the smallest firms. Geurts (2016) conducts several robustness checks to verify the sensitivity of measures of firm dynamics to alternative size thresholds and decision rules of the employee-flow method. She finds that they are not critical to the empirical results.

The linkages established by the two record linking methods are first used to identify continuing firms that are misclassified as entrants and exits. They are labeled as ‘spurious’ entrants and exits as opposed to de novo entrants and true exits. Panel (b) of Table A.1 shows that 78 percent of the spurious entrants we identify are simply incumbents that continue the same activities with a new identification code after a purely administrative or legal change. Another 18 percent are split-offs of another firm. 33 Second, for those firms that are involved in an ID change or restructuring, administratively recorded employment changes from one period to the next do not reflect internal job growth but are but artificially inflated or deflated by the event. Therefore, as a further step in the data editing, employment of these firms is imputed in the years after the event. Our approach is to construct an aggregate event-level that includes all firm ID’s interlinked from \( t-1 \) to \( t \). Firm-level employment in \( t \) and \( t+n \) is then imputed by assuming the same growth rate for each firm involved in the event. The imputation procedure is extended to the sixth year of existence for de novo entrants. 34 For one-to-one ID changes, which represent the vast majority of events, the imputation method simply corresponds to replacing the new by the old ID code. With respect to more complex events, the imputation method treats break-ups and mergers of firms symmetrically and preserves the firm size distribution in the sample. Imputed employment histories more closely reflect actual job creation or destruction at the firm level and allow a more accurate estimate of post-entry exit and growth patterns by size.

Table 1 in the text showed that the two linkage methods are strongly complementary for the accurate identification of de novo entry across different size classes of firms. The traditional method is needed especially in the size class below five employees, where employee-flow links are absent by construction. Yet the employee-flow method is essential

---

33 Some administrative entrants are subsidiaries of foreign firms entering the Belgian market and are not de novo entrants either. Our linkage methods are unable to identify these FDI entrants. As it is an extremely small group, their presence is unlikely to affect the results. On a reduced sample, covering the 2005-2010 period, we find that they represent fewer than 1 percent of all de novo entrants.

34 We also impute employment for mature firms involved in an event to calculate consistent employment growth rates for them, which we use as a comparison for the evolution of de novo firms.
in larger size classes, where it identifies two to three times more spurious entrants than the traditional method. Table 1 further showed that the probability that a new ID code corresponds to a spuriousentrant dramatically increases with size. The size-distribution of de novo entrants is more strongly right-skewed than in the unedited data.

The linkage methods similarly divide the group of de novo young firms that disappear from the dataset into true economic and spurious exit. The extent of misclassification is somewhat lower than on the entry side, 4 percent of administrative exits are identified as spurious, but the likelihood is again increasing with firm size. In the working paper, see Geurts and Van Biesenbroeck (2014), we report those statistics and provide separate summary statistics for all the different groups of entrants and exiting firms.

B. Size classification

Regression-to-the-mean and sample selection may spuriously introduce a negative relation in estimates of the relationship between growth and size of surviving firms if firms are classified by their size in the base year t-1. The extent to which these problems bias actual empirical results, and possible solutions have been extensively debated in the literature, without reaching a unanimous conclusion so far.\footnote{For a discussion see for example Hall (1987), Baldwin and Picot (1995), Davis et al. (1996b), Davidsson et al. (1998), and Kirchhoff and Greene (1998).} As discussed before, both problems are exacerbated if growth rates are measured in a population of predominantly small firms, as is the case in our sample of de novo entrants. We therefore need to directly address these measurement problems. To avoid bias in the size-growth relationship, we use three alternative firm-size classifications that approximate a continuous size-growth relationship.

The first size classification method, and the one we use for our benchmark estimates, allocates employment gains and losses to each respective size class in which the growth or loss occurred. This ‘dynamic’ sizing is used by the U.S. Bureau of Labor Statistics to avoid base-year classification biases in the Business Employment Dynamics statistics (Butani et al. 2006), and is further discussed in Davidsson, Lindmark and Olofsson (1998) and de Wit and de Kok (2014). Firms are initially assigned to a size class based on employment in t-1, but are re-assigned to a new class when they cross a threshold. The growth from \(E_{it-1}\) to the threshold is assigned to the initial class and the remaining growth from the threshold to \(E_{it}\) is assigned to the next size class. Growth rates use average employment in the denominator as discussed in Section 4 of the main text, but use the intermediate size class thresholds as upper or lower limits. This methodology approximates instantaneous class re-assignment that would be feasible if size and growth were measured in continuous time. We choose the size class thresholds such that they imply symmetric and (almost) equal ranges of potential growth rates within each class between -0.67 and +0.67.\footnote{The size thresholds between the size classes [0.2], [2.4], [4.8], [8.16], [16,32], and [32,\infty] are 2, 4, 8, 16, and 32 for expansion and 1.85, 3.7, 7.4, 15, 31 for contraction. This yields growth ranges of [-0.60,0.67], [-0.67,0.67], [-0.67,0.67], [-0.68,0.67], [-0.70,0.67], [-0.68,0.67], and [-0.67,\infty] respectively.} This approach mitigates the negative bias in the size-growth relationship caused by regression-to-the-mean because symmetric growth
and decline are equally attributed to the same size classes. The problem of left-truncated growth rates in the smallest size classes is also mitigated because the range of growth rates within each size class is symmetric with mean zero. The equal ranges of potential growth rates further imply that no size class is favored when the sample exhibits on average positive (or negative) growth, avoiding the upward size-growth bias of the methodology used by Haltiwanger et al. (2013) discussed below.

The second classification method uses each firm twice in the regression, assigning a weight of one half to each observation. One observation uses the firm’s employment level at the beginning of the period both as a base for the growth rate and to determine the size class. The second observation uses the firm’s employment at the end of the period for both calculations. Growth rates of firms assigned to the same size class based on $E_{it-1}$ or $E_{it}$ contribute to the regression in a symmetric way as before. Firms assigned to different size classes can show a different size-growth relationship in each instance and both contribute equally to the average pattern identified in the regression. This approach has been proposed by Prais (1958) to avoid regression-to-the-mean bias and can be motivated similarly as the use of average wage shares in a Solow residual, i.e. as a discrete approximation to the continuous Divisia index of productivity growth (Caves, Christensen and Diewert, 1982).

For comparison with the results of Haltiwanger et al. (2013), our last classification method uses the average of firm size in years $t-1$ and $t$ as a proxy for the size over the intervening period. This size classification, proposed by Davis et al. (1996a, 1996b), reduces the regression fallacy and the truncation problem. If firm size fluctuates around a stable long-run size, using the average size classification would yield unbiased results. However, in a sample with an average positive growth rates, it introduces an upward bias between size and growth (Baldwin and Picot 1995). Rapidly growing firms are more likely to cross a size class border and their measured rate of growth will be entirely reassigned to a higher size class.

In Figure A.5, we report regression results on a simulated dataset where we imposed the same average growth rate for all size categories. We started from a cohort of de novo entrants that replicates the actual entry size distribution observed in the data. We then applied a stochastic growth rate to each observation that averaged 10 percent regardless of size, but with a large dispersion, as in the observed data. We then applied an exit rule that was stochastically decreasing in firm size, generating an exit probability that is negatively correlated with the growth rate. The size-growth relationship was then estimated using each of the size classification methodologies just discussed and also using the base-year classification. The graph plots the regression coefficients on the different size class dummies. The results confirm the strong downward bias in the size-growth relationship for the base-year classification and a much more constant relationship for the three alternatives, especially for firms with at least 4 employees.

---

37 For further discussion see also Davidsson et al. (1998) and Kirchhoff and Greene (1998).
C. Confirmed patterns

As found in many other countries, the annual entry rate is high but involves only a small fraction of the labor force. Statistics in Table A.2 show that de novo entrants represent 9 percent of all active employer firms in a given year, but only 1.5 percent of total employment. Most entrants are small. Average entry size is 1.9 employees, six times smaller than the average size of incumbents. In the years following entry, a large fraction of the entering cohort exits and the average firm size among survivors increases. Only half of all entrants are still around at age 6, at which time the average firm size in the surviving group has almost doubled. Job creation by survivors is substantial and almost compensates for job loss due to the exit of young firms. Total employment created by an entry cohort falls only slightly below its initial value in the five years after entry.

As the entry cohort matures, the size distribution becomes more concentrated as illustrated by the kernel density in Figure A.3. The strongly right-skewed distribution at entry gradually gets a fatter right tail, but at age 6 it has not yet converged to the distribution of incumbents.

A first mechanism generating this pattern of increased concentration in an entry cohort is selective survival. In line with the predictions of the passive learning model, we find high exit rates for young firms which are decreasing in age as well as in size. This is shown in panel (a) of Figure A.1, which plots the age-size coefficients for the exit regression representing job destruction rates for each age-size class.\textsuperscript{38} Exit rates are especially high in the first full year of existence, from age 1 to age 2, and then rapidly decrease with age. Five years after entry, exit rates have approximately halved, but they are still significantly higher than for incumbents, i.e. firms older than six years. The ordering of the lines for different size classes further shows that exit rates decline with size within every age cohort. The same pattern holds for each age group and is even true for incumbents. These results suggest that the selection process of the passive learning model—which predicts market exit of the least efficient and therefore the smallest firms—unfolds quickly in the first years after entry.

Panel (b) of Figure A.1 shows that a second prediction of the passive learning model is also borne out in the Belgian data. Surviving young firms exhibit high growth rates in the early years after entry, but growth slows down rapidly with age. In contrast with the exit probabilities which decline at a relatively constant pace, the growth slowdown is most pronounced in the first few years. The average growth rate declines convexly as it converges to a constant steady state. On average, surviving young firms at age 6 still show a positive growth rate of 4 percentage points while the average incumbent does not show any employment growth.

Much higher growth rates of young firms—which are overrepresented in smaller size classes—induce a negative relationship between growth and size in a cross-section of firms of all ages. Such a relationship has often been documented in the literature and it is also what we find for Belgium, as shown by the ‘all firms’ line in Figure 1 in the text. Average growth rates among all firms surviving from year $t-1$ to $t$ decline monotonically with the current size

\textsuperscript{38} Recall that all regression coefficients are estimated using employment weights.
of the firm. As incumbents dominate this population, the absolute growth rates are rather low, especially beyond the first two size classes.

It is instructive, however, to show the size-growth relationship separately for young firms that entered the sample at most five years ago, and older firms. The dashed line at the bottom of Figure 1 shows low growth rates for incumbents regardless of firm size. For them, absolute employment growth is proportional to the current size of the firm, confirming Gibrat’s law in our data set. In contrast, growth rates for young firms are not only higher, they clearly increase with size.

Except for this last finding for young firms, all patterns described so far are in line with results from other empirical studies based on large-scale firm-level datasets, even when no or little attempt has been made to distinguish between what we have labelled de novo and spurious entrants. It suggests that most patterns are fairly robust to less accurate identification of truly new and young firms. The positive relationship between growth and size that we observe among young de novo firms, however, is not replicated in the full sample of administrative entrants. Instead, as indicated by the light gray line in Figure 1, the raw, administrative data suggest that small young firms have higher growth rates than larger ones. In Section 5.3 below, we show that the difference is even more pronounced when growth rates are estimated conditional on age, that the pattern of the solid black line is robust, and how spurious entry biases the estimated relationship.

D. Alternative representations of the exit probability and growth of survivors patterns

In the main text we have discussed the growth rates in Figures 3 and 4 with size on the X-axis and separate curves for each age cohort. We could present the same coefficients with age on the X-axis, as we did for the exit rates in Figure A.1(a). Figure A.2 shows both possibilities for both dependent variables. For the exit rates, the two figures look extremely similar as the probability of exit is declining with age and size. The only difference is that the decline with age is linear, while rates decline more convexly with size.

Showing the growth rates against age with different curves for each size class, in panel (b.1), highlights the convergence in growth rates to approximately zero growth for incumbents that holds for all firm sizes. The declining growth rates with age now show up as downward-sloping lines, rather than a downward shift of the different curves in Figures 3 and 4. An important difference with the corresponding figure for exit rates is that the ordering of the different size-class curves is reversed. The curve for largest firms is at the top and curves for smaller firms are further down. Smaller entrants already start out closer to the steady state growth rate, while larger entrants grow rapidly at first and take more time to converge.

39 The patterns in both panels of Figure A.1 and those for incumbents and all firms in Figure 1 in the text are qualitatively the same when calculated using the raw administrative data, reported in Geurts and Van Biesebroeck (2014).
E. Why do many studies find a negative relationship?

Several reasons why previous studies did not find the same positive relationship between growth and size of young firms have been mentioned briefly in the text. This section provides a point by point discussion. A first reason is that not all studies condition on age, which is crucial. For example, Dunne et al. (1989) find a negative relationship but lump all firms up to age 5 in one group. Similarly, Mata (1994) finds a significant negative relationship only when young firms up to age 4 are pooled into one age class. Given the important share of young—on average high-growth—firms in smaller size classes, while larger size classes contain almost exclusively older—low-growth—firms, composition effects induce a negative relationship if firms of different ages are pooled. Pooling across all firms, incumbents and young firms, we also found a negative relationship in our dataset, see Figure 1 in the text.

A second reason is the inherent negative bias induced by the conventional base year classification. Most recent studies use a base year classification but control for potential bias using various other solutions than to one presented in this paper. Hence it remains unclear to what extent the difference in results is explained by different methodologies. One solution adopted by Mata (1994) is to omit all firms that enter with fewer than 10 employees to avoid truncated growth rates of the smallest firms. The same solution adopted to our sample of de novo entrants would imply to exclude 98.5 percent of the firms at entry. It is questionable whether the growth patterns of the 1.5 percent largest entrants are representative for those of total population of new firms entering the market. Evans (1987a) and Lotti et al. (2003) do include entrants of all sizes and use other estimation techniques to control for sample selection bias. Still, they report an inverse growth-size relationship for young firms even given age. Importantly, however, they also find convergence towards proportional growth rates for older firms, as we do.

A third reason is that spurious entrants are generally not adequately filtered out from the dataset. Since they are misclassified older firms, their growth rates tend to be much lower, resembling those of incumbents. As spurious entrants dominate in larger size classes, they introduce a downward bias in post-entry growth rates that is strongly increasing with firm size. This effect is shown in panel (b.3) of Figure A.2 which replicates Figure 3 on the full sample of administrative entrants. The bias is hardly noticeable in the smallest size classes where the share of spurious entrants is negligible. Yet, in larger size classes where spurious entrants represent the majority of administratively recorded entrants, their low growth rates swamp the high growth rates typically observed for de novo entrants. It obscures the positive relationship between growth and current size and even reverses it at age 2. Growth rates seem to be size invariant already from age 3 onwards.

Misclassified exits have a similar effect on the estimated pattern. Larger entrants that grow strongly are more likely to be involved in a restructuring that changes their firm ID, but is not economic exit. Some firms reorganize to cope with higher than expected growth rates, for example by splitting off some activities or adopting a different administrative structure. Other

---

40 In Geurts and Van Biesebroeck (2014) we show growth rates in all age-size classes separately for spurious entrants which highlights their uniformly low growth rates.
firms are taken over by rivals that see the growth potential. Misclassifying such events involving large firms that grow strongly as exits obscures the positive size-growth relationship.

Finally, most previous studies, e.g. Evans (1987a), Dunne et al. (1989), Mata (1994), Lotti et al. (2003), focus on the manufacturing sector where the positive relationship is weaker also in our dataset. We find the increasing relationship to be most pronounced in the sectors of ‘business support services’ and in ‘mixed business and household services,’ where entry costs are often lower—graphs by industry are shown in Geurts and Van Biesebroeck (2014). Firms in these sectors can easily enter with a very small size and gradually adjust to optimal scale. Consistent with a higher minimum efficient scale in manufacturing, we find firms to enter with higher average size and show a much weaker size-growth relationship.

F. Additional Figures and Tables
Table A.1 Employee flow links by decision rule

An employee-flow link between two firm identification numbers is established if a cluster of at least 5 employees moves from one firm ID in quarter q-1 (the ‘predecessor’) to another firm ID in quarter q (the ‘successor’), and if the decision rules in Table A.2(a) are met.

(a) Type of employee-flow linkages by decision rules

<table>
<thead>
<tr>
<th>Type of linkage</th>
<th>Number of predecessors to successors</th>
<th>Predecessor type</th>
<th>Successor type</th>
<th>Minimum absolute cluster size (n employees)</th>
<th>Minimum relative cluster size</th>
<th>Minimum relative cluster size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ID-change</td>
<td>1 to 1</td>
<td>-</td>
<td>-</td>
<td>5</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>2. Takeover 75%</td>
<td>1 to 1</td>
<td>exit</td>
<td>continuing</td>
<td>5</td>
<td>75%</td>
<td>-</td>
</tr>
<tr>
<td>3. Split-off 75%</td>
<td>1 to 1</td>
<td>continuing</td>
<td>entrant</td>
<td>5</td>
<td>-</td>
<td>75%</td>
</tr>
<tr>
<td>4. Takeover 50%</td>
<td>1 to 1</td>
<td>exit</td>
<td>continuing</td>
<td>10</td>
<td>50%</td>
<td>-</td>
</tr>
<tr>
<td>5. Split-off 50%</td>
<td>1 to 1</td>
<td>continuing</td>
<td>entrant</td>
<td>10</td>
<td>-</td>
<td>50%</td>
</tr>
<tr>
<td>6. Merger of exits</td>
<td>n to 1</td>
<td>all exits</td>
<td>entrant</td>
<td>5</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>7. Break-up into entrants</td>
<td>1 to n</td>
<td>exit</td>
<td>all entrants</td>
<td>5</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>8. Merger other</td>
<td>n to 1</td>
<td>-</td>
<td>entrant</td>
<td>5</td>
<td>-</td>
<td>25%, 50%</td>
</tr>
<tr>
<td>9. Break-up other</td>
<td>1 to n</td>
<td>exit</td>
<td>-</td>
<td>5</td>
<td>25%</td>
<td>50%</td>
</tr>
<tr>
<td>10. Cluster &gt;= 30</td>
<td>1 to 1</td>
<td>-</td>
<td>-</td>
<td>30</td>
<td>10%</td>
<td>10%</td>
</tr>
</tbody>
</table>

- Share of the sum of the clusters in total employment of the predecessors
- Share of each individual cluster in employment of successor
- Share of each individual cluster
- Share of the sum of the clusters

(b) Share of employee-flow linkages by type

<table>
<thead>
<tr>
<th>Type of linkage</th>
<th>All links</th>
<th>Spurious entrants</th>
<th>Transfers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ID-change</td>
<td>0.57</td>
<td>0.78</td>
<td>0.71</td>
</tr>
<tr>
<td>2. Takeover 75%</td>
<td>0.22</td>
<td>-</td>
<td>0.15</td>
</tr>
<tr>
<td>3. Split-off 75%</td>
<td>0.12</td>
<td>0.18</td>
<td>0.05</td>
</tr>
<tr>
<td>4. Takeover 50%</td>
<td>0.01</td>
<td>-</td>
<td>0.02</td>
</tr>
<tr>
<td>5. Split-off 50%</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>6. Merger of exits</td>
<td>0.01</td>
<td>-</td>
<td>0.01</td>
</tr>
<tr>
<td>7. Break-up into entrants</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>8. Merger other</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>9. Break-up other</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>10. Cluster &gt;= 30</td>
<td>0.03</td>
<td>0.00</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Note: Total sums to one in each column. Annual averages over the sample period.
<table>
<thead>
<tr>
<th></th>
<th>Entry rate</th>
<th>Employment share</th>
<th>Exit rate</th>
<th>Share of survivors</th>
<th>Employment share of survivors</th>
<th>Average size (employees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 1 (entry)</td>
<td>0.09</td>
<td>0.015</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.93</td>
</tr>
<tr>
<td>Age 2</td>
<td>0.21</td>
<td>0.79</td>
<td>0.98</td>
<td>2.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 3</td>
<td>0.15</td>
<td>0.68</td>
<td>0.98</td>
<td>2.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 4</td>
<td>0.13</td>
<td>0.60</td>
<td>0.98</td>
<td>3.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 5</td>
<td>0.11</td>
<td>0.54</td>
<td>0.98</td>
<td>3.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 6</td>
<td>0.10</td>
<td>0.49</td>
<td>0.98</td>
<td>3.61</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Annual averages over the 2003-2012 period. The year a firm enters the dataset is indicated by Age 1.
Table A.3  Coefficient estimates shown in Figures 3 and 4
(Growth rates of surviving de novo entrants by age and size)

(a) Dynamic size classification

<table>
<thead>
<tr>
<th>Firm size class (employment)</th>
<th>1</th>
<th>2-4</th>
<th>4-7</th>
<th>8-15</th>
<th>16-31</th>
<th>32+</th>
</tr>
</thead>
<tbody>
<tr>
<td>De novo entrants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age 2</td>
<td>0.133 (.002)</td>
<td>0.150 (.002)</td>
<td>0.163 (.002)</td>
<td>0.191 (.003)</td>
<td>0.196 (.004)</td>
<td>0.234 (.005)</td>
</tr>
<tr>
<td>age 3</td>
<td>0.068 (.003)</td>
<td>0.067 (.002)</td>
<td>0.077 (.002)</td>
<td>0.101 (.003)</td>
<td>0.113 (.004)</td>
<td>0.142 (.004)</td>
</tr>
<tr>
<td>age 4</td>
<td>0.037 (.004)</td>
<td>0.038 (.003)</td>
<td>0.049 (.003)</td>
<td>0.070 (.003)</td>
<td>0.069 (.004)</td>
<td>0.081 (.004)</td>
</tr>
<tr>
<td>age 5</td>
<td>0.026 (.004)</td>
<td>0.025 (.003)</td>
<td>0.029 (.003)</td>
<td>0.038 (.003)</td>
<td>0.038 (.004)</td>
<td>0.059 (.004)</td>
</tr>
<tr>
<td>age 6</td>
<td>0.019 (.005)</td>
<td>0.014 (.003)</td>
<td>0.025 (.003)</td>
<td>0.028 (.003)</td>
<td>0.033 (.004)</td>
<td>0.032 (.004)</td>
</tr>
<tr>
<td>Incumbents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age 6+</td>
<td>0.006 (.001)</td>
<td>0.003 (.001)</td>
<td>0.004 (0.000)</td>
<td>0.004 (0.000)</td>
<td>0.004 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
</tbody>
</table>

(b) Average of estimates using firm base size at \( t - 1 \) and \( t \)

<table>
<thead>
<tr>
<th>Firm size class (employment)</th>
<th>1</th>
<th>2-4</th>
<th>4-7</th>
<th>8-15</th>
<th>16-31</th>
<th>32+</th>
</tr>
</thead>
<tbody>
<tr>
<td>De novo entrants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age 2</td>
<td>0.166 (.004)</td>
<td>0.198 (.004)</td>
<td>0.229 (.004)</td>
<td>0.272 (.006)</td>
<td>0.312 (.007)</td>
<td>0.309 (.008)</td>
</tr>
<tr>
<td>age 3</td>
<td>0.048 (.004)</td>
<td>0.083 (.003)</td>
<td>0.105 (.003)</td>
<td>0.145 (.004)</td>
<td>0.167 (.005)</td>
<td>0.161 (.005)</td>
</tr>
<tr>
<td>age 4</td>
<td>0.009 (.005)</td>
<td>0.046 (.004)</td>
<td>0.066 (.003)</td>
<td>0.094 (.004)</td>
<td>0.086 (.005)</td>
<td>0.104 (.005)</td>
</tr>
<tr>
<td>age 5</td>
<td>0.002 (.005)</td>
<td>0.023 (.004)</td>
<td>0.042 (.003)</td>
<td>0.051 (.004)</td>
<td>0.065 (.005)</td>
<td>0.061 (.004)</td>
</tr>
<tr>
<td>age 6</td>
<td>-0.011 (.006)</td>
<td>0.015 (.004)</td>
<td>0.034 (.003)</td>
<td>0.038 (.004)</td>
<td>0.038 (.005)</td>
<td>0.042 (.005)</td>
</tr>
<tr>
<td>Incumbents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age 6+</td>
<td>-0.029 (.001)</td>
<td>-0.006 (.001)</td>
<td>0.002 (.001)</td>
<td>0.006 (.001)</td>
<td>0.008 (.001)</td>
<td>0.001 (0)</td>
</tr>
</tbody>
</table>
Table A.4 Six main industries and NACE Rev. 2 classes

<table>
<thead>
<tr>
<th>Nace Rev. 2 classes</th>
<th>De novo entrants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of firms</td>
</tr>
<tr>
<td>1. Manufacturing and energy</td>
<td>777</td>
</tr>
<tr>
<td>Section B, C, D, E</td>
<td></td>
</tr>
<tr>
<td>2. Construction</td>
<td>2,730</td>
</tr>
<tr>
<td>Section F</td>
<td></td>
</tr>
<tr>
<td>3. Wholesale and retail trade</td>
<td>4,236</td>
</tr>
<tr>
<td>Section G</td>
<td></td>
</tr>
<tr>
<td>4. Accommodation and food services</td>
<td>2,793</td>
</tr>
<tr>
<td>Section I</td>
<td></td>
</tr>
<tr>
<td>5. Business support services</td>
<td>2,945</td>
</tr>
<tr>
<td>- Freight transport, handling and storage: Nace 49.2, 49.4, 49.5, 50.2, 50.4, 51.2, 52.1, 52.241, 52.249;</td>
<td></td>
</tr>
<tr>
<td>- IT programming and services: Nace 62, 63;</td>
<td></td>
</tr>
<tr>
<td>- Central banks, holdings, financial leasing, hedgefunds and auxiliary financial services: Nace 64.110, 64.2, 64.3, 64.910, 64.991, 64.992, 64.999, 66;</td>
<td></td>
</tr>
<tr>
<td>- Accounting: Nace 69.2;</td>
<td></td>
</tr>
<tr>
<td>- Head offices: Nace 70;</td>
<td></td>
</tr>
<tr>
<td>- Architecture and engineering: Nace 71;</td>
<td></td>
</tr>
<tr>
<td>- Advertising: Nace 73;</td>
<td></td>
</tr>
<tr>
<td>- Professional and technical support services: Nace 74;</td>
<td></td>
</tr>
<tr>
<td>- Professional rental and leasing: Nace 77.1, 77.3, 77.4;</td>
<td></td>
</tr>
<tr>
<td>- Security: Nace 80;</td>
<td></td>
</tr>
<tr>
<td>- Services to buildings except Cleaning: Nace 81 excl. 81.210, 81.220</td>
<td></td>
</tr>
<tr>
<td>- Administrative services: 82;</td>
<td></td>
</tr>
<tr>
<td>- Repair of ICT: Nace 95.1</td>
<td></td>
</tr>
<tr>
<td>6. Mixed business &amp; household services</td>
<td>2,011</td>
</tr>
<tr>
<td>- Passenger transport and transport services: Nace 49.1, 49.3, 50.1, 50.3, 51.1, 52.210, 52.220, 52.230, 52.290;</td>
<td></td>
</tr>
<tr>
<td>- Postal and courier activities: Nace 53;</td>
<td></td>
</tr>
<tr>
<td>- Publishing, Movies, radio and television: Nace 58, 59, 60;</td>
<td></td>
</tr>
<tr>
<td>- Telecommunication: Nace 61;</td>
<td></td>
</tr>
<tr>
<td>- Banks, credit, insurance instit.: Nace 64.19, 64.921, 64.92, 65;</td>
<td></td>
</tr>
<tr>
<td>- Real estate: Section L;</td>
<td></td>
</tr>
<tr>
<td>- Legal activities: Nace 69.1;</td>
<td></td>
</tr>
<tr>
<td>- Scientific research: Nace 72;</td>
<td></td>
</tr>
<tr>
<td>- Veterinary : Nace 75;</td>
<td></td>
</tr>
<tr>
<td>- Rental and leasing of household goods: Nace 77.2;</td>
<td></td>
</tr>
<tr>
<td>- Travel agencies: Nace 79;</td>
<td></td>
</tr>
<tr>
<td>- Repair of household goods: Nace 95.2;</td>
<td></td>
</tr>
<tr>
<td>- Personal service activities: Nace 99</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15,492</td>
</tr>
</tbody>
</table>

Note: Annual averages (2003-2012) of de novo entrants and employment in entry year. Firms not in the listed categories are excluded from the analysis, primarily quasi-public sector services and subsidized household help.
Figure A.1  Confirmed predictions of the passive learning model

(a) Exit rates of \textit{de novo} entrants by age and size

Note: Annual averages over the 2003-2012 period. Age 6+ refers to incumbents.

(b) Growth rates of surviving entrants by age

Note: Annual averages over the 2003-2012 period. Age 6+ refers to incumbents.
Figure A.2 Alternative representations of the exit probability and growth of survivors patterns

(a) Exit probabilities
(a.1) Exit-age pattern for de novo entrants
(a.2) Exit-size pattern for de novo entrants
(a.3) Exit-age pattern for administrative entrants

(b) Growth rates of survivors
(b.1) Growth-age pattern for de novo entrants
(b.2) Growth-size pattern for de novo entrants
(b.3) Growth-size pattern for administrative entrants
Figure A.3 Evolution of the firm size distribution

kernel = epanechnikov, bandwidth = 0.5000
Figure A.4  Delayed adjustment of *de novo* entrants in exit and growth

(a) Delayed exit: Growth rates of survivors versus exiters

(b) Delayed growth: Growth rates of firms surviving till age 5

Note: sample contains only *de novo* entrants. In panel (b) we use the dynamic size classification.
Figure A.5  Estimated size-growth relationships on simulated data with constant growth rate

Note: Calculations on simulated dataset with growth rates that are size-invariant.