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

Karen Guerts & Johannes Van Biesebroeck

Document Identifier

D3.6 New job creation and entrepreneurship

Version

1.0

Date Due

M36

Submission date

15/05/18

WorkPackage

3

Lead Beneficiary

KU Leuven

Grant Agreement Number 649378

International Journal of Industrial Organization 49 (2016) 59–104

Contents lists available at ScienceDirect

International Journal of Industrial Organization

www.elsevier.com/locate/ijio

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

Karen Geurts[a], Johannes Van Biesebroeck[a,b,*]

[a] *University of Leuven (KU Leuven), Belgium*
[b] *CEPR, United Kingdom*

ARTICLE INFO

Article history:
Received 21 January 2016
Revised 23 August 2016
Accepted 31 August 2016
Available online 10 September 2016

JEL classification:
L11
L25
M13

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; in line with previous studies, we find that growth rates become independent of size as a cohort matures. 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*, *i.e.* newly registered firms that start new operations and hire their first

☆ The authors are grateful to the Belgian National Social Security Office (NSSO), in particular Peter Vets, for collaboration on the development of the employee-flow method, and to Statistics Belgium, in particular Youri Baeyens and Antonio Fiordaliso, for providing firm linkages created for the Eurostat Structural Business Statistics. We would like to thank several people for valuable comments and suggestions along the way: Eric Bartelsman, John Earle, Joep Konings, Otto Toivanen, six anonymous referees and seminar participants at the OECD Workshop Understanding Productivity growth, the Belgian Federal Planning Bureau, the Bank of Finland, the Bank of Italy, Tilburg University, Copenhagen Business School, the Düsseldorf Institute for Competition Economics, CREPP Université de Liège, 2015 ECORE Summer School on Market Power, the IWH Workshop on Entrepreneurship and the Labour Market, ECDEF (Entrepreneurship, Culture, Finance and Economic Development) Conference 21016, and the Third International ZEW Conference on the Dynamics of Entrepreneurship. Financial support from ERC Grant no. 241127 and KU Leuven Program Financing is gratefully acknowledged.

* Corresponding author at: Department of Economics, University of Leuven (KU Leuven), Naamsestraat 69, 3000 Leuven, Belgium.

E-mail addresses: Karen.Geurts@kuleuven.be (K. Geurts), Jo.VanBiesebroeck@kuleuven.be (J. Van Biesebroeck).

http://dx.doi.org/10.1016/j.ijindorg.2016.08.002

Keywords:
Firm dynamics
Passive learning model
Growth

employee. Using two state-of-the-art record linking methods, we distinguish them from pre-existing companies that merely re-register as a new firm, for example following an ID change or merger. The extremely narrow size distribution that we observe for *de novo* entrants provides further support for the passive learning model.

1. Introduction

New firms entering the economy are generally both numerous and small. Empirical studies have consistently documented that many young firms fail shortly after entry and firms that expand have a higher probability of survival than firms that stay small (Evans 1987a; Dunneet al., 1989; Mata et al., 1995).[1] 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 discover gradually by operating in the market. Firms that learn they are highly efficient grow and survive, while the inefficient exit.

Less consensus exists on the growth patterns prevailing among young firms that are able to survive. 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 the model of Jovanovic (Evans 1987a; Haltiwanger et al., 2013; Mata and Portugal, 2004). But it is unclear whether within an entry cohort smaller firms grow faster and to some extent catch up in size, or whether larger firms have higher growth rates. Knowing the form of this relationship is important, as theoretical models of firm dynamics often assume or imply a specific relation between growth and size.

In the general version of Jovanovic's model, the size–growth relationship is undetermined. 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 et al. (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. When using their preferred methodology, Haltiwanger et al. (2013) even find that the size–growth relationship within a given age cohort is positive, both for young and older firms.[2]

[1] Note that some studies have used firms as their unit of analysis while others used plants or establishments. In our analysis, we do not make cross-country comparisons, 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 only have a single plant or establishment, this definition covers a very similar sample of entrants that plant-level studies would identify.

[2] The general results presented in Haltiwanger et al. (2013) are shown in greater detail for young firms in Decker et al. (2014).

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. Importantly, we show that this unique relationship is confined to the very first years of operation. When entrants mature, the empirical pattern converges to growth rates that are proportionate to size. This convergence confirms previous studies showing that growth rates are independent of size 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.

Two measurement problems that we explicitly address are worth highlighting as they illustrate the empirical pitfalls in estimating the relationship of interest. First, the estimated pattern is highly sensitive to the identification of truly new firms, which we call *de novo* entrants, in the data. Large-scale firm-level datasets are typically contaminated by spurious entry and exit, due to incumbent firms changing administrative ID code or restructuring. It is widely recognized that such events lead to an overestimation of firm and employment dynamics (Haltiwanger et al., 2013; Geurts, 2016). The bias they introduce in post-entry growth patterns has received less attention. We make use of two state of the art record linking methods to minimize these problems. They enable us to accurately identify *de novo* entrants as new firms starting new operations, and to trace their complete histories from the moment they hire their first employee till they cease activities, *i.e.* true economic exit.

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. Furthermore, we show that failing to identify even a small fraction of entrants as spurious has major implications for the estimated post-entry growth patterns.

Second, it is well-known that regression to the mean as well as sample selection may bias the relationship between size and growth for surviving firms. In particular, they 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 are not that important for larger firms, the statistical side-effects of the base-year size classification are greatly exacerbated in a sample of small firms, such as our sample of *de novo* entrants. To avoid bias in the size–growth relationship, we use three alternative firm-size classifications and find a robust positive size–growth relationship for each of them.

Both the concentrated entry distribution and the subsequent growth patterns that we observe are consistent with an augmented passive learning model. The positive size–growth relationship for very young firms can be rationalized within the Jovanovic (1982) framework if 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 (Cabral and Mata, 2003; Beck et al., 2006). A similar delay before weak performers exit the industry will tend to reduce the growth rate of small firms and contribute to the observed positive relationship.[3] As firms mature and gradually learn their true efficiency, additional information becomes less informative and firm size converges to the optimal steady state level.

The remainder of the paper is organized as follows. Section 2 starts with a brief discussion of firm entry in theoretical models and previous empirical findings. It also reviews predictions of the model of Jovanovic (1982) and the stylized facts on entry and post-entry dynamics consistent with this model. 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 measurement of firm size are discussed. The results are presented in Section 5, first showing the size distribution of *de novo* entrants, followed by the post-entry growth patterns by age and size. Section 6 discusses how these patterns may be explained by delayed adjustment, and explores some alternative interpretations. Section 7 concludes.

2. Facts and theory

2.1. 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 for their industry. Since entrants know the population distribution, they have the same prior beliefs and all enter at the same size.[4] Firms gradually discover their own efficiency level from operating in the market. This leads to divergence in firm size as a cohort matures.

This approach contrasts with models that incorporate heterogeneity among firm size already at startup. Lucas (1978) 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. In the model of Hopenhayn (1992), firms also receive a random productivity draw from a known 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 at a similar

[3] 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.

[4] 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.

scale, but entrants are typically observed over a broad range of size classes.[5] Deviations might simply be due to the stylized assumptions of the model, but two measurement issues help explain the discrepancy between the prediction and empirical observations. First, the group of entrants in administrative datasets typically includes some established firms that merely re-register with a new identification code. A new ID code may for example be assigned after a change of ownership or legal form, when firms merge or divest, or for other reasons discussed below. These other modes of entry are certainly economically relevant, but differ from the concept of genuine entry assumed in the models above. We label them as spurious entrants, as opposed to *de novo* entrants which we study in this paper. Several studies have demonstrated that spurious entrants fundamentally differ from *de novo* entrants (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 good 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.

Second, some variation in entry size can reflect selection and growth effects occurring between the moment a firm is established and the first time it is observed in the dataset. In many administrative datasets like the one we use, new firms are observed in the first year they record positive employment on a fixed day in the year. 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 already have had some chance to learn about their innate productivity and choose a different size in response. Alternatively, initial size differences can reflect some prior knowledge entrants have about their ability even before they start as a corporation and hire their first employee. Helfat and Lieberman (2002) show that even *de novo* entrants may possess some knowledge about their resources and intrinsic quality which affects both entry decisions and subsequent success. In either case, the observed entry size distribution should (at least partly) be regarded as the outcome of an initial selection and size adjustment process. To highlight this, we will denote the year of entry in the dataset as age 1, and the unknown moment of the firm's establishment as age 0.

2.2. How do firms grow after entry?

Empirical studies for several countries have documented a number of regularities in the survival and growth patterns of young firms.[6] Young firms exhibit high failure rates

[5] For example, the Business Dynamics Statistics of the U.S. Census Bureau shows entry sizes between 1 and +2500 employees (Haltiwanger et al., 2013, Table 1). The Research Data on Business Employment Dynamics of the U.S. Bureau of Labor Statistics shows entry sizes between 1 and +500 employees, see http://www.bls.gov/bdm/business-employment-dynamics-data-by-age-and-size.htm

[6] See for example Evans (1987a) and Dunne et al. (1989) for U.S. manufacturing plants, Haltiwanger et al. (2013) for U.S. manufacturing and services; Mata et al. (1995) and Mata and Portugal (2004) for Portugal. Overviews are provided by Siegfried and Evans (1994), Geroski (1995), and Caves (1998).

which strongly decline with age. Among firms of the same entry cohort, exit rates decrease with firm size. Surviving firms exhibit high growth rates in the early years after entry, but growth slows down rapidly with age. These post-entry dynamics lead to a rapid increase in concentration in a given entry cohort. A typical pattern is that 5–10 years after entry, average firm size has doubled, but only half of an entry cohort survives.[7] As shown below, we find the same regularities in the sample of *de novo* entrants in the Belgian private sector.

The stylized patterns of post-entry growth and exit rates suggest a stochastic process in which firms make their entry decision uncertain of their success and do not enter immediately at their optimal size. Jovanovic's passive learning model (Jovanovic, 1982) has been widely used as a fruitful explanation for these empirical regularities. [8] In the model, firms only discover their own innate efficiency level from operating in the market. Initially, they have the same beliefs about this and all enter at the same size. Realized profits depend on a firm's actual efficiency and idiosyncratic cost shocks. Each period a firm observes its profitability, uses Bayes' rule to update its beliefs about its own productivity level, and expands or contracts to the appropriate size given the new beliefs. Firms that gradually discover they are highly 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' efficiency. As firms mature and gradually learn their true efficiency, additional information becomes less informative and firm sizes converge to a steady state.[9]

The model generates testable predictions about exit and growth patterns in relation to the firm's age and size, which are in line with the stylized facts described above. First, the noisy selection process implies an inverse relationship 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.

[7] See for example Dunne et al. (1988) for the U.S., Wagner (1994) and Boeri and Cramer (1992) for Germany; Mata et al. (1995) for Portugal.

[8] The active learning model of Ericson and Pakes (1995) provides an alternative explanation. In that model, growth is a function of firms' actions as they can make investments to raise productivity.

[9] 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.

Third, because smaller firms are on average younger and young firms grow faster, the model also predicts an inverse relationship between growth rates and size in a cross-section of firms that encompasses all age cohorts. Several empirical studies find evidence for this inverse relationship and Jovanovic (1982) cites it as a key motivation for the model.

Without additional assumptions, the model does not imply a systematic relationship between growth rates and size among firms of the same age. Only when the cost function is assumed to take the Cobb–Douglas form does the model predict that growth rates for mature firms are independent of firm size, consistent with Gibrat's law.[10] At each point, a firm's size reflects its best estimate of its efficiency. Subsequent adjustments depend only on future information which is by definition random. The existing evidence on which pattern prevails in the early post-entry process has been inconclusive. Earlier studies found an inverse relationship between growth rates and size of young firms, which suggests that smaller firms to some extent catch up in size with larger ones, but recent evidence by Haltiwanger et al. (2013) does not confirm this pattern.

One set of 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 within this broad age class. Dunne et al. (1989) and Almus and Nerlinger (2000) find that smaller plants or firms grow faster than larger ones. Wagner (1994) finds growth rates to be independent of size.[11] 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.[12]

The few studies that have investigated post-entry growth *conditional on single-year age classes* 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 to diminish with age and converge towards growth that is proportionate with size for older firms. Mata (1994) finds a similar, but weaker negative relationship.[13] While these previous studies are limited to manufacturing firms, a recent study by Haltiwanger et al. (2013) which also includes the service sector, tells a different story. They report a negative as well as a positive pattern, depending on the size classification method. When using their preferred methodology, they find larger

[10] 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. However, 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.

[11] The three studies group together all firms younger than, respectively, 5, 6, or 10 years.

[12] 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.

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

firms to grow more rapidly than smaller ones among young survivors of the same age. Moreover, their results show no convergence towards size-invariant growth for older firms.

We discuss below that estimates of the size–growth relationship of young firms are sensitive to three elements, each of which may partly explain the contrasting results. First, an accurate identification *de novo* entrants matters greatly. Misclassifying older firms biases the observed pattern towards that of incumbents. Second, the potential negative bias in the size–growth relationship induced by sample selection and regression-to-the-mean is exacerbated in a sample of very small firms. Finally, the patterns can differ by industry, especially between manufacturing and services, reflecting different entry costs and post-entry size adjustments.

After presenting our empirical results, we will revisit the theoretical interpretation. The stylized framework of the Jovanovic model necessarily implies some simplifications and we will discuss two elements that are hard to reconcile with some patterns in the data. A first assumption is that adjustment to new information is instantaneous. When firms update their estimate of their own productivity, they can immediately adjust to the new optimal size. In reality, learning takes time and adjustments need to overcome frictions. A second assumption is that firms' innate productivity is constant over time and not under the firm's control. In active learning models like Ericson and Pakes (1995) firms not only make operational decisions given their current productivity, they also make investments to influence their future productivity. We will discuss some alternative models that could explain why the passive learning model fits less well for some sectors than for others.

3. Data

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

De novo entrants are defined as new firms starting new operations. We identify their point of entry as the year they hire their first employee. We distinguish them from spurious entrants, by making use of two state-of-the-art record linking methods. The linkage methods are further used to trace the complete histories of *de novo* entrants from the moment they start operating till they cease activities, *i.e.* true economic exit. For *de novo* entrants that change identification code or restructure in the years following entry, 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 summarized below and explained in full in Appendix A. We also show that the size range of *de novo* entrants differs dramatically from the size range at entry in 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.[14] Firms may change identification code after a legal change or a restructuring. Examples include ownership changes, legal changes for tax or liability reason, mergers, takeovers and divestitures, *etc.* 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 minimize these problems, we use two record linking methods cumulating the missing linkages we identify.

The first method 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 misclassified as exits and entrants in consecutive years. These 'spurious' exit and entry events differ from true economic exit and *de novo* entry, and are removed from the dataset by linking the appropriate firm-year observations. It is especially important to recognize that spurious entrants are pre-existing firms that exhibit characteristics similar to other incumbents. Failing to distinguish them from *de novo* entrants biases the size and growth patterns of young firms towards those of incumbents. Panel (b) of Table A.1 in Appendix A shows that 78% of the spurious entrants we identify are simply incumbents that continue with a new ID code after a purely administrative or legal change. Another 18% are split-offs of another firm.

Next, for *de novo* entrants involved in an ID change or restructuring in the years following entry, employment is imputed up to the sixth year after entry. For firms that simply change ID code, no imputation is needed. For more complex events, such as mergers or split-offs, we impute firm-employment in the following periods by assuming the same growth rate for each firm involved in the event. An important advantage of

[14] The problem that large-scale firm-level data suffer from spurious entry and exit due to administrative or legal changes, has long been recognized. However, only with the recent development of sophisticated record linkage methods has the extent of the problem and the profound impact on empirical results become clear. Most previous studies that examined the size–growth relationship for 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. Evans (1987a, 1987b) uses U.S. data from Dunn and Bradstreet which are known to suffer from data problems with respect to young and small firms (Davis et al., 1996b). Almus and Nerlinger (2000), Lotti et al. (2003) and Mata (2004) do not report the use of linkage methods to identify spurious entry and exit. Wagner (1994) recognizes that entry with a large firm size is unlikely and excludes the largest firms from the entry sample, ignoring spurious entrants in other size classes. Dunne et al. (1989), using the U.S. Census of Manufacturers, correct for ownership changes but not for other administrative changes or changes in firm structure. Only Haltiwanger et al. (2013) use a dataset edited by traditional record linking methods, additionally relying on physical addresses to more accurately identify entry and exit of multi-establishment firms. It is therefore not surprising that our results are more in line with that study.

Table 1
Share of *de novo* and spurious entrants in all administratively recorded entrants.

	Total	By firm size class					
		1–4	5–9	10–19	20–49	50–49	100+
Number of administrative entrants	*17,283*	*15,368*	*1209*	*446*	*190*	*39*	*32*
Share of *de novo* entrants	0.91	0.95	0.64	0.41	0.26	0.17	0.03
Share of spurious entrants, identified by:							
Both methods combined (total)	0.09	0.05	0.36	0.59	0.74	0.83	0.97
Traditional method	0.06	0.05	0.12	0.16	0.21	0.32	0.44
Employee-flow method	0.05	–	0.32	0.57	0.72	0.82	0.97

Note: annual averages over the sample period. Firm size classes are based on the number of employees.

this imputation procedure is that it preserves the firm size distribution in $t-1$ when calculating growth rates from $t-1$ to t.

In the first row of Table 1, we report the number of entrants in an average year of the sample period, as observed in the raw administrative data. The next rows presents the fraction of these firms that are identified as either *de novo* or spurious entrants. The last rows further illustrate 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.

Spurious entrants only represent 9% of all administratively recorded entrants, 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, accounting for more than one third of administrative entrants with 5–9 employees and even two thirds of those with 10 or more employees. *De novo* entrants with more than 50 employees are extremely rare. The presence of spurious entrants in the unedited data would introduce a bias in post-entry patterns by size.

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.[15]

Following Davis et al. (1996a), firm-level growth rates are calculated as discrete-time employment changes relative to the average of employment in years $t{-}1$ and t. Denoting employment of firm i in year t as E_{it}, the growth rate over the preceding year equals $g_{it} = (E_{it} - E_{it-1})/\bar{E}_{it}$, with $\bar{E}_{it} = (E_{it} + E_{it-1})/2$. These growth rates range from -2 for exits to $+2$ for entrants, show job creation and destruction symmetrically, and are bounded away from infinity.[16] 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,∝].[17] All observations with more than 32 employees are in the same size class because few *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 describe 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 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.

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:

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

[15] Some survivors have zero employment in a given year ('dormant' firms). We omit them from the regressions in those years.

[16] This growth rate is close to the more commonly used logarithmic growth rate $g_{it} = \ln\,(E_{it}/E_{it-1})$ for values around 0. 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 can fluctuate 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.

[17] Due to the use of average employment and imputed employment levels, size is a continuous variable.

where the dummy variable $1[age_{it}=j]$ 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_{it}^{d} enter both additively and interacted with the full set of age–size interactions. As we impose that $\sum_d \beta_{jk}^{d}=0$, the average effect of age and size on growth and exit is captured by the uninteracted α_{jk} coefficients, while the β_{jk}^{d} coefficients allow for heterogeneity across industries. The additive year dummies γ_t control for business cycle effects.

4.1. Size classification of surviving firms

We approalternative approaches to allocate surviving firms into 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 unbounded. It especially affects smaller firms which will already exit when hit with a moderate negative shock, leading 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) 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, like 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 the 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.[18] Instead, we use three alternatives to allocate firms in a given size class. The objective is to approximate firm growth in continuous time and we refer to the 'current' size of the firm. A more detailed discussion of the econometric problems and the classification methodologies we use is in Appendix B; here we provide a brief overview.

The first size classification method, and the one we use for the 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 reassigned 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

[18] 94% of *de novo* entrants have fewer than 5 employees at age 2 and still 82% at age 6.

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 et al., 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.[19]

5. Results

A first set of results confirms that the well-established patterns of post-entry growth and exit rates discussed above also hold for our sample of *de novo* entrants.[20] We summarize them here, but only show them in Appendix C, see Table A.2 and Fig. A.1, as they mainly serve to illustrate that the novel findings discussed below are not an artifact of the Belgian dataset.

Young firms exhibit high exit rates, which are decreasing in age and within every age cohort also decreasing in firm size. The selection process of the passive learning model – which predicts market exit of the least efficient and therefore the smallest firms – appears to unfold 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 have 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. Only half of all entrants survive to age 6, at which time the average firm size in the surviving group has almost doubled. Appendix C contains a more detailed discussion of these patterns.

[19] The weights we use in the growth regressions follow are adjusted for 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) $\bar{E}_{it}$.

[20] These patterns are in line with previous results even as often little attempt has been made to distinguish between *de novo* and spurious entrants. It suggests that most patterns are robust to less accurate identification of truly new and young firms.

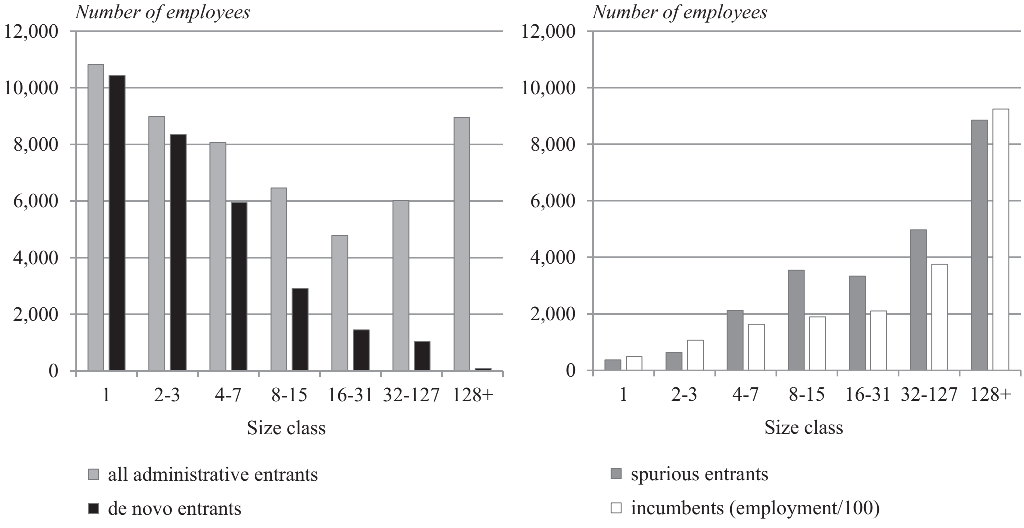

Fig. 1. Employment distribution of entrants.
Note: total employment of entrants and incumbents by firm size class. Annual averages over the sample period.

We find two novel patterns that we discuss in detail below. First, *de novo* entry is confined to a much narrower range of small size classes than usually found. Second, growth rates of surviving young firms of the same age cohort are increasing with firm size. This pattern rapidly converges to growth rates that are proportional to size when a cohort matures.

5.1. Entry distribution

Although summary statistics based on all administrative entrants or limited to the set of *de novo* entrants look very much alike (Table A.2), a closer examination of both samples reveals some fundamental differences. This is because spurious entrants – pre-existing firms that 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 *de novo* entrants correctly is readily seen from the employment distribution at entry by firm size class. Fig. 1 shows total employment of entrants by 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.

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 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 right panel shows the employment distribution of spurious entrants (dark) relative to that of incumbents (light). 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 discussed, the sample of administrative entrants that uses untreated firm-data mixes two distinct populations of firms. Failing to distinguish between them has two implications. First, given that spurious entrants account for 44% of total employment in the sample of administrative entrants, they lead to an inflated impression of the importance of new firms for job creation. 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 2.6%, 1.8 times higher.[21]

Second, the size distribution of administrative entrants has a much more dispersed shape than the strong right-skew we observe for *de novo* start-ups. The focus on *de novo* entrants shrinks the firm entry sizes 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.

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-exit and size–growth pattern for entrants. The bias is barely noticeable in exit probabilities by size, see Fig. A.2 in Appendix D, since exit rates are decreasing in size both for young and older firms. In contrast, the bias is pronounced for growth estimates by size where young and older firms strongly differ. This is the topic we turn to next.

5.2. Post-entry growth

In a cross-section of firms of all ages, growth rates of firms that survive from year $t-1$ to t decline monotonically with the current size of the firm. Such a relationship has often been documented in the literature and it is also what we find for *de novo* entrants Belgium, as shown by the 'all firms' line in Fig. 2. As young firms have much higher growth rates and are overrepresented in smaller size classes, sample composition alone could generate such a negative relationship between growth and size in a cross-section of firms.

[21] Geurts (2016) compares job creation rates by *de novo* entrants with official rates reported for OECD countries.

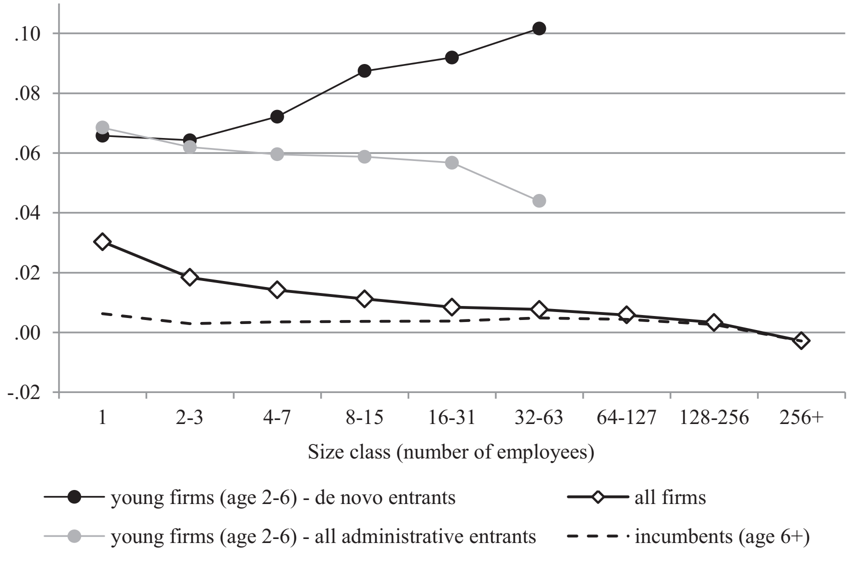

Fig. 2. Growth rates of surviving firms by size: young firms versus incumbents.
Note: the figure plots the coefficients on the firm-size category dummies (using the dynamic size classification) from the employment growth regressions for surviving firms, without including age dummies and estimated separately for each sample. For young firms the 32–63 size class is really 32+.

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 Fig. 2 shows that incumbents on average only achieve low growth rates regardless of firm size. For them, absolute employment growth is proportional to their current size, confirming an empirical regularity that has often been found for mature firms. In contrast, the two lines at the top show not only much higher growth rates for young firms, the relationship with firm size clearly depends on the group of firms we consider.

Young *de novo* firms show a positive relationship between growth and size (black line), while the raw sample of administrative entrants suggests that growth rates for young firms are decreasing with size (gray line). Note that either pattern is consistent with the negative size–growth relationship for all firms as the share of young firms in the population decreases strongly with size.[22] The difference between *de novo* and administrative entrants is even more pronounced when growth rates are estimated conditional on age, as discussed below.

5.3. 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 Fig. 3. We then assess the robustness of the pattern in Fig. 4, illustrate the sensitivity to measurement bias in Fig. 5, and compare the pattern for different industries in Fig. 6.

[22] Fig. A.3 in Appendix D shows almost identical patterns for incumbents and all firms when calculated using the raw administrative data.

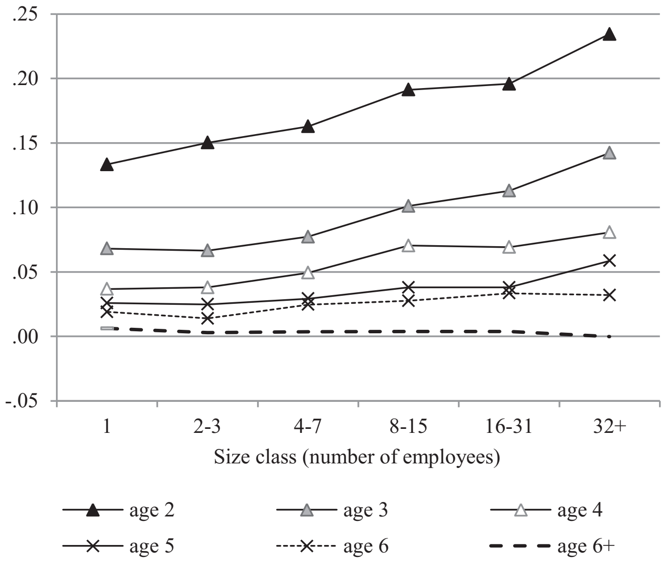

Fig. 3. Growth rates of surviving *de novo* entrants by age and size.
Note: the figure plots the coefficients on age–size category interaction dummies from the employment growth regressions for surviving firms using the dynamic size classification. Point estimates and standard errors are reported in Table A.3. Age 6+ refers to incumbents.

Fig. 3 plots the coefficients from the employment growth regression of *de novo* entrants that survive from period $t-1$ to t. Due to the employment weights, they represent the net employment growth rates of the entire group of survivors within each age–size class. These benchmark results use the dynamic size classification to assign firms to a size class, while results using two alternative classification methods follow below. For clarity, we do not show confidence bounds but report all coefficient estimates and standard errors in Table A.3 in Appendix D. Coefficients are estimated extremely precisely and almost all point estimates for successive age or size categories 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 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 growth rates of incumbents (labeled age 6+) has not been completed entirely when entrants reach age 6, *i.e.* when we have observed them for five years.

The more remarkable pattern in Fig. 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 be close to zero in all size classes. For the smallest firms, growth has basically stalled after five years while for larger firms growth will remain positive for a few years longer.

The unique growth pattern we observe is confined to the very first years after entry and rapidly converges to growth rates that are proportional to the current size of the firm. This contrasts with the exit probabilities, which are inversely related to size even for older cohorts. Convergence to size-invariant growth for older firms is consistent with many studies that have found that Gibrat's law is a good approximation of the size–growth relationship in a sample of older firms or among firms that have exhausted scale economies (Mansfield, 1962; Hall, 1987; Geroski, 1995).

Empirical studies have also documented how selective survival, *i.e.* higher exit rates for smaller firms, leads to a rapid increase in concentration in a given entry cohort. The positive size–growth relationship that we document for surviving entrants will contribute to this process. Firms that expand in the very first years of their life cycle are likely to grow faster in subsequent years, while firms that do not expand early on are more likely to remain small.

Although specific to the Belgian economy in the 2003–2012 period, it is instructive to consider how the firm-level patterns translate into aggregate job creation. The average cohort of *de novo* entrants represent approximately 30,000 new jobs or 1.5% of total private employment. Five years after entry, about half of all entrants have failed, while only 1% of entrants have expanded beyond 20 employees. Total employment of an entry cohort has dropped below its initial level of 30,000 jobs by age 6 and is still mainly located among the smallest firms. Fast-growing firms that expanded beyond 20 employees represent 20% of these jobs, which is a disproportionate, but still only a modest share.

5.4. 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. 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 Fig. 4 estimates based on two alternative size classifications which also approximate a continuous size–growth relationship. The results in panel (a) 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.

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.

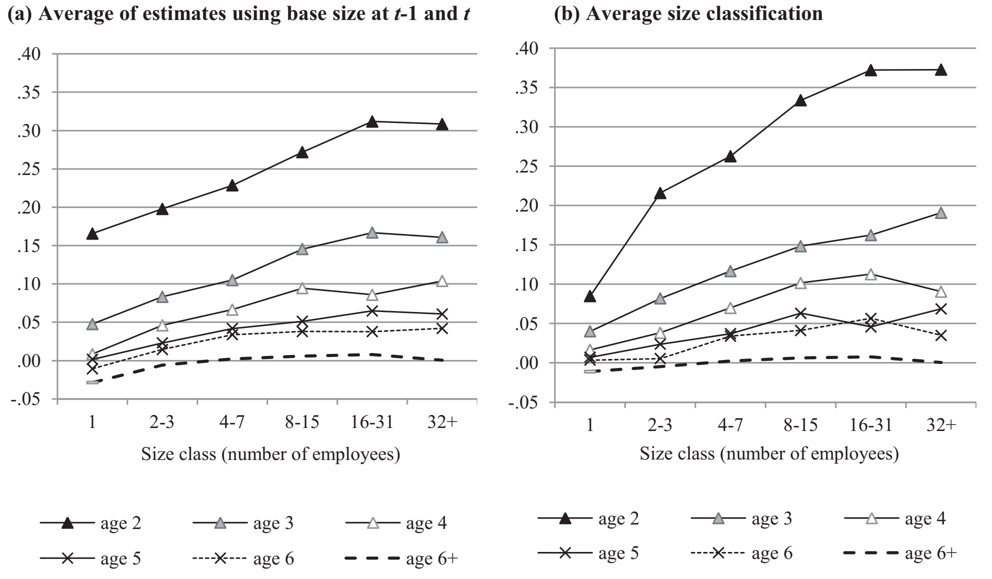

Fig. 4. Alternative size classifications: growth rates of surviving *de novo* entrants by age and size.
Note: the figure plots the coefficients on age–size category interaction dummies from the employment growth regressions for surviving firms using two alternative methods to classify firms by size. Point estimates and standard errors are reported in Table A.3. Age 6+ refers to incumbents.

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 are 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.[23]

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. An important difference, however, is that the growth rates they report do not evolve to size-invariant growth as firms mature, but remain positively related to firm size also among older cohorts. Such a persistent positive size–growth pattern cannot be a steady state as the firm size distribution would collapse.

5.5. A relationship sensitive to measurement problems

The relationship between growth and size among young firms of the same age cohort is highly sensitive to the identification of new firms in the data and to the use of the firm size classification. Fig. 5 shows how inappropriate measurement on either dimension induces a negative bias in the relationship.

[23] Separate results for pre and post-crisis entrants are shown in Geurts and Van Biesebroeck (2014).

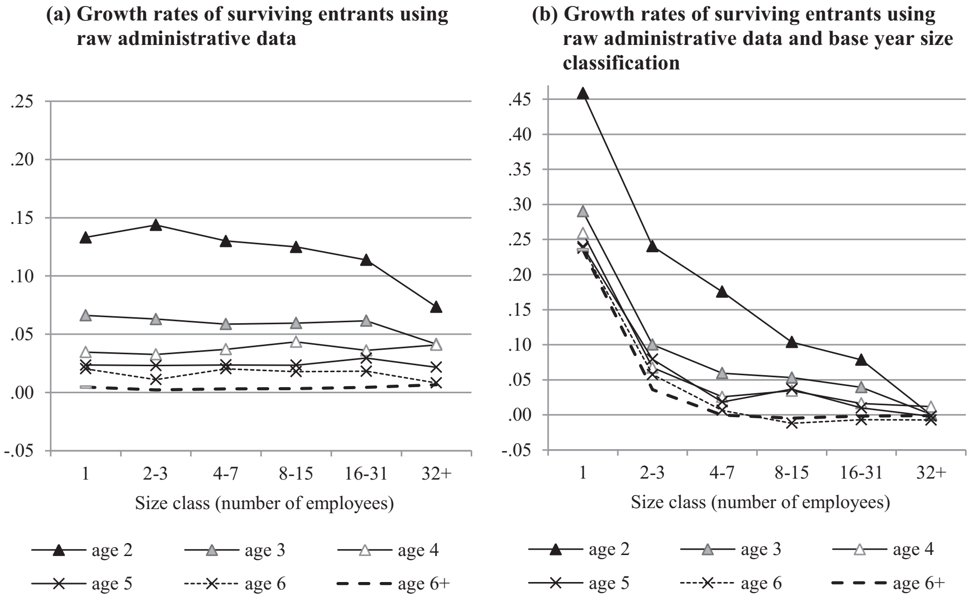

Fig. 5. The sensitivity of growth rates of young firms by age and size to measurement bias.
Note: the figure plots the coefficients on age–size category interaction dummies from the employment growth regressions for surviving firms. Estimates in both panels are based on the sample of all administratively recorded entrants, including spurious entrants. In panel (a) we use the dynamic size classification; in panel (b) we use the base year size in $t-1$ to classify firms by size. Age 6+ refers to incumbents.

In many applications on administrative data, spurious entrants are not adequately filtered out from sample. As they are misclassified older firms, their growth rates tend to be much lower, resembling those of incumbents.[24] It introduces a downward bias in post-entry growth rates because spurious entrants dominate in larger size classes. This effect is illustrated in panel (a) of Fig. 5 which replicates Fig. 3 on the raw sample of administrative entrants. The downward bias is barely noticeable in the smallest size classes where the share of spurious entrants is negligible. But 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 appear to be size invariant already from age 3 onwards.[25]

Panel (b) of Fig. 5 is again based on the raw administrative sample, but uses a conventional base-year size classification which measures firm size at the start of the period $(t-1)$ over which growth rates are calculated.[26] The estimated relationship is now

[24] Fig. A.4 in Appendix D shows growth rates separately for spurious entrants, which highlights their uniformly low growth rates.
[25] Imputing employment for *de novo* entrants that change ID code or restructure in one of the years after entry has little impact on the estimated size–growth relationship, see Fig. A.5 in Appendix D. Ignoring such events, which would misclassify them as exits, would makes the positive relationship slightly more pronounced.
[26] Fig. A.6 in Appendix D shows the comparable results for the sample of *de novo* entrants.

strongly negative in the first two years after entry, and also at older ages in the smallest size classes. Regression-to-the-mean and sample selection bias spuriously induce a negative relation between growth and size when the explanatory variable is size at the start of the period. While Hall (1987) found that these measurement issues are relatively unimportant in a sample of publicly traded (large) firms, the statistical biases are exacerbated in our population of young firms, which are predominantly small and exhibit large variation in growth rates.

In small firm size classes, even the smallest discrete employment changes will imply large positive or negative growth rates and easily induce firms to cross size-class borders. Random variation due to measurement error or transitory fluctuations will thus spuriously lead to a negative relationship between growth and size. The left-truncation of the range of possible growth rates of surviving firms is also concentrated in the smallest size classes. It is especially important for surviving one-employee firms which cannot even have negative growth rates. Given that the smallest size classes represent the vast majority of young firms, mean growth rates estimated from a regression on relatively broad size classes or on last period's firm size are highly sensitive to the biases induced by these small firms.

Some studies use a base year classification but use a different approach to control for potential bias. For example, Mata (1994) omits all firms that enter with fewer than 10 employees to avoid truncated growth rates of the smallest firms. In our sample, this would necessity the exclusion of 98.5% of all *de novo* entrants, leaving an unrepresentative sample of the total population of new firms. Evans (1987a) and Lotti et al. (2003) include entrants of all sizes and find an inverse size–growth relationship for young firms conditional on age. It is unclear to what extent their alternative estimation techniques are sensitive to the biases we discussed, but their lack of control for spurious entry is certainly problematic.

5.6. By industry

Finally, we verify whether the positive relationship between growth and size holds in all sectors. Our empirical specification allowed for heterogeneity across sectors in average growth rates and for the pattern by age and size, but Figs. 3–5 only showed the average pattern. In Fig. 6, we show separate results for the six industries groups: manufacturing, construction, wholesale and retail trade, accommodation and food services, business support services, and mixed household and business services. Growth rates of surviving *de novo* entrants in most industries show broadly the same pattern as in Fig. 3. They are high in the first year, but decrease quickly with age within each size class. Only in accommodation and food services there appears to be less room for size diversification after entry. The average firm size at entry is relatively small, and from age 3 onwards growth rates are uniformly close to zero. In all other industries, growth rates are increasing with size in the first year after entry, while for older cohorts the pattern moves to a more proportional distribution.

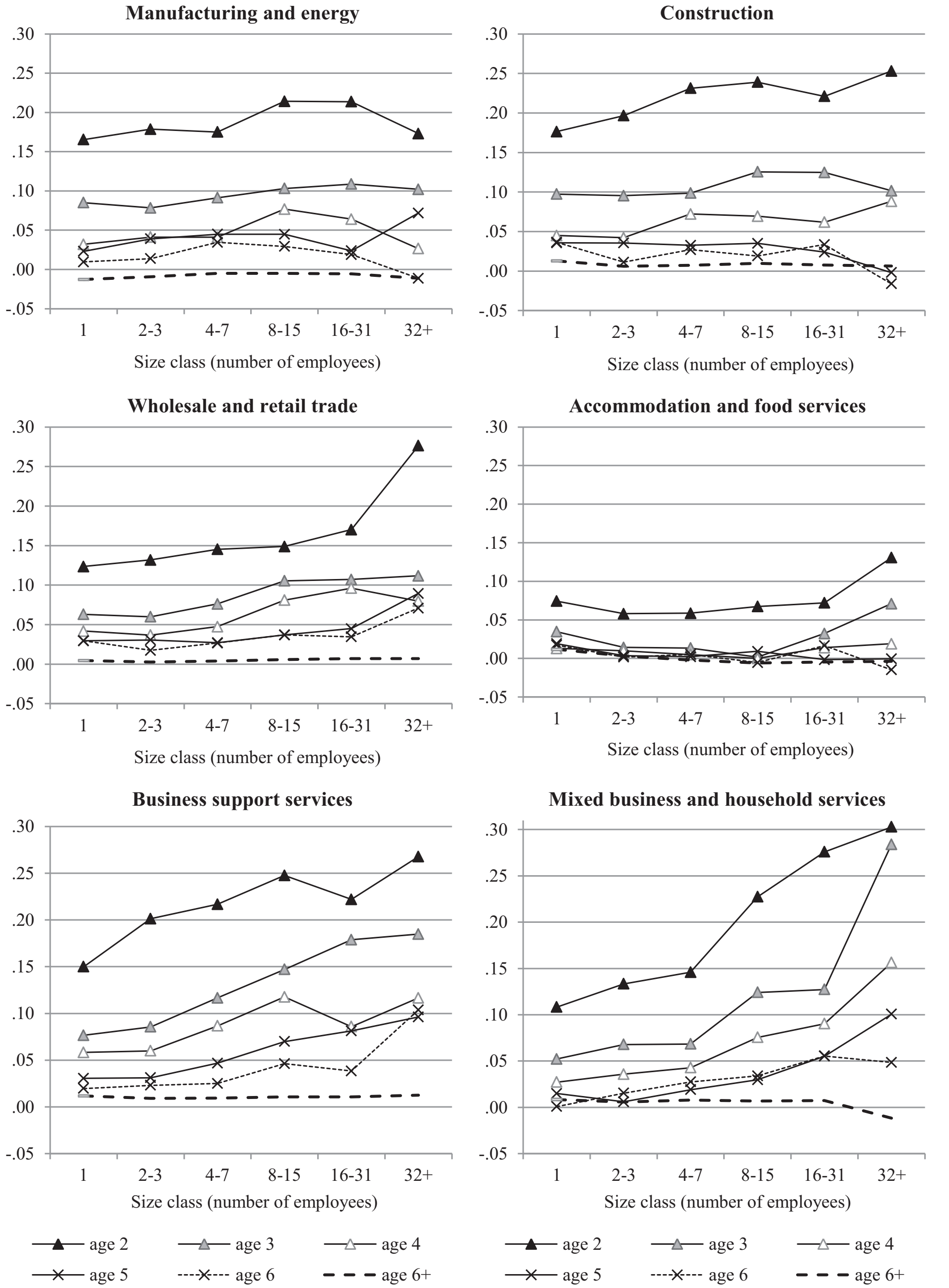

Fig. 6. Growth rates of surviving *de novo* entrants in six industry groups.
Note: the figure plots the coefficients on age–size category interaction dummies from the employment growth regressions for surviving firms, using the dynamic size classification. The model is estimated separately for each industry group. Age 6+ refers to incumbents.

The increasing relationship is most pronounced in trade, business services, and the last category of mixed household and business services. We expect entry costs to be lower for these activities and firms can more easily enter at a small size and gradually adjust to an optimal scale. In the three sectors, growth rates are monotonically increasing with size at every single age, and the convergence to low growth rates that are proportional to firm size has not been completed by age 6. The pattern relative to that of older firms suggests that smaller firms have approximated their steady state growth rates by age 6, but larger firms will continue to grow faster for some more years.

The increasing pattern is least pronounced in manufacturing. Consistent with a higher minimum efficient scale in manufacturing, we find higher average size at entry and a negative size–growth relation for size classes above 16 employees in most age cohorts. The results for this sector differ less from many of the previous studies that were based on samples of young firms in manufacturing only, and found a negative size–growth relationship (Evans, 1987a; Lotti et al., 2003; Mata, 1994).

6. Discussion

The narrow size range at entry and the exit and growth patterns we observe in the data are consistent with the predictions of the passive learning model of Jovanovic (1982). The size–growth relation among young firms of the same entry cohort is, however, not defined in that model. Below we argue that the positive relationship we find for young firms is not at odds with the predictions of the model if one takes into account that firms need some time to adjust their size to the new information. Next, we explore some alternative explanations and discuss why the passive learning model fits less well for some sectors than for others.

In the stylized framework of the Jovanovic model, a firm's current size fully reflects all past information regarding its own ability. 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 ratios of factor prices and marginal productivities and between observed and optimal output levels are widespread. Asker et al. (2014) have shown that these deviations partially reflect the dynamic adjustment of quasi-fixed production factors.

Similarly, some of the young firms in our sample will not be able to immediately adjust to their desired size when they revise the estimate of their innate efficiency. External constraints can impose barriers that need to be overcome before a firm can expand its operations. Or firms that learn they are more efficient than previously realized might choose not to adjust completely to this new information right away. For some expanding firms, current size will be below desired size until all positive information is incorporated into their size. Some of the positive news leads to instantaneous growth and raises a firm's current size. Some of the growth rate is postponed to future years when frictions are overcome or the positive news is confirmed. Such partially delayed growth will induce a positive correlation between past and current growth, and thus between firm size and growth.[27]

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 simply hang around for the business cycle to improve rather than exit, or continue operations before eventually deciding to withdraw from the market, even when they make losses. In the administrative data set we even observe some firms with no employees for a few years. We omitted them from the analysis, but it suggests that merely surviving might not be all that costly. As firms gradually adjust their size downward but postpone exit, it leads to low or negative growth rates for smaller firms.

Fig. A.7 in Appendix D provides some evidence for this 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 similar for firms of different ages. 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 gray curve indicates that the difference in growth rates already appears two years before exit, 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. Panel (b) in the same figure shows that delayed exit does not explain the observed positive size–growth relationship entirely. Excluding all *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 convergences to a size-invariant pattern.

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 et al., 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 et al. (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. It takes time for additional capacity to become operational. Risk aversion can contribute to a pattern of gradual adjustment as it might induce firms to wait an extra period for the positive information to be confirmed. 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.

Delayed adjustment due to growth constraints encountered after entry can rationalize the observed size–growth pattern. However, frictions could also influence firms' initial

[27] In a Markov perfect equilibrium, the value of current state variables are sufficient statistics for the entire firm history (Ericson and Pakes, 1995). With adjustment frictions this is not necessarily the case anymore.

entry size. A prominent example is the model of Evans and Jovanovic (1989) where heterogeneity of firm size at startup reflects liquidity constraints. Some constrained firms cannot finance their optimal level of capital and are forced to enter below their desired size and gradually grow into their optimal size using retained earnings to expand. If the presence of the initial friction is uncorrelated with ability, the most constrained firms will enter at a very small size and have the largest gap with their optimal size and thus the highest growth potential. It would lead to a negative relation between initial size and subsequent growth, as in Audretsch et al. (1999). Cabral (1995) contains an alternative mechanism that generates the same prediction. If production capacity requires substantial sunk costs that are foregone when firms exit, smaller firms that are more likely to exit will choose to invest gradually and enter at even smaller scale. Both the limited size dispersion for *de novo* entrants and the positive size–growth relationship post entry are inconsistent with the predictions of these models.

Both the limited size heterogeneity at entry and the subsequent positive size–growth relation are consistent with the presence of initial credit constraints if these are strongly correlated with firms' efficiency.[28] In Evans and Jovanovic (1989), entrepreneurs know their own ability, but lenders limit credit to a fixed proportion of the firms' assets. However, if lenders observe even before startup a noisy signal that is correlated with a firm's efficiency, *e.g.* as in Bonnet and Cressy (2016), more efficient entrants will receive higher financing, although not as much as they need to enter at their optimal size. They will enter with an above average size and display high growth rates as they gradually expand into their optimal size, using either retained earnings to finance expansion or additional external financing as the accuracy of the lenders' signals improve with operations.

This mechanism can also generate a positive relationship between size and growth among young firms. Note that, somewhat counterintuitively, the largest entrants are most constrained in this situation. Earlier, we discussed that the observed size heterogeneity at entry can be regarded as the result of early selection and growth effects occurring between the unknown moment of the firm's genuine startup (denoted as age 0) and the first time we observe it in the dataset (age 1). The correlated credit constraints we just discussed constrain the firm at age 1. For some firms, it will be a post-entry constraint as they learned more about their ability than the external lender from operating between ages 0 and 1 and the financing friction delays expansion. Other firms might have learned their own efficiency before startup and the inability to communicate this information credibly to the lender is a genuine entry constraint.

We have used the passive learning model of Jovanovic (1982) as the main framework for interpreting our results. Heterogeneous firm models along the lines of Hopenhayn (1992) do not incorporate firm-specific stochastic elements that give rise to systematic heterogeneity in growth rates. In those models firms enter immediately at their optimal size and later adjustments are responses to random productivity shocks firms have no control over. Abbring and Campbell (2005) add persistence in post-entry shocks to the

[28] We thank an anonymous referee for this alternative explanation.

model which does lead to serial correlation in growth rates. It could induce a positive size–growth relationship, but this would hold almost by construction.

A different class of models that feature heterogeneous firms are based on the active learning model of Ericson and Pakes (1995). A firm's optimal size is again determined by its state variable, but firms can undertake investments to raise their efficiency or productivity. Depending on the exact specification of the primitives of technology and the market environment, it might be the case that only high-productivity firms find it profitable to make positive investments leading to a positive correlation in firm size and growth rates. To finance such fixed cost investments, these models generally assume the industry is oligopolistic, such that firms make (variable) economic profits, but it also makes the optimal firm size a function of the state variables of competitors. The difficulty of solving and estimating these dynamic games has led to few empirical applications.[29] As the positive size–growth relationship we documented is limited to very young firms and the gradual adjustment to a steady state size unfolds very rapidly in the very first years of operation, we preferred the passive learning model. To understand growth among incumbents, active learning is likely to be more important.

Different models of firm behavior may be more appropriate to describe the dynamics in different industries. Pakes and Ericson (1998) find that the passive learning model is more consistent with services, while active learning fits better for manufacturing firms. Our results are suggestive of a similar distinction. As shown in Fig. 6, the positive size–growth relationship is most pronounced in wholesale and retail trade, business support services and Household services. These sectors are characterized by entry of predominantly small firms and high exit rates of young firms. Low entry costs lead to a large number of entrants which may have very little prior knowledge about their profitability and a high risk of early failure. Survivors need several years to form a precise posterior about the efficiency level they are endowed with, and gradually adjust to an optimal scale. Active learning is likely to play less of a role in these sectors, where R&D investment is low and firms are predominantly oriented to local markets (Kuusisto 2008).

This contrasts with manufacturing. Both learning prior to entry and active learning post-entry may explain why the positive size–growth relationship for young firms is less pronounced. Consistent with higher entry costs and a higher minimum efficient scale, we find firms to enter at a higher average size and exhibit lower exit and growth rates. It suggests that the moment of entry we observe is already the result of a prior selection process in which firms have gained knowledge about their efficiency. It might also be the case that entering at a small scale is often not an option as key machinery is an indivisible input. Moreover, productivity-enhancing activities along the lines of the active learning model of Ericson and Pakes (1995) have been studied extensively for manufacturing firms. In this model the information firms received about their productivity from prior

[29] Aw et al. (2011) for Taiwanese electronics firms and Hashmi and Van Biesebroeck (2016) for the global automotive industry are two examples. Both papers focus on recovering dynamic parameters of a structural model, respectively the fixed and sunk costs of exporting and investing and the R&D costs of innovation. Adjustments in firm size contribute to the identification of these structural parameters, but are not studied explicitly.

observations will erode as time passes. Past values of firm size should have no long-lasting effect on future firm size, consistent with some of the evidence in Pakes and Ericson (1998).

Finally, several studies have highlighted that the growth-oriented firm founder assumed in models of firm dynamics, who is driven by growing sales and profit expectations, is not the only type of entrepreneur (see Vivarelli and Audretsch, 1998; Hurst and Pugsley, 2011). Surveys reveal that many new entrepreneurs have low growth ambitions and are motivated by non-economic factors such as social status or the desire to be independent. This type of entry aligns more with self-employment theory, where new firm formation is modeled as the choice of an individual between employment and self-employment, based on his degree of risk aversion and expected earnings in either situation (Parker, 1996; Kihlstrom and Laffont, 1979). The growth pattern we observe in accommodation and food services, with growth rates close to zero from age 2 onwards, hints at the predominance of this type of entrepreneur.

Santarelli and Vivarelli (2007) point out that models of passive or active learning are consistent with a world where firm founders are heterogeneous in terms of capabilities and motivations. Market selection will effectively single out growth-oriented entrepreneurs and initial motivation or entrepreneurial 'taste' can be considered an indication of the firm's innate productivity and hence as a predictor of post-entry performance. The positive size–growth relationship we observe is consistent with the prevalence of different types of entrepreneurs in other sectors as well. As an entry cohort matures and firms reveal their type, entrepreneurs with low growth ambitions will become dominant in the smallest size categories, while more ambitious and productive firms will grow.

7. Conclusion

Constructing the dataset, we have taken great care to identify a sample of firms that start new operations, corresponding to actual new firm creation. Combining different firm linkage methods, we filtered out older firms misclassified as entrants to avoid biasing the patterns of interest. For the remaining group of *de novo* entrants, we confirm several patterns from the literature. In particular, exit rates are strongly declining with age and size; growth rates for survivors decline with age and also with size if we pool across age cohorts.

More importantly, we document two novel findings. First, we find that firm entry sizes are confined to a narrow range of small size classes. Second, growth rates of *de novo* entrants are increasing with size in the first years and then converge to size-invariant growth as an entry cohort matures. Our firm size distribution at entry differs more markedly the distribution of mature firms than is usually the case in large-scale datasets, but the positive size–growth pattern accelerates the tendency towards increased concentration in an entry cohort.

All entry, exit and growth patterns are remarkably regular. We estimated them over a 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.

These patterns are consistent with firms having an imperfect knowledge of their productivity at entry, as in the passive learning model of Jovanovic (1982). It assumes that firms have a constant efficiency level, but they only discover this from operating in the market. While not all firms can freely choose their size – *e.g.*, a large literature documents financial constraints at entry – the post-entry growth patterns suggest that by and large small firms choose to be small. 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. In continuous time, one can think of firm entry as the moment the first employee is hired. Some entrants add additional employees right away, while others take years. A large size today means that a firm previously received positive news about its profitability and expanded in response. With adjustment frictions, some of this expansion might spill over into the next period and generate a positive size–growth relationship.

In terms of policy conclusions, a few elements are worth emphasizing. Acs et al. (2016) highlight that many public policies subsidize individuals to become entrepreneurs and Guner et al. (2008) provide evidence that government policies often favor small firms. In Europe, this policy focus has recently been strengthened by the Entrepreneurship 2020 Action Plan, which sets out a number of initiatives to encourage the creation of new businesses and support growth of existing entrepreneurs (European Commission, 2013). These policies are rationalized on the assumption that new and small firms are the engine of job creation in the economy. While this may be true for the Schumpeterian innovative entrepreneur, most new firms do not fit this type. Previous literature has already highlighted that one should not confuse the (conditional) effects of age and size: it tends to be young, not small firms which are vital for job creation. Our findings highlight that job creation by startups is much lower than commonly believed on the basis of official statistics, casting further doubt on the employment growth potential of small entrants.

A recent literature has documented that especially in less developed economies, production factors are often stuck at unproductive firms (Hsieh and Klenow, 2009). This misallocation lowers potential output and aggregate productivity. If new firms do not know their own likelihood of success very well, it is inevitable that some unproductive entrants end up at least temporarily with too much resources. Policies that encourage new firms may support the entry and survival of more inefficient firms and add to market distortions rather than resolving them. Santarelli and Vivarelli (2007) point to the deadweight and substitution effects from entry subsidies. Even though our results confirm Haltiwanger et al. (2013) and show that a small subset of fast-growing young firms contribute disproportionally to job creation, one should be careful when drawing policy inferences from this empirical regularity. Brown et al. (2015), for example, demonstrate that there is no clear link between firm-level growth rates and the constraints they experience. It is uncertain whether firms with the most potential would respond most strongly to policy interventions.

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 have restricted the analysis to firms in the private, non-farm sector and also exclude highly subsidized sectors which receive strong support from government programs.[30] 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% per year till 2008, dropped by 2.5% 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. Because the extent of the problem is register-specific, it also hampers comparative analysis.[31]

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.[32] 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.

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 statisti-

[30] 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.

[31] Davis et al. (1996b) and Baldwin et al. (1992) discuss the linkage problem in detail. Vilhuber (2008) and Bartelsman et al. (2009) discuss the implications for cross-country comparisons.

[32] See Benedetto et al. (2007) for a discussion of these practices in the U.S.

cal 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.[33]

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 et al. (1992) for Canada and implemented for the United States by Benedetto et al. (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% of the workforce of both the disappearing and the newly appearing ID code. A second

[33] 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.

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 panel (a) are met.

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

		Decision rules					
		Number of predecessors to successors	Predecessor type	Successor type	Minimum absolute cluster size *(n employees)*	Minimum relative cluster size	
						Share in predecessor employment (%)	*Share in successor employment (%)*
Type of linkage							
1.	ID-change	1 to 1	Any	Any	5	50	50
2.	Takeover 75%	1 to 1	Exit	Continuing	5	75	–
3.	Split-off 75%	1 to 1	Continuing	Entrant	5	–	75
4.	Takeover 50%	1 to 1	Exit	Continuing	10	50	–
5.	Split-off 50%	1 to 1	Continuing	Entrant	10	–	50
6.	Merger of exits	n to 1	All exits	Entrant	5	50[a]	50
7.	Break-up into entrants	1 to n	Exit	All entrants	5	50	50[b]
8.	Merger other	n to 1	Any	Entrant	5	–	25[c], 50[d]
9.	Break-up other	1 to n	Exit	Any	5	25[c], 50[d]	–
10.	Cluster >= 30	1 to 1	Any	Any	30	10	10

(b) Share of employee-flow linkages by type

		All links	Spurious entrants
1.	ID-change	0.57	0.78
2.	Takeover 75%	0.22	–
3.	Split-off 75%	0.12	0.18
4.	Takeover 50%	0.01	–
5.	Split-off 50%	0.01	0.01
6.	Merger of exits	0.01	–
7.	Break-up into entrants	0.01	0.01
8.	Merger other	0.01	0.01
9.	Break-up other	0.00	0.00
10.	Cluster >= 30	0.03	0.00

Note: total sums to one in each column. Annual averages over the sample period.
[a] Share of the sum of the clusters in total employment of the predecessors.
[b] Share of each individual cluster in employment of successor.
[c] Share of each individual cluster.
[d] Share of the sum of the clusters.

rule identifies takeovers, allowing the receiving ID code to exist already, but requiring a cluster of at least 75% 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. When estimating the post-entry patterns of young firms, spurious entrants are removed from the sample. Panel (b) of Table A.1 shows that 78% 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% are split-offs of another firm.[34] 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 *de novo* entrants that change identification code or restructure in the years following entry 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.[35] 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 shows 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 in larger size classes, where it identifies two to three times more spurious entrants than the traditional method. Table 1 further shows that the probability that a new ID code

[34] 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% of all *de novo* entrants

[35] 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.

corresponds to a spurious entrant 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% of administrative exits are identified as spurious, but the likelihood is again increasing with firm size. In the working paper, see Geurts and Van Biesebroeck (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.[36] 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 et al. (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 $+0.67$.[37] This approach mitigates the negative bias in the size–growth relationship caused by regression-to-the-mean because symmetric growth and decline are

[36] For a discussion see for example Hall (1987), Evans (1987b), Baldwin and Picot (1995), Davis et al. (1996b), Davidsson et al. (1998), and Kirchhoff and Greene (1998). 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).

[37] The size thresholds between the size classes [0,2], [2,4], [4,8], [8,16], [16,32], and [32,∝] 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,∝], respectively.

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 et al., 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 on average positive growth rates, it introduces an upward bias between size and growth (Baldwin and Picot, 1995). [38] 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 Fig. A.8, 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% 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.

[38] For further discussion see also Davidsson et al. (1998) and Kirchhoff and Greene (1998).

Table A.2
Summary statistics of entrants.

	Entry rate	Employment share	Exit rate	Share of survivors	Employment share of survivors	Average size (employees)
	(a) *De novo* entrants					
Age 1 (entry)	0.088	0.015		1.00	1.00	1.93
Age 2			0.21	0.79	0.98	2.39
Age 3			0.15	0.68	0.97	2.78
Age 4			0.13	0.60	0.95	3.10
Age 5			0.11	0.54	0.93	3.38
Age 6			0.10	0.49	0.92	3.61
	(b) All administratively recorded entrants					
Age 1 (entry)	0.097	0.026		1.00		
Age 2			0.20	0.80	0.98	3.84
Age 3			0.15	0.68	0.95	4.38
Age 4			0.13	0.60	0.93	4.89
Age 5			0.11	0.54	0.91	5.29
Age 6			0.10	0.49	0.87	5.63

Note: annual averages over the sample period. The year a firm enters the dataset is indicated by age 1.

C. Confirmed patterns

Empirical studies for various countries have found entry rates of new firms in manufacturing and services to vary between 5 and 15% 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% 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 5–10 years after entry, average firm size has doubled, but only half of an entry cohort survives.[39] 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).

Our results for Belgium confirm these empirical regularities. 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 8.8% of all active employer firms in a given year, but only 1.5% of total employment. 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 firm size among survivors increases. Only half of all entrants are still around at age 6, at which

[39] See for example Dunne et al. (1988) for the U.S., Wagner (1994) and Boeri and Cramer (1992) for Germany; Mata et al. (1995) for Portugal.

Fig. A.1. Confirmed predictions of the passive learning model.
(a) Note: the figure plots the coefficients on age-size category interaction dummies from the exit regressions using the size in $t-1$ to classify firms by size. Age 6+ refers to incumbents.
(b) Note: the figure plots the coefficients on the firm-age category dummies from the employment growth regressions for surviving firms, without including size dummies. Age 6+ refers to incumbents.

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 Fig. A.9. 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.

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 entry. Two patterns are highly robust: (i) exit rates are decreasing in firm size and (ii) survival rates increase as firms mature.[40]

In line with previous studies and 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 Fig. A.1, which plots the age–size coefficients for the exit regression representing job destruction rates for each age–size class.[41] 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.

Another well-established fact is that young surviving firms exhibit remarkably high growth rates which decline with age.[42] Panel (b) of Fig. A.1 shows this 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 Fig. 2 in the text. Average growth rates among all firms surviving from year $t-1$ to t decline monotonically with the current size 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 Fig. 2 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

[40] See for example Evans (1987a) and Dunne et al. (1989) for U.S. manufacturing plants, Haltiwanger et al. (2013) for U.S. manufacturing and services; Mata et al. (1995) for Portugal.
[41] Recall that all regression coefficients are estimated using employment weights.
[42] See the same studies for the U.S.; Mata and Portugal (2004) for Portugal.

when no or little attempt has been made to distinguish between what we have labeled *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 Fig. 2, the raw, administrative data suggest that small young firms have higher growth rates than larger ones.

D. Additional figures and tables

Table A.3
Growth rates of surviving *de novo* entrants by age and size. Standard errors in parentheses.

	Firm size class (number of employees)					
De novo entrants	1	2–4	4–7	8–15	16–31	32+
(a) Dynamic size classification						
Age 2	0.133 (.002)	0.150 (.002)	0.163 (.002)	0.191 (.003)	0.196 (.004)	0.234 (.005)
Age 3	0.068 (.003)	0.067 (.002)	0.077 (.002)	0.101 (.003)	0.113 (.004)	0.142 (.004)
Age 4	0.037 (.004)	0.038 (.003)	0.049 (.003)	0.070 (.003)	0.069 (.004)	0.081 (.004)
Age 5	0.026 (.004)	0.025 (.003)	0.029 (.003)	0.038 (.003)	0.038 (.004)	0.059 (.004)
Age 6	0.019 (.005)	0.014 (.003)	0.025 (.003)	0.028 (.003)	0.033 (.004)	0.032 (.004)
Incumbents						
Age 6+	0.006 (.001)	0.003 (.001)	0.004 (.000)	0.004 (.000)	0.004 (.000)	0.000 (.000)
(b) Average of estimates using firm base size at $t-1$ and t						
Age 2	0.166 (.004)	0.198 (.004)	0.229 (.004)	0.272 (.006)	0.312 (.007)	0.309 (.008)
Age 3	0.048 (.004)	0.083 (.003)	0.105 (.003)	0.145 (.004)	0.167 (.005)	0.161 (.005)
Age 4	0.009 (.005)	0.046 (.004)	0.066 (.003)	0.094 (.004)	0.086 (.005)	0.104 (.005)
Age 5	0.002 (.005)	0.023 (.004)	0.042 (.003)	0.051 (.004)	0.065 (.005)	0.061 (.004)
Age 6	−0.011 (.006)	0.015 (.004)	0.034 (.003)	0.038 (.004)	0.038 (.005)	0.042 (.005)
Incumbents						
Age 6+	−0.029 (.001)	−0.006 (.001)	0.002 (.001)	0.006 (.001)	0.008 (.001)	0.001 (.000)
(c) Average size classification						
Age 2	0.085 (.004)	0.216 (.003)	0.263 (.004)	0.334 (.005)	0.372 (.007)	0.373 (.008)
Age 3	0.040 (.004)	0.082 (.004)	0.117 (.004)	0.148 (.005)	0.162 (.006)	0.191 (.006)
Age 4	0.016 (.005)	0.038 (.004)	0.070 (.004)	0.102 (.005)	0.113 (.006)	0.091 (.006)
Age 5	0.007 (.006)	0.024 (.004)	0.037 (.004)	0.063 (.005)	0.046 (.006)	0.069 (.005)
Age 6	0.003 (.007)	0.005 (.005)	0.034 (.004)	0.041 (.005)	0.056 (.006)	0.035 (.006)
Incumbents						
Age 6+	−0.011 (.001)	−0.005 (.001)	0.002 (.001)	0.006 (.001)	0.008 (.001)	0.000 (.000)

Note: the tables show the coefficients (and standard errors) on age–size category interaction dummies from the employment growth regressions for surviving firms.

Table A.4
Six main industries and NACE Rev. 2 classes.

Nace Rev. 2 classes	*De novo* entrants		
	Number of firms	Number of employees	Average entry size (employment)
1. Manufacturing and energy	777	1996	2.6
Section B, C, D, E			
2. Construction	2730	5150	1.9
Section F			
3. Wholesale and retail trade	4236	7497	1.8
Section G			
4. Accomodation and food services	2793	6570	2.4
Section I			
5. Business support services	2945	5143	1.7
- Freight transport, handling and storage: Nace 49.2, 49.4, 49.5, 50.2, 50.4, 51.2, 52.1, 52.241,52.249;			
- IT programing and services: Nace 62, 63;			
- 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;			
- Accounting: Nace 69.2;			
- Head offices: Nace 70;			
- Architecture and engineering: Nace 71;			
- Advertising: Nace 73;			
- Professional and technical support services: Nace 74;			
- Professional rental and leasing: Nace 77.1, 77.3, 77.4;			
- Security: Nace 80;			
- Services to buildings except Cleaning: Nace 81 excl. 81.210, 81.220			
- Administrative services: Nace 82;			
- Repair of ICT: Nace 95.1			
6. Mixed business and household services	2011	3459	1.7
- Passenger transport and transport services: Nace 49.1, 49.3, 50.1, 50.3, 51.1, 52.210, 52.220, 52.230, 52.290;			
- Postal and courier activities: Nace 53;			
- Publishing, movies, radio and television: Nace 58, 59, 60;			
- Telecommunication: Nace 61;			
- Banks, credit, insurance instit.: Nace 64.19, 64.921, 64.92, 65;			
- Real estate: Section L;			
- Legal activities: Nace 69.1;			
- Scientific research: Nace 72;			
- Veterinary: Nace 75;			
- Rental and leasing of household goods: Nace 77.2;			
- Travel agencies: Nace 79;			
- Repair of household goods: Nace 95.2;			
- Personal service activities: Nace 99			
Total	15,492	29,815	1.9

Note: number 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. Annual averages over the sample period.

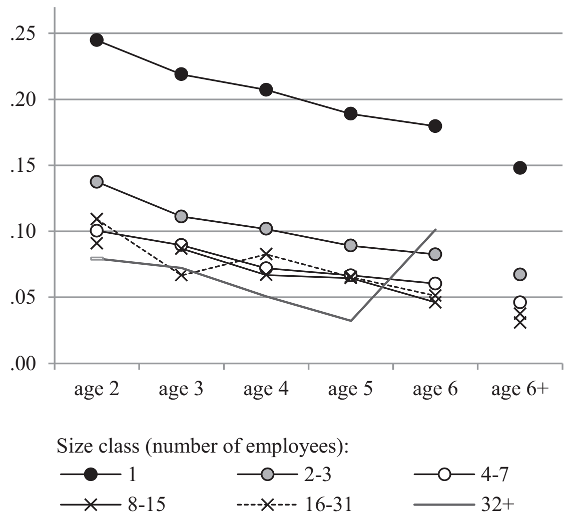

Fig. A.2. Exit rates of all administratively recorded entrants by age and size.
Note: the figure plots the coefficients on age–size category interaction dummies from the exit regressions using the size in $t-1$ to classify firms by size. Age 6+ refers to incumbents.

Fig. A.3. Growth rates of surviving firms by size using raw administrative data: young firms versus incumbents.
Note: the figure plots the coefficients on the firm-size category dummies (using the dynamic size classification) from the employment growth regressions for surviving firms, without including age dummies and estimated separately for each sample. For young firms the 32–63 size class is really 32+.

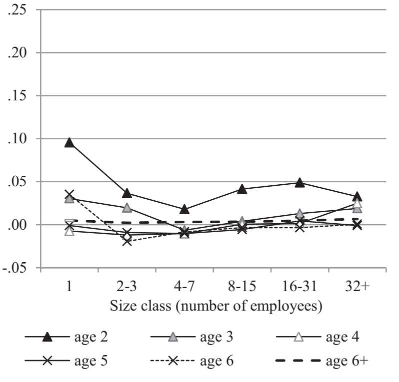

Fig. A.4. Growth rates of surviving spurious entrants by age and size.
Note: the figure plots the coefficients on age–size category interaction dummies from the employment growth regressions for surviving firms using the dynamic size classification. Age 6+ refers to incumbents.

Fig. A.5. Growth rates of surviving *de novo* entrants by age and size when post-entry employment is not imputed.
Note: the figure plots the coefficients on age–size category interaction dummies from the employment growth regressions for surviving firms using the dynamic size classification. Age 6+ refers to incumbents.

Fig. A.6. Growth rates of surviving *de novo* entrants by age and size using base year size classification.
Note: the figure plots the coefficients on age–size category interaction dummies from the employment growth regressions for surviving firms using the base-year size classification. Point estimates and standard errors are reported in Table A.3. Age 6+ refers to incumbents.

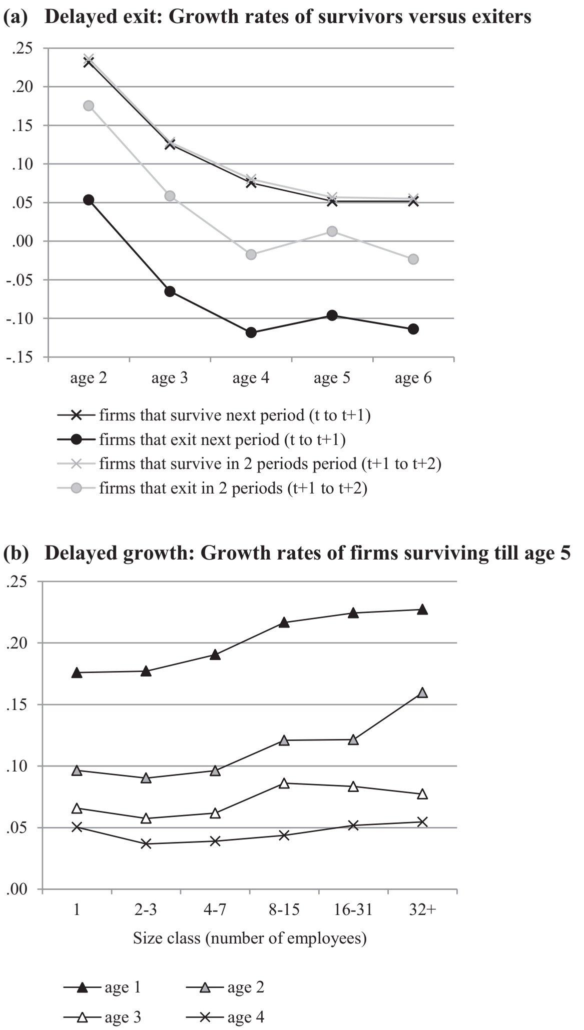

Fig. A.7. Delayed adjustment of *de novo* entrants in exit and growth.
(a) Note: the figure plots the coefficients on the firm-age category dummies from the employment growth regressions for surviving firms without including size dummies. Age 6+ refers to incumbents.
(b) Note: the figure plots the coefficients on age-size category interaction dummies from the employment growth regressions for surviving firms for the subset of *de novo* entrants that survive till age 5. We use the dynamic size classification to classify firms by size.

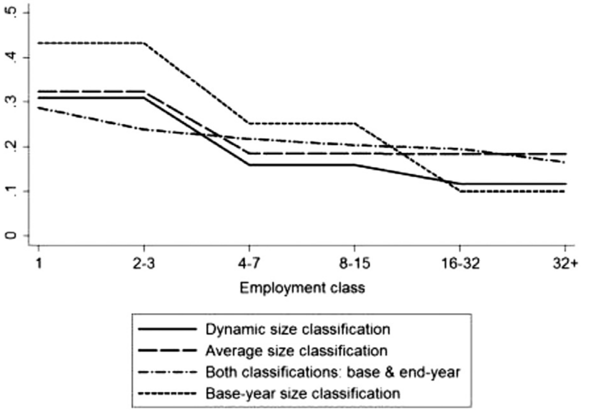

Fig. A.8. Estimated size–growth relationships on simulated data with constant growth rate.

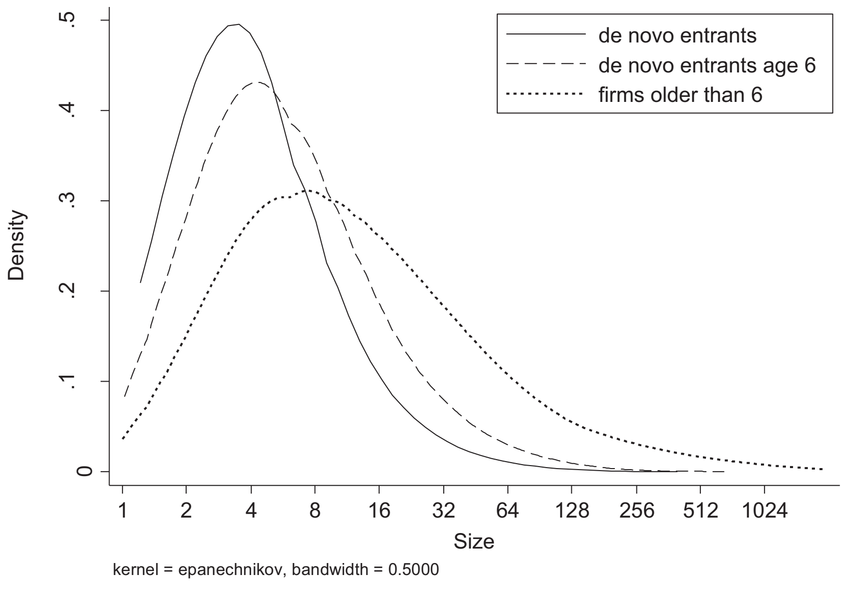

Fig. A.9. Evolution of the firm size distribution.

References

Abbring, J.H. and J.R. Campbell (2005). "A Firm's First Year." Tinbergen Institute Discussion Paper No. 2005-046/3.

Acs, Z., Astebro, T., Audretsch, D., Robinson, D.T., 2016. Public policy to promote entrepreneurship: a call to arms. Small Business Economics 47, 35–51.

Almus, M., Nerlinger, E.A., 2000. Testing "Gibrat's law" for young firms – empirical results for West Germany. Small Business Economics 15, 1–12.

Angrist, J.D., Pischke, J.S., 2009. Mostly Harmless Econometrics. An Empiricist's Companion. Princeton University Press.

Asker, J., Collard-Wexler, A., De Loecker, J., 2014. Dynamic inputs and resource (mis)allocation. Journal of Political Economy 122 (5), 1013–1063.

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

Aw, B.Y., Roberts, M.J., Yi Xu, D., 2011. R&D investment, exporting, and productivity dynamics. American Economic Review 101 (4), 1312–1344.

Baldwin, J.R., Dupuy, R., Penner, W., 1992. Development of longitudinal panel data from business registers: the Canadian experience. Statistical Journal of the United Nations 9, 289–303.

Baldwin, J.R., Gorecki, P.K., 1987. Plant creation versus plant acquisition. The entry process in Canadian manufacturing. International Journal of Industrial Organization 5, 27–41.

Baldwin, J.R., Picot, G., 1995. Employment generation by small producers in the Canadian manufacturing sector. Small Business Economics 7 (4), 317–331.

Bartelsman, E., Haltiwanger, J.C., Scarpetta, S., 2009. Measuring and analyzing cross-country differences in firm dynamics. In: Dunne, T., Jensen, J.B., Roberts, M.J. (Eds.), Producer Dynamics: New Evidence from Micro Data. University of Chicago Press.

Beck, T., Demirgüç-Kunt, A., Laeven, L., Maksimovic, V., 2006. The determinants of financing obstacles. Journal of International Money and Finance 25, 932–952.

Benedetto, G., Haltiwanger, J.C., Lane, J., McKinney, K., 2007. Using worker flows to measure firm dynamics. Journal of Business and Economic Statistics 25 (3), 299–313.

Bilsen, V., Konings, J., 1998. Job creation, job destruction, and growth of newly established, privatized, and state-owned enterprises in transition economies: survey evidence from Bulgaria, Hungary, and Romania. Journal of Comparative Economics 26 (3), 429–445.

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

Bonnet, J., Cressy, R., 2016. An Encompassing Model & Test of the Evans-Jovanovic Credit Constraints Hypothesis. Center for Research in Economics and Management, Université de Caen Normandie and Université de Rennes 1 WP 2016-01.

Brockhaus, R.H., 1980. Risk taking propensity of entrepreneurs. Academy of Management Journal 23 (3), 509–520.

Brown, J.D., Earle, J.S., Morgulis, Y., 2015. Job Creation, Small vs. Large vs. Young, and the SBA. NBER Working Paper No. 21733.

Butani, S.J., Clayton, R.L., Kapani, V., Spletzer, J., Talan, D.M., Werking Jr, G.S., 2006. Business employment dynamics: tabulations by employer size. Monthly Labor Review 2 (1), 3–22.

Cabral, L., 1995. Sunk costs, firm size and firm growth. Journal of Industrial Economics 43 (2), 161–172.

Cabral, L., Mata, J., 2003. On the evolution of the firm size distribution: facts and theory. American Economic Review 93 (4), 1075–1090.

Caves, D.W., Christensen, L.R., Diewert, W.E., 1982. The economic theory of index numbers and the measurement of input, output, and productivity. Econometrica 50 (6), 1393–1414.

Caves, R.E., 1998. Industrial organization and new findings on the turnover and mobility of firms. Journal of Economic Literature 36 (4), 1947–1982.

Davidsson, P., Lindmark, L., Olofsson, C., 1998. The extent of overestimation of small firm job creation – an empirical examination of the regression bias. Small Business Economics 11 (1), 87–100.

Davis, S.J., Haltiwanger, J.C., Schuh, S., 1996a. Job Creation and Destruction. MIT Press.

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

de Wit, G., de Kok, J., 2014. Do small businesses create more jobs? New evidence for Europe. Small Business Economics 42 (2), 283–295.

Decker, R., Haltiwanger, J., Jarmin, R., Miranda, J., 2014. The role of entrepreneurship in US job creation and economic dynamism. Journal of Economic Perspectives 28 (3), 3–24.

Dunne, T., Roberts, M.J., Samuelson, L., 1988. Patterns of firm entry and exit in U.S. manufacturing industries. RAND Journal of Economics 19 (4), 495–515.

Dunne, T., Roberts, M.J., Samuelson, L., 1989. The growth and failure of U.S. manufacturing plants. Quarterly Journal of Economics 104 (4), 671–698.

Ericson, R., Pakes, A., 1995. Markov-perfect industry dynamics: a framework for empirical work. Review of Economic Studies 62 (1), 53–82.

European Commission (2013). “The entrepreneurship 2020 action plan. Reigniting the entrepreneurial spirit in Europe.” (Available online at http://ec.europa.eu/growth/smes/promoting-entrepreneurship).

Eurostat-OECD (2007). Manual on business demography statistics. (Available online at http://www.oecd.org/dataoecd/8/8/39974460.pdf).

Evans, D.S., 1987a. The relationship between firm growth, size, and age: estimates for 100 manufacturing industries. Journal of Industrial Economics 35 (4), 567–581.

Evans, D.S., 1987b. Tests of alternative theories of firm growth. Journal of Political Economy 95 (4), 657–674.

Evans, D.S., Jovanovic, B., 1989. An estimated model of entrepreneurial choice under liquidity constraints. Journal of Political Economy 97 (4), 808–827.

Evans, D.S., Leighton, L.S., 1989. Some empirical aspects of entrepreneurship. American Economic Review 79 (3), 519–535.

Fazzari, S, Hubbard, R.G., Petersen, B., 1988. Financing constraints and corporate investment. Brookings Papers on Economic Activity 1, 141–206.

Friedman, M., 1992. Do old fallacies ever die? Journal of Economic Literature 30 (4), 2129–2132.

Geroski, P.A., 1995. What do we know about entry? International Journal of Industrial Organization 13 (4), 421–440.

Geurts, K., 2016. Longitudinal firm-level data: problems and solutions. Small Business Economics 46 (3), 425–445.

Geurts, K., Van Biesebroeck, J., 2014. Job Creation, Firm Creation, and de novo Entry. CEPR Discussion Paper No. 10118.

Guner, N, Ventura, G., Xi, Y., 2008. Macroeconomic implications of size-dependent policies. Review of Economic Dynamics 11, 721–744.

Hall, B.H., 1987. The relationship between firm size and firm growth in the U.S. manufacturing sector. Journal of Industrial Economics 35 (4), 583–606.

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

Hashmi, A.R., Van Biesebroeck, J., 2016. The relationship between market structure and innovation in industry equilibrium: a case study of the global automobile industry. Review of Economics and Statistics 98 (1), 192–208.

Helfat, C.E., Lieberman, M.B., 2002. The birth of capabilities: Market entry and the importance of prehistory. Industrial and Corporate Change 11 (4), 725–760.

Hethey, T., Schmieder, J., 2013. Does the use of Worker Flows Improve the Analysis of Establishment Turnover? Evidence from German Administrative Data. NBER Working Paper No. 19730.

Hopenhayn, H., 1992. Entry, exit, and firm dynamics in long run equilibrium. Econometrica 60 (5), 1127–1150.

Hsieh, C.-T., Klenow, P.J., 2009. Misallocation and manufacturing TFP in China and India. Quarterly Journal of Economics 124 (4), 1403–1448.

Hurst, E., Pugsley, B.W., 2011. What do Small Businesses do?. NBER Working paper 17041.

Jarmin, R.S., Miranda, J., 2002. The Longitudinal Business Database. Center for Economic Studies Discussion Paper No. CES-WP-02-17.

Jovanovic, B., 1982. Selection and the evolution of industry. Econometrica 50 (3), 470–649.

Kihlstrom, R.E., Laffont, J.-J., 1979. A general equilibrium entrepreneurial theory of firm formation based on risk aversion. Journal of Political Economy 87 (4), 719–748.

Kirchhoff, B.A., Greene, P.G., 1998. Understanding the theoretical and empirical content of critiques of U.S. job creation research. Small Business Economics 10, 153–169.

Konings, J., 1995. "Gross job flows and the evolution of size in UK establishments. Small Business Economics 7 (3), 213–220.

Konings, J., Lehmann, H., Schaffer, M., 1996. Job creation and job destruction in a transition economy: ownership, firm size, and gross job flows in Polish manufacturing 1988–91. Labour Economics 3 (3), 299–317.

Kuusisto, J., 2008. R&D in Services – Review and Case Studies. European Commission Paper submitted for the CREST R&D in Services Working Group.

Lotti, F., Santarelli, E., Vivarelli, M., 2003. Does Gibrat's law hold among young, small firms? Journal of Evolutionary Economics 13 (3), 213–235.

Lucas Jr., R.E., 1978. On the size distribution of business firms. Bell Journal of Economics 9 (2), 508–523.
Mansfield, E., 1962. Entry, Gibrat's law, innovation, and the growth of firms. American Economic Review 52 (5), 1023–1051.
Mata, J., 1994. Firm growth during infancy. Small Business Economics 6 (1), 27–39.
Mata, J., Portugal, P., 2004. Patterns of entry, post-entry growth and survival. Small Business Economics 22 (3–4), 283–298.
Mata, J., Portugal, P., Guimaraes, P., 1995. The survival of new plants: start-up conditions and post-entry evolution. International Journal of Industrial Organization 13 (4), 459–481.
Pakes, A., Ericson, R., 1998. Empirical implications of alternative models of firm dynamics. Journal of Economic Theory 79 (1), 1–45.
Parker, S.C., 1996. A time series model of self-employment under uncertainty. Economica 63 (251), 459–475.
Prais, S.J., 1958. The statistical conditions for a change in business concentration. Review of Economics and Statistics 40 (3), 268–272.
Santarelli, E., Vivarelli, M., 2007. Entrepreneurship and the process of firms' entry, survival and growth. Industrial and Corporate Change 16 (3), 455–488.
Siegfried, J.J., Evans, L.B., 1994. Empirical studies of entry and exit: a survey of the evidence. Review of Industrial Organization 9 (2), 121–155.
Vilhuber, L., 2008. Adjusting imperfect data: overview and case studies. In: Lazear, E.P., Shaw, K.L. (Eds.), The Structure of Wages: An International Comparison. University of Chicago Press.
Vivarelli, M., Audretsch, D., 1998. The link between the entry decision and post-entry performance: evidence from Italy. Industrial and Corporate Change 7 (3), 485–500.
Wagner, J., 1994. The post-entry performance of new small firms in German manufacturing industries. Journal of Industrial Economics 42 (2), 141–154.